Individual Differences in Lexical Learning Across Two Language Modalities: Sign Learning, Word Learning, and Their Relationship in Hearing Non-Signing Adults

David Martinez[1] and Jenny L. Singleton[2]

School of Psychology, Georgia Institute of Technology

Corresponding author:

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

This is a preprint of an article published by Elsevier in Acta Psychologica available online at https://doi.org/10.1016/j.actpsy.2019.102892

Martinez, D., & Singleton, J. (2019). Individual differences in lexical learning across two language modalities: Sign learning, word learning, and their relationship in adult non-signers. Acta Psychologica. https://doi.org/10.1016/j.actpsy.2019.102892

[1] David Martinez is now at the Applied Research Laboratory for Intelligence and Security (ARLIS), University of Maryland, College Park.
[2] Jenny L. Singleton is now at the Department of Linguistics, University of Texas at Austin

Abstract

A considerable amount of research has been devoted to understanding individual differences in lexical learning, however, the majority of this research has been conducted with *spoken* languages rather than *signed* languages and thus we know very little about the cognitive processes involved in sign learning or the extent to which lexical learning processes are specific to *word* learning. The present study was conducted to address this gap.

Two-hundred thirty-six non-signing adults completed 25 tasks assessing word learning and sign learning (via associative learning paradigms) as well as modality-specific phonological short-term memory, working memory capacity, crystallized intelligence, and fluid intelligence.

Latent variable analyses indicated that, when other variables were held constant, fluid intelligence was predictive of both word and sign learning, however, modality-specific phonological short-term memory factors were only predictive of lexical learning within modality—none of the other variables made significant independent contributions. It was further observed that sign and word learning were strongly correlated. Exploratory analyses revealed that all lexical learning tasks loaded onto a general factor, however, sign learning tasks loaded onto an additional specific factor. As such, this study provides insight into the cognitive components that are common to associative L2 lexical learning regardless of language modality and those that are unique to either signed or spoken languages. Results are further discussed in light of established and more recent theories of intelligence, short-term memory, and working memory.

Keywords: Associative learning; lexical learning; phonological short-term memory; sign language

1 Introduction

We have all engaged in associative lexical learning at one point or another—perhaps while studying vocabulary for a test or to prepare for a trip to a country where a foreign language is spoken. While learning in this *decontextualized* manner is not how we have developed the bulk of our lexicon (Hulstijn, 2003; Krashen, 1989; Nation, 1980), it does have its place. Associative lexical learning allows one to select items for focused study, facilitating long-term retention (Rohrer, Taylor, Pashler, Wixted, & Cepeda, 2005; Seibert, 1930; Thorndike, 1908) and fluent use (Elgort, 2011; Yang, 1997), even before mastering the phonology or grammar of a target language, as is the case in second language learning. Moreover, associative lexical learning ability has been found to correlate moderately to strongly with other linguistic variables—such as grammar learning (Cooper, 1964; Gardner & Lambert, 1965; K. I. Martin & Ellis, 2012; O'Brien, Segalowitz, Collentine, & Freed, 2006), second language (L2) learning aptitude (Cooper, 1964; Li, 2015), and verbal ability (Hundal & Horn, 1977)—and more generally with intelligence (Hundal & Horn, 1977; Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009; Lilienthal, Tamez, Myerson, & Hale, 2013; Tamez, Myerson, & Hale, 2008; B. A. Williams & Pearlberg, 2006).

Importantly, however, the vast majority of individual differences research on lexical learning in adults has been conducted with *spoken* languages (e.g., Hundal & Horn, 1977; Kyllonen, Tirre, & Christal, 1991; Underwood, Boruch, & Malmi, 1978) and has largely overlooked lexical learning in *signed* languages (cf., Martinez & Singleton, 2018; Stone, 2017; J. T. Williams, Darcy, & Newman, 2016a). As a consequence, we know very little about the cognitive factors engaged while learning signs and whether they are similar and relied upon to the same degree as those employed during word learning. This is unfortunate given the popularity of American Sign Language as an L2 in secondary (Pufahl & Rhodes, 2011) and postsecondary US schools (Goldberg, Looney, & Lusin, 2015) and the significant advances that can and have been made via the comparative study of signed and spoken languages (e.g., Bavelier et al., 2008; Campbell, MacSweeney, & Waters, 2008; Cardin et al., 2016; Emmorey, 2002; Hirshorn, Fernandez, & Bavelier, 2012; Klima & Bellugi, 1979; Malaia & Wilbur, 2018; Mayberry, 2010; Poizner, Klima, & Bellugi, 1987; Wilson, 2001).

With the above in mind, the present study was conducted to extend individual differences research from L2 word learning to L2 sign learning and to examine the relationship between the two constructs in hearing non-signing adults. A necessary first step was to consider the components involved in associative lexical learning. In brief, effective lexical learning relies on encoding and maintaining information via domain-specific and general processes long enough and with sufficient fidelity to generate associations and encode them in long-term memory. As will be reviewed below, these components implicate phonological short-term memory, working memory, and complementary action by fluid and crystallized intelligence.

1.2 Phonological Short-Term Memory in Signed and Spoken Languages

Phonological short-term memory (PSTM) refers to the ability to encode verbal (or *phonological*) information and retain it in some form for a brief period of time. A large body of research indicates that this ability to maintain phonological information is necessary for the explicit learning of lexical forms in L1 vocabulary development and L2 lexical learning (Baddeley, Papagno, & Vallar, 1988; Gupta, 2003; Hummel & French, 2016; K. I. Martin & Ellis, 2012; O'Brien et al., 2006; O'Brien, Segalowitz, Freed, & Collentine, 2007; Papagno, Valentine, & Baddeley, 1991; Papagno & Vallar, 1992), presumably because short-term storage facilitates long-term retention mechanisms (Baddeley, Gathercole, & Papagno, 1998).

Spoken-PSTM is typically assessed via span tasks in which individuals are asked to remember sets of verbal items (e.g., words, digits, or pseudowords) and recall them in the order they were presented. *Signed*-PSTM is assessed in a similar fashion as spoken-PSTM (e.g., Bellugi, Klima, & Siple, 1974; Boutla, Supalla, Newport, & Bavelier, 2004; Conrad, 1970), however, rather than maintaining sequences of speech-sounds, individuals maintain one or more signs composed of the simultaneous presentation of the following major phonological parameters: handshape, movement, and location (Brentari, 1998; Klima & Bellugi, 1979). A fourth parameter, hand orientation, is incorporated in some models of sign phonology as an independent parameter (e.g., Brentari, 1998) and in others as a feature of handshape (e.g., Sandler, 1989).

To illustrate sign phonology, consider the American Sign Language (ASL) signs APPLE and CANDY. APPLE is signed by touching the knuckle of the flexed index finger to the cheek and rotating the wrist back and forth (see Figure 1). The ASL sign CANDY is articulated in the same location (touching the cheek) and with the same movement (rotating the wrist) but the index finger is fully extended. Thus, in ASL, APPLE and CANDY are minimal pairs (lexical items that differ by one phoneme; for an English example, consider *rat* and *bat*).

Figure 1. ASL APPLE

At least one study has found that *signed*-PSTM tasks, utilizing sign-like material, are related to sign learning (Martinez & Singleton, 2018). Given the scarcity of research investigating signed-PSTM and sign learning, it is also worth noting that signed-PSTM tasks are related to other language outcomes in deaf children who sign (Marshall et al., 2015; Mason et al., 2010) as well as in hearing sign language interpreters (Gómez, Molina, Benítez, & de Torres, 2007; Shaw, 2011)—relationships that are analogous to those observed in spoken language research (Daneman & Merikle, 1996; Gathercole & Baddeley, 1990).

Whether signed-PSTM and spoken-PSTM are facets of a single construct or are independent is an empirical question. Gathercole (2006) theorized that PSTM performance is multiply determined by phonological, perceptual, and motor processes. In theory, phonological processing is amodal, as the information being processed are abstract linguistic units (Baddeley, 2015; Baddeley et al., 1998; for a counter argument, see Jones, Hughes, & Macken, 2006). In fact, neuroimaging studies have revealed that the same "classic language areas" that are activated by spoken language processing are active during sign language processing in *fluent* signers (Bavelier et al., 1998; Söderfeldt et al., 1997; J. T. Williams, Darcy, & Newman, 2015). However, there is evidence that hearing non-signers do not immediately process signs linguistically, instead processing them as nonverbal movements (Martinez & Singleton, 2018; Newman-Norlund, Frey, Petitto, & Grafton, 2006; Siple, Caccamise, & Brewer, 1982; J. T.

Williams, Darcy, & Newman, 2016b). The other two common processing components, perceptual and motor processes, are undoubtedly different across signed and spoken languages. The lack of phonological processing in non-signers learning signs along with differences in the perceptual and motor processes recruited to perceive and produce languages across modalities implies that in hearing non-signers, PSTM for signed material relies on processes that are at least partially distinct from those utilized to encode and maintain spoken language. As such, one would expect to observe different relationships between signed-PSTM and spoken-PSTM and criterion measures.

1.3 Working Memory Capacity

Working memory capacity (WMC) is defined and operationalized in a variety of ways (Cowan, 2017; Oberauer et al., 2018). Here, WMC is defined as a domain-general ability that allows individuals to maintain a limited amount of information in a highly accessible state, even in the face of interference (Engle, 2002; Shipstead, Harrison, & Engle, 2016); it is best assessed by tasks that require short-term retention of information *and* prevent or disrupt motor rehearsal such as speech-motor (i.e., articulatory; Baddeley, Thomson, & Buchanan, 1975) or gaze-based (Tremblay, Saint-Aubin, & Jalbert, 2006) rehearsal, forcing individuals to rely on the control of attention, or *executive attention*, to maintain durable representations (Cowan, 2008; see also La Pointe & Engle, 1990, p. 1130).

To be sure, WMC, as defined here, is similar to STM and therefore PSTM—both WMC and STM are defined in part by the ability to maintain information in memory for a brief period of time. Moreover, WMC and STM tasks are often operationalized similarly (Cowan, 2008). In fact, modeling studies investigating the relationship between WMC and STM have observed correlations approaching unity (e.g., Colom, Shih, Flores-Mendoza, & Quiroga, 2006), however, at the latent variable level, researchers generally find correlations equal to or less than .80 (Cowan, 2008; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004; Unsworth & Engle, 2007), suggesting that the two are highly related but different constructs.

The relationship between WMC and STM is at least partly due to the fact that both are supported by executive attention (Kane et al., 2004; Unsworth, 2010). The two are distinguished, however, by the fact that WMC depends on executive attention to a greater degree than STM and STM tends to depend on domain-specific processes to a greater degree than WMC (Kane et al., 2004). The distinction between WMC and STM is further supported by studies reporting independent contributions from WMC and STM to language-based outcomes (e.g., Cantor, Engle, & Hamilton, 1991; Engle et al., 1999; K. I. Martin & Ellis, 2012; Verhagen & Leseman, 2016).

Notably, researchers have found that WMC is predictive of L2 word learning (K. I. Martin & Ellis, 2012) and spoken L2 learning more generally (Li, 2015; Linck, Osthus, Koeth, & Bunting, 2014), though, research on WMC as a predictor of sign learning is scant. As with the relationship between WMC and STM, the relationship between WMC and lexical learning is likely due, in part, to the control of attention. To elaborate, one needs to control attention to stay focused on the task at hand and avoid attending to irrelevant information from the environment, our own thoughts, or from within the task itself; when our attention is pulled to irrelevant stimuli, then the encoding of target material is compromised and interference increases (Hasher & Zacks, 1988; Hasher, Zacks, & May, 1999). Poor encoding and increased interference leads to a lower probability of a correct response, be that in a memory task, lexical learning task, or some other task. Individuals with high WMC, however, are better able to use executive attention to prevent encoding failures and the accumulation of interference (Kane & Engle, 2000), resulting in easily accessible and durable memory representations (Shipstead & Engle, 2013; Unsworth, Brewer, & Spillers, 2013; Unsworth, Spillers, & Brewer, 2012; Wilhelm, Hildebrandt, & Oberauer, 2013).

1.4 Crystallized and Fluid Intelligence

According to the relation-construction principle, "the strength of a bond between a pair of items (which governs the success of retrieval of that pair) is determined by the *quantity* and *quality* of the relations constructed between the items during study (Kyllonen et al., 1991, p. 58)." The greater the number of relations formed—or the more *elaborative*—the greater the number of cues that can be used to retrieve the appropriate response. Of course, these relations are of little use if they do not uniquely index the items under study or if they have weak association values such that self-initiated cues are unlikely to be recalled at a later time (Glaze, 1928; Jenkins, 1985; Noble, 1952).

Crystallized intelligence and fluid intelligence are both implicated in the construction of relationships and therefore support associative lexical learning. Crystallized intelligence refers to acquired knowledge and skills (Cattell, 1943)—it provides the "network of facts and associations into which new facts and associations might be interwoven (Kyllonen & Woltz, 1989, p. 246)." Fluid intelligence refers to the ability to solve novel problems and reason in novel situations (Cattell, 1943). According to Shipstead et al. (2016), fluid intelligence tasks place a premium on the ability to *disengage* from outdated information. When inducing a relationship between familiar and unfamiliar information, an individual must consider possible relations and be able to abandon those that are inadequate, lest they block one from constructing more appropriate associations. Indeed, both crystallized and fluid intelligence generally show moderate relationships with associative learning (e.g., Hundal & Horn, 1977; Kyllonen & Tirre, 1988; Unsworth, 2019) and are predictive of L2 learning as well (Gardner & Lambert, 1965; Sasaki, 1993; Sparks, Humbach, Patton, & Ganschow, 2011).

To illustrate the impact of crystallized and fluid intelligence, suppose one was studying a list of words and their meanings and one of the items was *gloaming-twilight.* One may note that *gloaming* and *twilight* both have 8 letters but this is not likely unique to this pair of words; this relation may then be abandoned in favor of one that relates *gloaming* and *twilight* via "glow," which sounds similar to *gloaming* and relates to the level of light present at *twilight*. Assuming no other words in the list relate to dim lighting and/or sound similar to glow, then relating *gloaming* and *twilight* via *glow* will likely result in correct recall. Note, neither the word *glow* nor the concept of luminosity were explicit, rather, this information was drawn from prior knowledge and a relationship was induced. If it so happens that *gloomy* is another term in the list, then it would behoove one to abandon the previous relation (further implicating fluid intelligence) as *gloomy, gloaming,* and *glow* share sound similarities and all relate to dim lighting conditions, resulting in increased interference, and consequently affecting the likelihood of correct recall. While this example illustrates the learning of a low frequency English word (*gloaming*), we expect similar processes are engaged in L2 lexical learning (for an example using Turkish, see the discussion section).

To our knowledge, no study has investigated the relationship between fluid intelligence and sign learning in hearing individuals acquiring a sign language and only one study has investigated crystallized intelligence as a predictor. J. T. Williams et al. (2016a) administered an English vocabulary test (amongst other measures) to 25 individuals enrolled in an ASL course.

The English vocabulary test, an indicator of crystallized intelligence, was significantly related to sign learning, though caution must be taken given the small sample size.

1.5 The Present Study

The preceding review has provided evidence indicating that associative lexical learning is related to a number of other abilities, namely: PSTM, WMC, crystallized intelligence, and fluid intelligence. The vast majority of support for these claims, however, has come from research with spoken languages, leading one to question how these constructs relate to sign learning and if and how word learning is related to sign learning.

The present study had two aims. The first aim was to extend individual differences research in L2 word learning to the sign domain. The second aim was to directly investigate the relationship between sign learning and word learning in hearing non-signing adults. In order to accomplish these aims, structural equation modeling was used. Structural equation modeling is a statistical modeling technique that allows for the simultaneous estimation of relationships amongst a number of observed and latent variables (Loehlin, 1998).

2. Method

2.1 Participants

Participants were recruited from the Georgia Tech School of Psychology research participant pool and surrounding community, including local colleges and universities. Georgia Tech students received course credit and an additional $15 if they completed the entire study. Community participants received up to $65. In accordance with the Georgia Tech Institutional Review Board, informed consent was always obtained prior to participation.

In order to participate in the study, participants had to be between the ages of 17-35, fluent in English, have resided in the USA since at least the age of five, and have normal or corrected-to-normal hearing and vision. Due to the nature of the tasks and the aims of this study, participants were excluded if they indicated fluency in ASL or Turkish, were diagnosed with a language disorder, or if they possessed an upper-body injury or movement disorder that could affect their ability to rehearse movements (if they so chose to do so).

Our aim was to obtain a sample size of 240 individuals—the minimum sample size needed to detect a moderate effect size with alpha and power level of .05 and .80, respectively (Westland, 2010). In total, 286 individuals consented to participate in the study. Of those individuals, 34 did not return for the second session of the study, 13 indicated poor English fluency, and three individuals were removed from the analysis because they were observed answering their cell phone, copying to-be-remembered items, or skipping task instructions—the final sample consisted of 236 participants. Additionally, one individual indicated having studied ASL as a child but reported very limited fluency and so was retained; no other individual reported experience with a signed language.

Within the final sample, 234 answered a demographic questionnaire, though not necessarily all questions. Based on the information provided, the mean age was 21.24 years (SD = 3.57, n = 233); approximately 63% of individuals (147/233) identified as female; all individuals who provided information about their education (233) indicated that they had at least a high school diploma and nearly all (94.4%) indicated that they had at least some college education with 122 participants (52.3%) identifying as Georgia Tech students at the time of participation—the remaining 47.6% were community members, including students from local colleges and universities.

2.2 Procedure

The study consisted of two sessions, with nearly all tasks completed on a PC running E-Prime 3.0 (Schneider, Eschman, & Zuccolotto, 2002) in a room with up to five participants; only a reading test and demographic questionnaire were completed on paper.

The first session lasted up to 2.5 hours and consisted of eight associative lexical learning tasks (four sign learning tasks and four word learning tasks) and six PSTM tasks (three signed-PSTM tasks and three spoken-PSTM tasks). The second session lasted up to 2 hours and consisted of eight intelligence tasks (four crystallized intelligence tasks and four fluid intelligence tasks), three WMC tasks, an imagery questionnaire, the Object-Spatial Imagery Questionnaire (OSIQ; Blajenkova, Kozhevnikov, & Motes, 2006), and a language experience and demographics questionnaire. As is typical of individual differences research, task administration order was fixed and tasks were blocked by construct (see Table 1; for task descriptions see the following section), however, within these constraints, care was taken to minimize the effects of interference (e.g., alternating sign and spoken language tasks) and fatigue/motivation (e.g., placing an optional break after completing the somewhat monotonous WMC tasks). Note, the OSIQ and language experience portion of the language experience and demographic questionnaire are not relevant to the present study and will not be discussed further.

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 5 min Break]	[Optional 5 min 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 pseudoword 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.

2.3 Tasks

All computerized tasks developed in-house (those administered during Session 1) are available for download at osf.io/xmype. Computerized tasks began with instructions and at least one example item. Feedback was always provided during practice trials, however, the extent of the feedback ranged between simply stating whether the response was correct or giving a brief but detailed explanation. A research assistant was always present to observe participants as they completed each task and to answer any questions.

2.3.1 Lexical learning

All lexical learning tasks utilized a similar associative learning paradigm with blocks consisting of study and test trials. First, each task began with instructions introducing the task, followed by a single example item. Within the instructions, participants were told what language they were learning or, if items were contrived, they were told that the items were "fake." Next, participants were encouraged to use elaborative rehearsal strategies—either imagery or sentence generation—to aid their learning.

During the learning phase, a target item (word or sign) was presented aurally or visually, depending on the language modality being tested, and immediately followed by its associate, a single English word, presented on screen in its written form for 1000ms for all tasks except the three-term tasks, which presented the word for 2000ms. After a number of pairs were presented, the testing phase would begin.

During the testing phase, participants viewed randomly selected stimulus items followed by a response screen with all the English words encountered in the task. The participant was to click on the appropriate English word or guess. Once the participant made a response, the next item was presented, and so on. If the participant did not respond correctly to 100% of the items in a task then the task would continue with another block of trials until 100% of items were answered correctly or the maximum number of blocks (dependent on the task) was reached—whichever came first. Participants were never given explicit feedback or shown the correct associate during the test phase.

Scores were always calculated as the total number of correct responses across trials, however, because participants vary in the number of trials necessary to reach the criterion, superfluous trials were awarded the maximum number of points.

All English words 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 mimic what adults typically encounter when they first attempt to learn a new language.

Target L2 items were either drawn from natural languages, namely ASL and Turkish, or contrived. The decision to use items drawn from ASL and Turkish was made to provide evidence of ecological validity, while the decision to use pseudosigns and pseudowords was made to allow for greater control over item characteristics. All pseudowords were selected from the English Lexicon Project (Balota et al., 2007) and obeyed English phonotactics. Pseudosigns, on the other hand, were created using a parametric approach with handshape, location, and movement as parameters and did not necessarily adhere to the phonotactics of any particular language. Note well, as a helpful reviewer pointed out, in linguistics, the prefix *pseudo-* is typically used in reference to a particular language, as was the case for our *English* pseudowords; our pseudosigns, however, may be better termed nonsigns or simply gestures. For further discussion, see the section on limitations.

ASL sign learning (ASL-SL) task. In the ASL-SL task, participants had up to two trials to learn 24 ASL signs and their associated English word pairs (Figure 2). ASL signs were selected from the ASL-LEX database (Caselli, Sehyr, Cohen-Goldberg, & Emmorey, 2016) such that 1) their English glosses conformed to the specifications listed above (e.g., concrete and familiar), 2) the ASL signs were low in iconicity (the mapping of form and meaning), and 3) signs were visually distinct. 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.

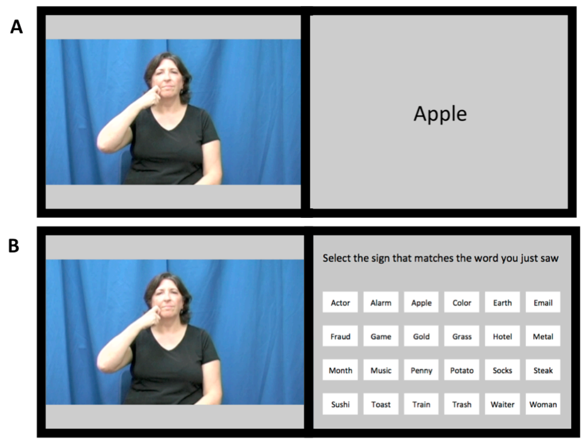


Figure 2. Depiction of the ASL-SL Task. Panel A depicts a study trial. Panel B depicts a test trial. In both cases, the sign is shown first and is immediately followed by either the response word in the study phase or the response screen in the test phase.

Pseudosign learning (PSL) and delayed pseudosign learning (DPSL) tasks. Like the ASL-SL task, the PSL is a paired-associate task, however, it differs from the ASL-SL task in a number of ways. First, pseudosigns are used instead of real signs. Using pseudosigns confers greater control over such variables as iconicity and sign complexity. Second, the model performing the sign varied between the study and test phase of a block (see Figure 3). This reduced the possibility that participants could rely on extraneous details (i.e., the model slouching in one video while sitting straight in all others) and placed greater focus on the linguistic features of the signs. A hearing native signer (the second author) performed all signs used during the study phase; test phase signs were reproductions of the study phase signs and were performed by a non-signer (the first author). Third, a dropout procedure was used in which once a participant correctly identified a sign, it no longer appeared in any future block (i.e., study or test trial). This was done to control for the positive effect that overlearning has on retention (Driskell, Willis, & Copper, 1992)—an important consideration for the Delayed Pseudosign Learning (DPSL) task.

The DPSL task consists of a single block of PSL test trials administered after four intervening tasks, approximately 30 min. after the PSL. As such, this task was intended to measure *retention* of lexical items, a construct that is substantially related to initial learning (Kyllonen & Tirre, 1988).

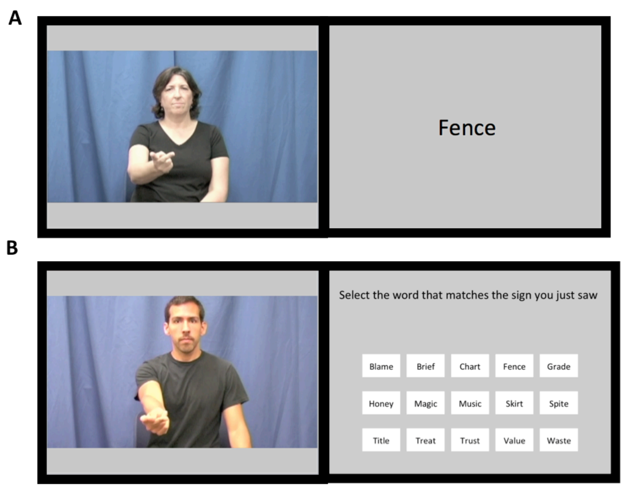


Figure 3. Depiction of the PSL task. Panel A shows a study trial. Panel B shows a test trial.

There were 15 PSL items and scores were calculated over a maximum of three trials for a possible score of 45. The DPSL, on the other hand, consisted of a single test block of 15 items, however, in order to avoid penalizing participants for pairs they had not learned and to further remove variance due to a participant's rate of learning, DPSL scores were calculated as a percentage of the number of pairs learned in the PSL over all PSL trials. Thus the denominator used to calculate the DPSL score for an individual who correctly responded to 10/15 PSL items was 10. In the final analysis, only one individual had a score above 100% on the DPSL—this score was adjusted to 100%.

Three-term sign learning (3TSL) task. The 3TSL is a complex associative learning task adapted from B. A. Williams and Pearlberg (2006) in which a stimulus is associated with three possible responses, contingent on a cue (see Figure 4). For example, during the study phase the stimulus pseudosign, *S*, may be associated with *tree*, *bone*, and *fork*, and each response word is associated with the cues *1, 2*, and *3*, respectively. During the first study trial, pseudosign *S* would be presented and immediately followed by instructions to press the 1-key. Once the button was pressed or after 2000ms had elapsed, the associated English word would be revealed and displayed for 2000ms. Next, the *same* pseudosign would be replayed, immediately followed by instructions to press the 2-key, and so on. During the test phase, a stimulus (e.g., *S*) and cue (e.g., *2*) would be presented followed by instructions to identify the associated English word (*bone* in this example). During the study phase, pseudosigns were presented randomly, however, all English words associated with a particular pseudosign were presented sequentially. During the test phase, pseudosign-cue combinations were presented randomly.

In the 3TSL, there were 6 pseudosigns, each with three associated words and cues and scores were calculated based on performance over a maximum of three blocks. A non-signer performed all pseudosigns and the same movie clips were used for the study and test phases. The maximum score was 54 (6 stimulus pseudosigns x 3 response words x 3 blocks).

Importantly, prior research has found that three-term associative learning is strongly correlated with paired associate learning (e.g., Kaufman et al., 2009), thus we include this measure as an indicator of sign learning.

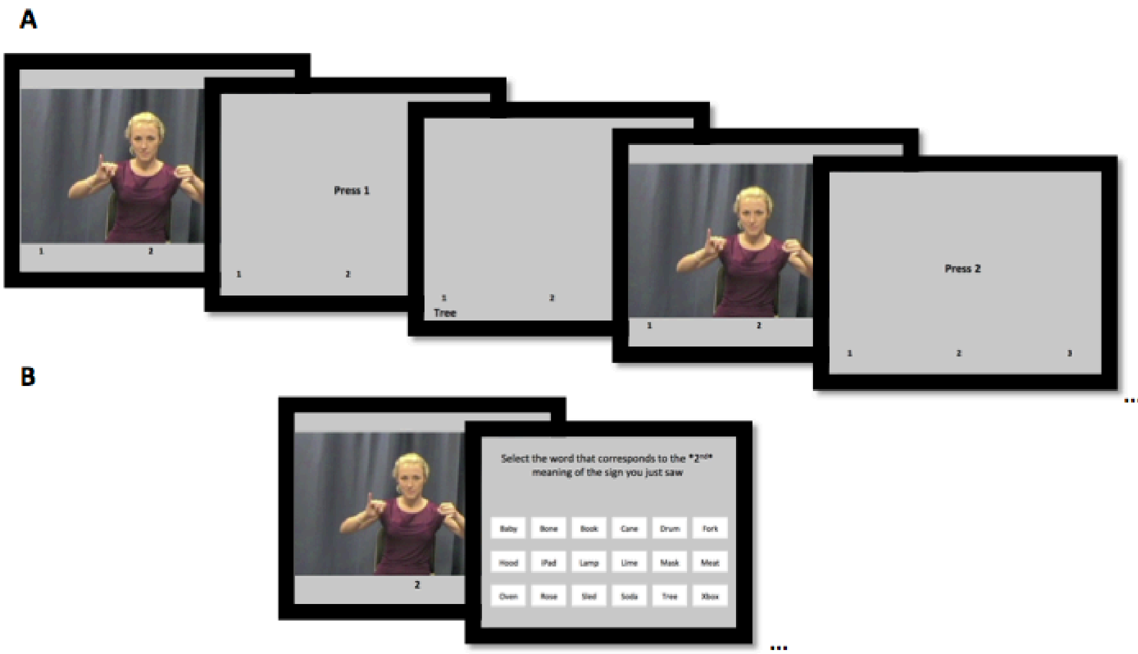


Figure 4. Depiction of the 3TSL. Panel A depicts part of a study trial. Panel B depicts part of a test trial.

Turkish word learning (TWL) task. Participants attempted to learn 15 Turkish-English word pairs over a maximum of three blocks. The Turkish words were spoken by a native Turkish speaker from Istanbul and presented over headphones; the same audio clips were used for both study and test trials. The maximum possible score was 45.

Pseudoword learning (PWL) and delayed pseudoword learning (DPWL) tasks. Like the PSL, the PWL employed a dropout procedure and two different people (in this case, two different female research assistants) produced the study and test items. All pseudowords were presented aurally over headphones. There were 15 pairs and participants' scores were calculated as the total correct over 3 blocks, for a maximum score of 45.

The DPWL learning consisted of a single test block of the PWL learning test items administered after four intervening tasks, approximately 40 min after the PWL. The maximum possible score was a percentage of the total number of items a participant had learned across all PWL trials.

Three-term word learning (3TWL) task. The 3TWL is similar to the 3TSL: six pseudowords (presented over headphones) were each associated with three English words and cues; the same audio clips were used during study and test trials; scores were calculated over a maximum of three trials for a total possible score of 54.

2.3.2 Phonological short-term memory

All PSTM tasks were either span tasks or discrimination (same-different) tasks. In a span task, a participant is presented with a set of items and is tasked with recalling the items in the order presented. Items were always selected from a limited pool of 9 to 12 items and the complete pool of items used in a task were always on display when participants responded. In order to reduce the role of WMC, effort was made to reduce within-task item similarity (see Oberauer, Farrell, Jarrold, & Lewandowsky, 2016), either acoustically (Baddeley, 1966; Conrad & Hull, 1964) or visually (Wilson & Emmorey, 1997), depending on the variant of PSTM the task was intended to assess. In order to maximize individual differences in performance, sets varied in length and a partial credit scoring procedure with unit weighting was used (for details, see Conway et al., 2005, pp. 775-777). In partial credit unit scoring, participants receive credit for each item of a set recalled in its correct serial position, however, the amount awarded is equal to one over the total number of items in the set.

In the discrimination tasks, participants judged whether a target item or sets of items were the same or different from a reproduction of either a single target item or an entire set of target items, depending on the task. Relative to the span tasks, discrimination tasks had more trials and items were drawn from larger pools (28 for the NSPT and NWRec and 16 for the ProSign) and, as such, it was difficult to limit within-task item similarity, though effort was made to limit within-*set* item similarity.

During the response portion of a task, the response screen appeared simultaneously with the reproduction and participants were to use the computer mouse to click on buttons (i.e., text boxes) with the words “same” or “different” inscribed. The *same* button always appeared on the right hand side and the *different* button on the left. Participants were able to make their judgments as soon as they recognized a difference and were warned that they should not make a *same* judgment until the entire reproduction was presented. Finally, it should be noted that during *same* trials, the exact same stimuli were used for both the target and reproduction.

All but one PSTM task, the letter span task, used pseudosigns or pseudowords. As with the lexical learning tasks, pseudosigns did not necessarily follow the phonotactics of any particular language, while pseudowords obeyed English phonotactics. To maintain consistency with the field (e.g., Gathercole, 2006; Mann, Marshall, Mason, & Morgan, 2010), we have labeled some tasks utilizing pseudowords or pseudosigns as *nonword* or *nonsign* tasks, respectively (e.g., nonword recognition).

Nonsign paired task (NSPT). The NSPT used here is a shortened version of the original NSPT used by Martinez and Singleton (2018). In the NSPT, participants must judge whether target pseudosigns differ from their reproductions (see Figure 5). Martinez and Singleton (2018) observed moderate to strong correlations between it, a sign learning task, two visuospatial STM tasks, and another putative task of signed-PSTM, the Nonsign Repetition Task (Mann et al., 2010), providing evidence of the validity of the NSPT as a measure of signed-PSTM.

The NSPT begins with a 164 second instructional video. The video introduces participants to the task and three sign phonology parameters: handshape, orientation, and movement. It was explained that 50% of pairs would be faithful reproductions and should be classified as "same" while the other 50% of reproductions would differ on *one* of the forenamed parameters.

Figure 5. Depiction of the NSPT. In the NSPT, there were 28 target signs each with two reproductions, produced by different individuals.

Next, participants were told that there would be two blocks: the same target pseudosigns would be used across both blocks, however, two different individuals would perform the reproductions, and pairs would be presented in a different order from one block to the other. Next, participants completed two blocks of three practice trials with automated feedback. The automated feedback either informed participants that they were correct, or if they were wrong, displayed a screen with a brief text description of the error as well as static images of the target and reproduction with differences highlighted. The critical trials followed. The two critical blocks each consisted of 28 items for a maximum score of 56 points—feedback was never provided.

Probed sign (ProSign) task. In the ProSign task, participants viewed sets of pseudosigns followed by a cue (500ms) and a probe; participants were to indicate whether the probed pseudosign was in the set just viewed or if it was different (see Figure 6). If a probe was different, then, as in the NSPT, it differed from one of the other items in the set by one parameter: handshape, movement, or orientation; if it was the same, then the pseudosign (and video clip) was exactly the same as a pseudosign in the set.

Figure 6. Depiction of a ProSign item. The probe differs from the second pseudosign in the set.

There were 40 critical trials with 10 trials each at set lengths three through six. Half of all trials were *different* trials with six differing from the target in handshape, seven in orientation, and seven in movement. In an attempt to maximize individual differences in performance, the majority of the forty trials assessed memory for pseudosigns between the first and last pseudosigns in a set, as recall of items in the first and last positions of a set tend to be at or near ceiling (e.g., Jones, Farrand, Stuart, & Morris, 1995; Unsworth & Engle, 2007; Ward, Avons, & Melling, 2005; Wu & Coulson, 2014). In all, eight (20% of all trials) assessed memory for the first item, eight (20%) assessed memory for the final item, and 24 (60%) assessed memory for pseudosigns in between.

Sign configuration task (SignCon). The SignCon is a dual-task in which participants completed two span tasks: a letter span (described in detail in section 2.3.2.4, below) and a pseudosign span (see Figure 7). The critical portion of the SignCon is the pseudosign span portion, however, in this task, within-task item similarity was low, potentially enabling participants to effectively use a verbal mediation strategy (e.g., labeling) and articulatory rehearsal—the letter span portion of the task was meant to disrupt use of verbal mediation strategies. Moreover, participants were explicitly told not to attempt to label any of the pseudosigns. To check for compliance, 40% of trials assessed only the letter span portion and 60% assessed only the pseudosign portion.

Every trial of the SignCon began with participants viewing sets of letters followed by one to four pseudosigns. The length of the set of letters was always equal to the participant's letter span—the maximum number of letters that could be perfectly recalled in serial order for three trials—calculated from the participant's performance on the LetterSpan task completed earlier in the session minus one. For example, if a participant was able to perfectly recall three sets of set size seven but only two sets of set size eight and one set of set size nine, then in this task (the SignCon), the participant would always see six letters during the letter set portion. In this way, a participant's ability to rehearse should be prevented and the memory load should be *functionally* equivalent across participants.

After the set of letters were presented for a length of time equal to 500 ms per letter, participants viewed one to four pseudosigns. Nine pseudosigns were repurposed from the nonsign repetition task developed by Mann et al. (2010) and used as stimuli in this task. Next, participants were tested on either the letters or the pseudosigns. If they were tested on the letters, then the test trial proceeded just like the letter span task described below. If the participant were tested on the pseudosigns, then they viewed a 3 x 3 matrix with still images representing a key configuration (either an initial or final position) in each of the nine pseudosigns used in this task, one configuration per sign. The participant's task was to click on the images in the same order as the pseudosigns were originally presented. To reiterate, participants initially saw video clips of pseudosigns, however, at test, they only saw still images. Twelve of the 20 trials were pseudosign trials and there were three trials at each set length. Using the partial credit unit scoring procedure described above, the maximum possible score was 12.

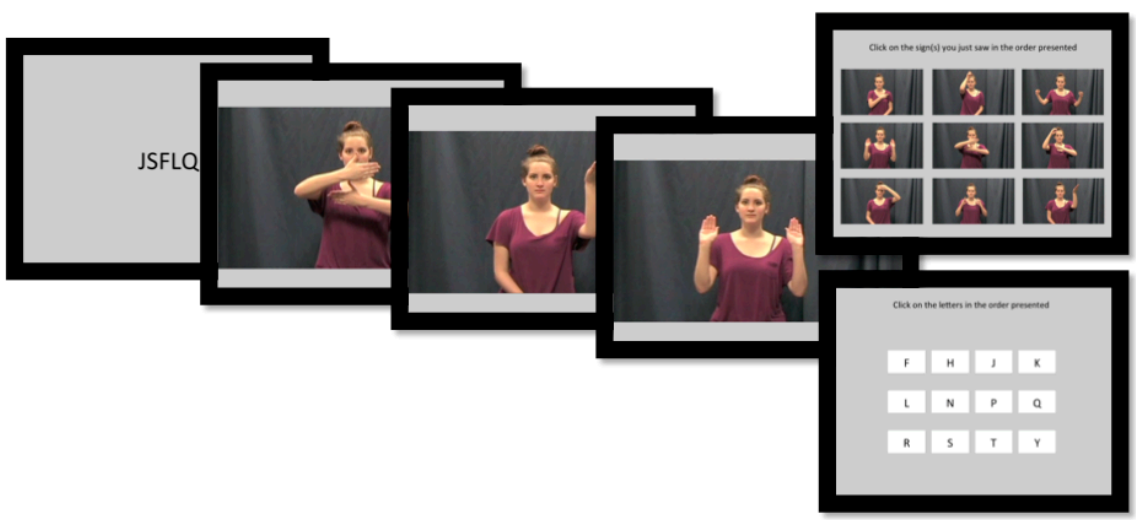


Figure 7. Depiction of the SignCon. Participants were either tested on the number of pseudosigns (60% of trials) or letters (40% of trials) they could recall in order.

Letter Span (LetSpan). In this task, participants attempted to recall four to nine letters in serial order. The pool of items consisted of 12 letters: F, H, J, K, L, N, P, Q, R, S, T, Y. The entire set of letters was presented on screen for a length of time equal to the set length times 500 ms (e.g., a set of 6 letters was presented for 3000 ms). There were three trials at each set length for a total of 18 sets. Using partial credit unit scoring, the maximum was 18 points.

Nonword recognition (NWRec) task. The NWRec task was adapted from Gathercole, Pickering, Hall, and Peaker (2001) and similar tasks have been used by others (e.g., K. I. Martin & Ellis, 2012; O'Brien et al., 2006). In the NWRec task, participants discriminate between two sequences of pseudowords presented aurally via headphones. If the sequences were different, then two neighboring pseudowords were transposed; if they were the same, then the exact same sequence of pseudowords was presented again. There were 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. Moreover, 1/3 of different trials contained a transposition of the first and second pseudowords, 1/3 were transpositions of the final and penultimate pseudowords, and the remaining were transpositions of pseudowords in between. Pseudowords were drawn from a pool of 28 items and were selected from Gathercole et al. (2001). The maximum score was 36.

Nonword span (NWSpan) task. In the NWSpan, participants heard a set of monosyllabic pseudowords over headphones and attempted to recall the pseudowords in the order presented by clicking on a response screen with the entire pool of words displayed. The pool of pseudowords 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.

2.3.3 Working memory capacity

The WMC tasks used here were all shortened versions of complex span tasks (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 serial order. As with the PSTM span tasks described above, WMC tasks were scored using partial credit unit scoring.

Operation span (OSpan). In the OSpan, participants were presented with a series of letters with math equations interleaved between letter presentations. Participants were to try to remember the letters in the order they were presented. There was one set at each set length of three through seven for a total of five trials. Using partial credit unit scoring, the maximum score was 5.

Symmetry span (SymSpan). In this task, the primary (memory) task was to remember the sequence of locations of a red square in a 4x4 matrix. The secondary task was to judge whether a figure composed of shaded squares in an 8x8 matrix was symmetrical along the vertical axis. The number of locations to be remembered varied from two to five per trial, for a total of four trials. Using partial credit unit scoring, the maximum score was 4.

Rotation span (RoSpan). The primary task in the RoSpan was to remember a sequence of arrows varying in size and direction. The Secondary task was to judge whether a rotated letter, when mentally rotated to its upright position, is displayed correctly or is mirrored. The number of arrows to be remembered varied between two and five, for a total of four trials. Using partial credit unit scoring, the maximum score was 4.

2.3.4 Intelligence

The following holds true for all intelligence tests used in this study: 1) test format was multiple-choice, 2) there was a time limit, 3) questions were generally ordered from easiest to hardest, and 4) participants were told that they should work quickly but accurately and, when necessary, guess.

Test of reading comprehension (Reading). Participants had up to 20 min to read 5 passages (varying in length from 112 words to 739 words) and answer 17 questions. All passages and their corresponding questions were drawn from released SAT and GRE tests and were selected to provide a range in item difficulty. The test was administered in paper format and participants were encouraged to use whatever strategies they normally use *except* answering questions out of order. The maximum score was 17.

Information (Info) test. The Info test consisted of two parts. In part 1, participants had up to 7 min to answer 40 general knowledge questions from the Information subscale of the Multidimensional Aptitude Battery II (Jackson, 1998). In part 2, participants were allowed 2 min to answer an additional 11 questions—these questions were developed in-house and were added to broaden the domains of knowledge assessed. Performance across both parts were summed to form one score, thus the maximum score was 51.

Extended range vocabulary (Vocab) test. In the vocab test (Ekstrom, French, Harman, & Dermen, 1976), participants are presented with a word and attempt to match it with one of five words that is closest in meaning. There were two parts, each with 24 items, and a time limit of 6 min. The maximum was 48.

Grammar and usage (Gram). The Gram test consisted of 21 "improving sentences" items selected from sections 5 and 10 of official SAT practice tests released between 2004 and 2013. Each item consisted of a sentence with a portion underlined; the participant was to select the answer choice that best rephrased the underlined portion or, if the original phrasing was the best choice, select the first answer choice, which always repeated the original phrasing. Participants had up to 10 min to complete the test. The maximum score was 21.

Raven's advanced progressive matrices, set II (Ravens). In Ravens, Participants were presented with 18 3x3 matrices with all but the lower right cell of each matrix containing figures. The figures in each matrix are arranged according to a rule (see Carpenter, Just, & Shell, 1990) and participants were to infer the rule and select which of eight figures presented below the matrix best completed the pattern. The 18 items used in this task were the odd items from set II of Raven's Advanced Progressive Matrices (Raven, Raven, & Court, 1998). Participants had 10 min to complete the task and the maximum score was 18.

Letter sets (LetSets). In the LetSets task (Ekstrom et al., 1976), participants were presented with five sets of letters, with each set consisting of four letters. The participant was to identify the set of letters that did not obey the same rule as the others. There were 30 problems and participants were given up to 7 min to complete the task. The maximum score was 30.

Number series (NumSeries). In NumSeries (Thurstone, 1938), participants were presented with a series of numbers that obeyed a particular rule. The participant's task was to complete the series by selecting the one answer choice (out of five) that would continue the series. Participants had up to 5 minutes to complete 15 items. The maximum score was 15.

Spatial learning ability test (SLAT). The SLAT used here is an adaptation of the SLAT described by Embretson (1992). In this version of the SLAT (see Figure 8), participants were presented with a representation of an unfolded cube. The six faces of the target contained simple shapes such as arrows and pentagons. The participant was to choose which of four cubes matched the target by mentally rotating and folding the target to compare with the four answer choices. Tasks such as these tend to correlate moderately to strongly with putative measures of fluid intelligence (Lohman, 1996; Marshalek, Lohman, & Snow, 1983; Varriale, van der Molen, & De Pascalis, 2018) and so it is being used here an indicator of that construct. There were 20 items and participants had up to 15 min to complete them. Note, the original SLAT is a dynamic tests consisting of a pretest, an intervention, and a posttest (for further details, refer to Embretson, 1992). In this version of the SLAT, there is only a single test and no intervention.

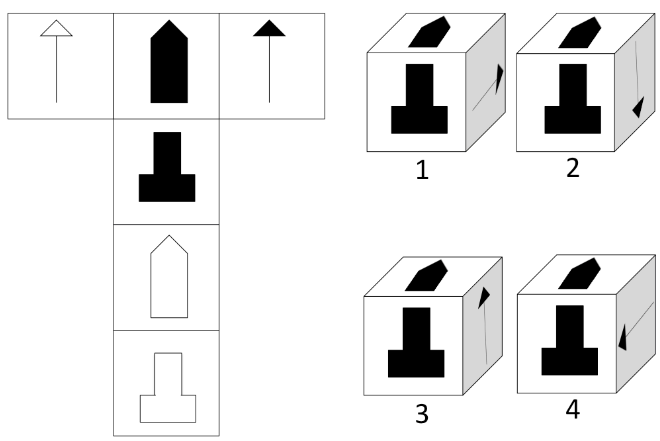


Figure 8. Depiction of a SLAT item.

3 Statistical Analyses

3.1 Data screening

There were 61 missing values and 11 values were classified as outliers because they were 3.5 standard deviations from the mean. Missing values were deemed *missing at random* and so the expectation-maximization (EM) algorithm was used to impute those values (Little & Rubin, 2014; Rubin, 1976). Outliers were treated by imputing the values using the EM algorithm and, when imputed values were still classified as outliers, by replacing the values with a score equal to 3.5 standard deviations from the mean. Two scores (both for the OSpan) were still below the cutoff after imputation and so they were replaced with a value equal to a z-score of 3.5.

Multivariate normality was assessed using a normalized version of Mardia's coefficient (Mardia, 1970). Bentler (2001) suggests that a value above five is suggestive of non-normality. As will be observed below, all values of Mardia's normalized estimate were below five and so no actions were taken to correct for non-normality.

3.2 Statistical procedure

Structural equation models were created and analyzed using EQS (Bentler, 2001). Because the data appeared to be normally distributed, model parameters were estimated using maximum likelihood, a method that yields the smallest errors when the data are normal (Ullman, 2006).

Model fit was assessed using several statistics recommended by Kline (2016): model chi-square (with associated degrees of freedom and p-value), Comparative Fit Index (CFI), Standardized Root Mean Square Residual (SRMR), and the Root Mean Square Error of

Approximation (RMSEA). The chi-square test, SRMSR, and RMSEA are "badness-of-fit" tests—lower values indicate good fit. The CFI, on the other hand, is a *goodness-of-fit* test with values closer to 1 indicating good fit.

Finally, estimated parameters (e.g., path coefficients) were assessed using significance tests; a value of 0.05 was considered significant.

4 Results

4.1 Observed Variable Analyses: Descriptive Statistics, Reliability, and Correlations

Descriptive statistics and internal consistency coefficients (Cronbach's alpha) are provided in Table 2. Tasks were generally sufficiently difficult for individual differences research and the data were approximately normally distributed. Nearly all internal consistency coefficients were at or near .80, suggesting acceptable reliability (cf., Draheim, Mashburn, Martin, & Engle, 2019). Only four tasks, the NSPT, ProSign, SymSpan, and RoSpan had coefficients below .70, however, these tasks tended to show strong correlations with tasks measuring the same or similar constructs, providing evidence for their validity.

Bivariate correlations are provided in Table 3. All tasks were significantly correlated to each other at $p < .01$. More importantly, the correlation matrix shows evidence of discriminant and convergent validity.

Table 2. Descriptive statistics and reliabilities

Task	Mean (SD)	Range	Skew	Kurtosis	α
1. ASL-SL	.71 (.22)	.08-1.00	-.96	.27	.93
2. PSL	.71 (.20)	.07-1.00	-.78	-.02	.87
3. DPSL	.70 (.22)	.00-1.00	-.57	.01	.82
4. 3TSL	.63 (.27)	.04-1.00	-.53	-.79	.95
5. TWL	.54 (.24)	.02-1.00	-.24	-.88	.91
6. PWL	.64 (.25)	.07-1.00	-.46	-1.03	.91
7. DPWL	.58 (.25)	.00-1.00	-.40	-.42	.83
8. 3TWL	.49 (.33)	.00-1.00	.03	-1.53	.98
9. NSPT	.80 (.07)	.57-.98	-.53	.51	.66
10. ProSign	.67 (.11)	.40-.90	-.25	-.31	.57
11. SignCon	.61 (.17)	.06-.92	-.67	.50	.75
12. LetSpan	.81 (.09)	.49-1.00	-.65	.45	.78
13. NWRec	.80 (.12)	.42-1.00	-.61	-.12	.75
14. NWSpan	.70 (.11)	.40-.99	-.15	-.04	.76
15. OSpan	.82 (.18)	.18-1.00	-1.29	1.63	.74
16. SymSpan	.74 (.23)	.00-1.00	-.99	.87	.67
17. RoSpan	.61 (.22)	.00-1.00	-.75	.18	.64
18. Reading	.50 (.22)	.00-1.00	-.00	-.55	.79
19. Info	.60 (.13)	.14-.90	-.85	.98	.83
20. Vocab	.53 (.15)	.15-.85	-.03	-.36	.84
21. Gram	.45 (.20)	.00-.95	.15	-.52	.78
22. NumSeries	.67 (.20)	.13-1.00	-.44	-.60	.77
23. LetSets	.57 (.16)	.17-.90	-.39	-.41	.86
24. Ravens	.57 (.21)	.06-1.00	-.37	-.41	.80
25. SLAT	.54 (.25)	.00-.95	.11	-1.17	.86

Note: Scores were converted to percentages. ASL-SL = ASL sign learning; PSL = pseudosign learning; 3TSL = three-term sign learning; DPSL = delayed pseudosign learning; LetSpan = letter span; NWRec = Nonword Recognition; NWSpan = Nonword Span; TWL = Turkish word learning; PWL = pseudoword learning; 3TWL = three-term word learning; DPWL = delayed pseudoword learning; NSPT = nonsign paired task; ProSign = Probed Sign recognition task; SignCon = sign configuration task; 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.

Table 3. Zero-order correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1. ASL-SL																								
2. PSL	.70																							
3. DPSL	.65	.63																						
4. 3TSL	.73	.70	.61																					
5. TWL	.62	.66	.51	.72																				
6. PWL	.58	.70	.52	.67	.78																			
7. DPWL	.52	.55	.53	.58	.64	.66																		
8. 3TWL	.57	.65	.50	.76	.77	.75	.64																	
9. NSPT	.47	.41	.38	.53	.42	.46	.37	.48																
10. ProSign	.49	.47	.44	.53	.50	.48	.44	.47	.62															
11. SignCon	.56	.51	.49	.59	.52	.54	.53	.57	.54	.58														
12. LetSpan	.23	.31	.23	.37	.42	.37	.29	.39	.36	.34	.27													
13. NWRec	.44	.48	.39	.54	.60	.54	.45	.55	.42	.48	.50	.55												
14. NWSpan	.39	.42	.39	.50	.60	.52	.46	.55	.47	.47	.52	.58	.63											
15. Reading	.41	.35	.34	.43	.44	.39	.39	.38	.51	.41	.34	.36	.45	.39										
16. Info	.41	.33	.31	.42	.40	.38	.31	.34	.45	.30	.29	.31	.32	.38	.55									
17. Vocab	.48	.42	.33	.47	.55	.50	.42	.47	.51	.43	.38	.33	.40	.43	.61	.66								
18. Gram	.42	.33	.28	.44	.50	.47	.38	.45	.45	.40	.35	.44	.50	.43	.58	.57	.62							
19. NumSeries	.42	.41	.33	.53	.50	.47	.42	.46	.43	.38	.38	.40	.43	.46	.46	.55	.49	.50						
20. LetSets	.48	.44	.37	.53	.52	.56	.44	.46	.43	.41	.46	.41	.46	.42	.40	.43	.44	.55	.66					
21. Ravens	.57	.51	.42	.62	.56	.54	.51	.55	.53	.48	.53	.26	.43	.46	.43	.46	.50	.50	.61	.57				
22. SLAT	.47	.45	.43	.53	.46	.44	.43	.47	.48	.45	.48	.27	.40	.42	.43	.47	.50	.45	.59	.55	.64			
23. OSpan	.32	.37	.34	.38	.39	.32	.25	.36	.42	.34	.34	.42	.43	.50	.34	.30	.28	.36	.47	.44	.37	.37		
24. SymSpan	.29	.35	.30	.38	.38	.33	.27	.38	.44	.37	.39	.33	.35	.36	.31	.29	.28	.40	.49	.41	.49	.46	.43	
25. RoSpan	.24	.27	.25	.38	.32	.25	.19	.36	.33	.31	.37	.27	.28	.32	.28	.31	.20	.29	.44	.33	.41	.33	.40	.54

Note: See Table 2 for key.

4.2 Latent Variable Analyses

To assess the validity of the tasks used here, the manifest variables in this study were grouped into factors and structural equation modeling was used to model the relationships amongst the latent variables. Model fit was good (Table 4, *Corr* model), however, inspection of the results of the Lagrange Multiplier test offered by EQS (Bentler, 2001) revealed that two pairs of tasks shared a significant proportion of variance: 1) 3TSL and 3TWL and 2) LetSets and NumSeries. These pairs of tasks are very similar in format and so it was deemed appropriate to account for this *method variance* by correlating their residuals. As can be seen (Table 4, *Corr-LM* model), these corrections resulted in a significantly better fitting model, $\Delta\chi^2$ (2) = 29.587, p < .001, and so they were retained.

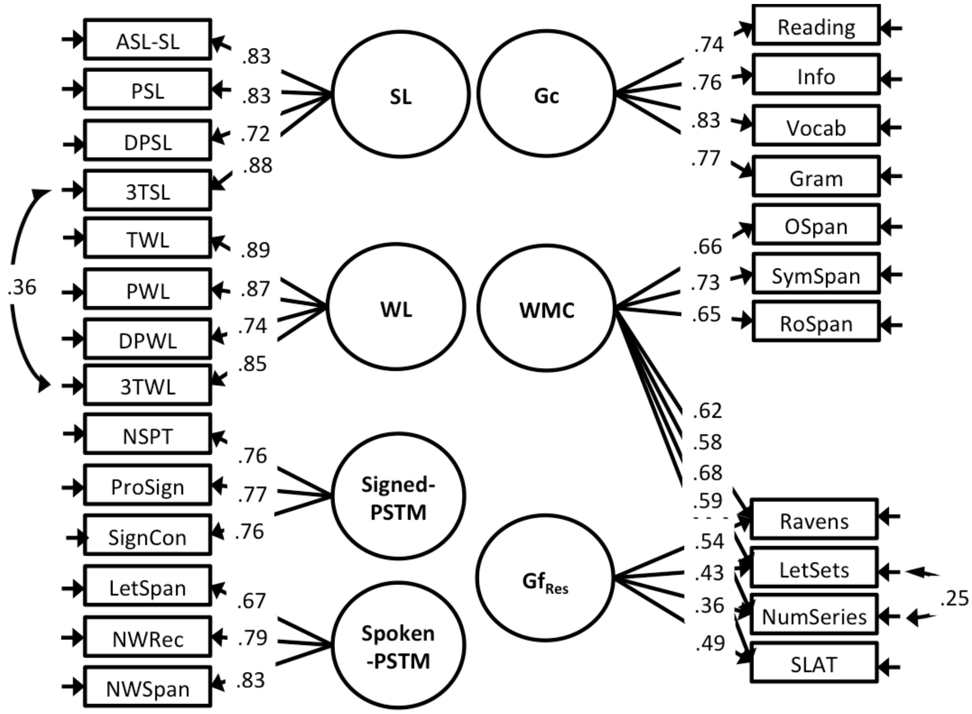
Table 3. Correlated factors model fit statistics

Model	Mardia's	X^2	df	CFI	SRMR	RMSEA (95% CI)
Corr	3.16	426.03	251	.953	.044	.054 (.045, .063)
Corr-LM	3.16	396.44	249	.961	.044	.050 (.041, .059)

Note: Corr = correlated factors model; Corr-LM = correlated factors model with corrections suggested by the Legrange Multiplier test.

Figure 9A, illustrates all latent variables (circles) and their corresponding observed variables (rectangles) as entered into the Corr-LM model, along with estimated path coefficients. For clarity, correlations are shown separately in Table 5.

A



B

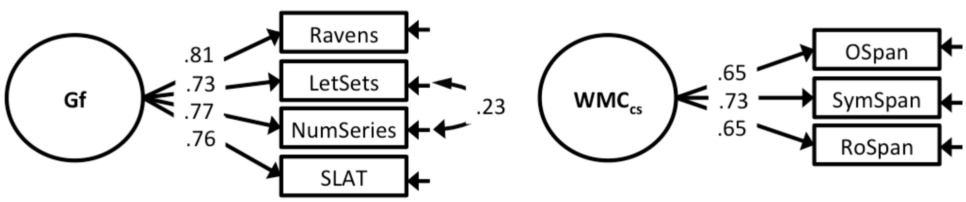


Figure 9. Latent variable and their indicators. Panel A shows the estimated path coefficients derived from analyzing Model Corr-LM. For reference, panel B shows the unresidualized fluid intelligence factor (Gf) and its indicators as well as a WMC factor derived from the variance of complex span tasks only (WMC_{CS}).

Table 4. Latent variable correlations

	SL	WL	Signed -PSTM	Spoken -PSTM	Gc	Gf_{Res}
1. SL						
2. WL	.88					
3. Signed-PSTM	.79	.74				
4. Spoken-PSTM	.65	.76	.75			
5. Gc	.62	.66	.68	.65		
6. Gf_{Res}	.54	.50	.40	.22	.54	
7. WMC	.57	.56	.69	.69	.57	---

Note: SL = sign learning; WL = word learning; PSTM = phonological short-term memory; Gc = crystallized intelligence; Gf_{Res} = residualized fluid intelligence factor with variance accounted for by WMC partialled out; WMC = working memory capacity.

Table 5. Correlations with fluid intelligence and WMCcs

	SL	WL	Signed -PSTM	Spoken -PSTM	Gc	Gf
Gf	.77	.75	.78	.68	.79	
WMC_{CS}	.57	.56	.71	.68	.56	.80

Note: Note: SL = sign learning; WL = word learning; PSTM = phonological short-term memory; Gc = crystallized intelligence; Gf = fluid intelligence; WMC_{Cs} = latent variable constructed from the variance due to complex span tasks only.

There are several things to note. First, following recent research (J. D. Martin et al., 2019), the fluid intelligence factor was residualized by partialling out the variance accounted for by WMC. This was done so that the residualized fluid intelligence factor would primarily represent individual differences in the ability to disengage from information while the WMC factor would represent a domain-general ability to maintain information. WMC and fluid intelligence also tend to be strongly correlated (Ackerman, Beier, & Boyle, 2005; Engle et al., 1999; Kane et al., 2004; Kyllonen & Christal, 1990)—and, in fact were here as well (see Table 6)—which can result in multicollinearity. Second, the fact that the Corr-LM model fits the data well and nearly all of the path coefficients between observed and latent variables were strong provides evidence of the validity of these tasks as measures of their intended constructs; only three task loadings were below .50, however, this is an outcome of the variance due to these tasks being split between the residualized fluid intelligence and WMC factors. Third, the correlations amongst the latent variables and in particular those concerning the WMC and residualized fluid intelligence factors speak to the appropriateness of modeling WMC and fluid intelligence as was done here and elsewhere (viz., J. D. Martin et al., 2019). Specifically, the WMC factor is most strongly correlated with two other memory factors, Signed- and Spoken-PSTM, while the residualized fluid intelligence factor correlates strongly with those factors that involve complex cognition. As further evidence, it should be noted that neither the complex span task loadings nor correlations with other latent variables changed substantially from what was observed when a WMC factor was constructed with only complex span tasks loading onto it (compare WMC in Figure 9A with WMC_{cs} in Figure 9B and Tables 6 and 7, respectively). Finally, it should be noted that the correlation between the sign learning (SL) and word learning (WL) factors was very strong (.88) but not perfect, suggesting that these latent variables are at least somewhat distinguishable (how to *best* model performance on the lexical learning tasks will be explored below).

4.2.1 Predicting sign and word learning

In this analysis, sign learning and word learning were set as outcome variables and the other latent variables were entered as predictors in a step-wise fashion. The first model was intended to assess the contribution of intelligence, indicated by fluid intelligence (unresidualized) and crystallized intelligence. Model fit was good (Table 7, Model 1). As can be seen in Figure 10, fluid intelligence significantly predicted both sign learning and word learning (indicated by solid lines) but crystallized intelligence did not make a significant contribution above and beyond fluid intelligence. Together, the predictors accounted for 60% of sign learning variance and 57% percent of word learning variance. Moreover, the disturbance terms were significant, indicating that a significant proportion of variance was left unaccounted for in both sign and word learning. Additionally, the disturbance terms were significantly correlated (.72; indicated by curved arrows).

Table 6. Fit statistics for predictive models

Model	Mardia's	X^2	*df*	CFI	SRMR	RMSEA (95% CI)
1	2.58	144.69	96	.98	.035	.046 (.030, .061)
2	3.65	208.41	137	.98	.039	.047 (.034, .059)
3	3.50	288.43	189	.97	.042	.047 (.036, .058)
4	3.16	396.44	249	.96	.044	.050 (.041, .059)

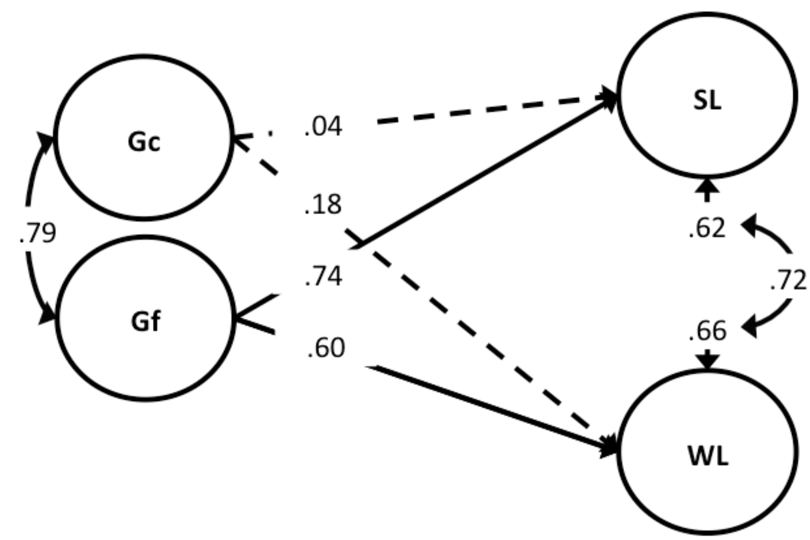


Figure 10. Model 1—the effect of intelligence

Next, WMC was added to the model and fluid intelligence was residualized (see Figure 11). Model fit was good (Table 7, Model 2), however, this model is largely a restructured version of Model 1; that is, the variance in sign learning and word learning explained by WMC is simply a portion of that which was already accounted for by the fluid intelligence variable in Model 1. This is supported by the fact that the proportions of variance accounted for in Model 1 and 2 are nearly identical, with Model 2 explaining 60% of the variance in sign learning and 58% in word learning. Still, by partitioning the variance in this way, we can see that those processes that are unique to fluid intelligence tasks are predictive of sign learning and word learning.

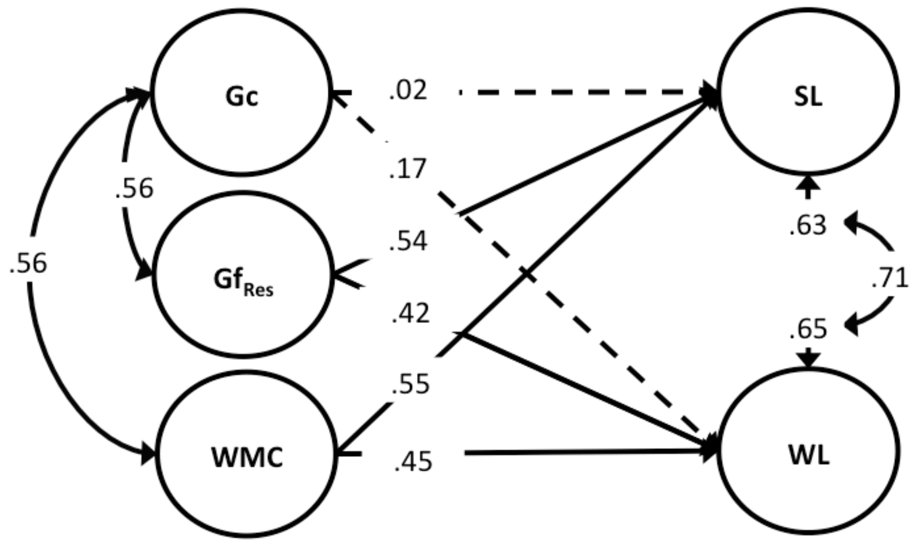


Figure 11. Model 2—accounting for WMC

In Model 3, Spoken-PSTM was added (see Figure 12). Model fit was good (Table 7) and the proportion of sign learning and word learning variance accounted for increased to 66% and 72%, respectively, while the correlation between the disturbance terms dropped to .67. Importantly, the inclusion of the Spoken-PSTM factor resulted in the path between WMC and word learning becoming insignificant. This suggests that, in relation to word learning, Spoken-PSTM assesses very similar processes as WMC, however, Spoken-PSTM assesses other relevant processes above and beyond those assessed by WMC.

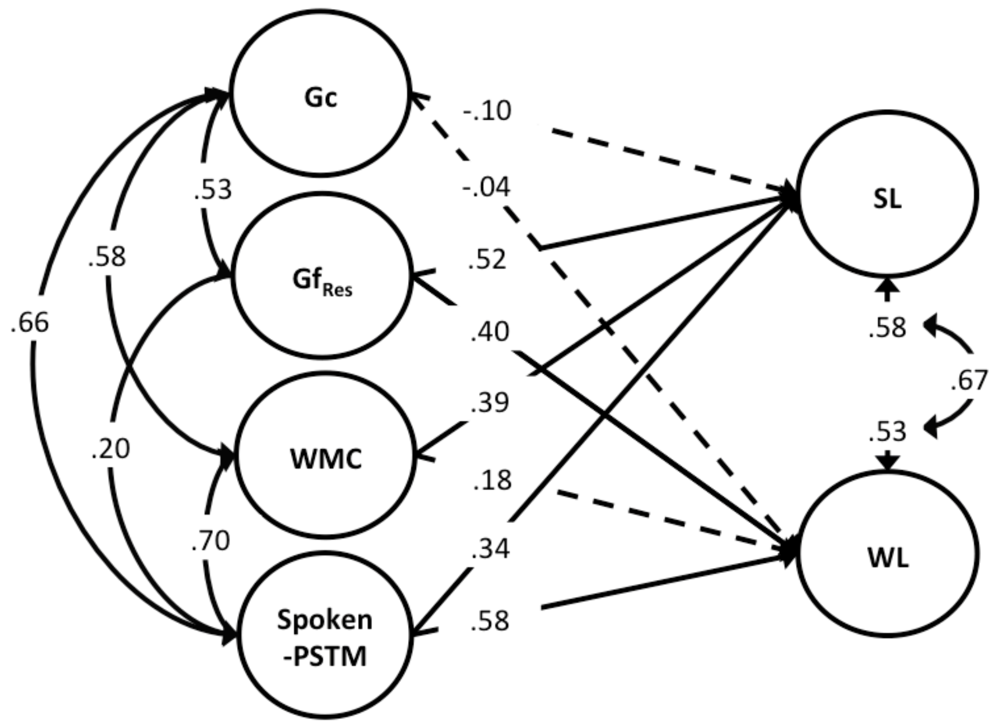


Figure 12. Model 3—Accounting for spoken-PSTM

In Model 4, Signed-PSTM was added. Model fit was good (Table 7); the proportion of sign learning and word learning variance accounted for were 71% and 72%, respectively; and the correlation between the disturbance terms was .67. Here, adding Signed-PSTM resulted in WMC and Spoken-PSTM no longer being significantly predictive of sign learning. Ultimately, it was only the residualized fluid intelligence factor and Signed-PSTM that significantly predicted sign learning while the residualized fluid intelligence factor and Spoken-PSTM were the only significant predictors of word learning (see Figure 13).

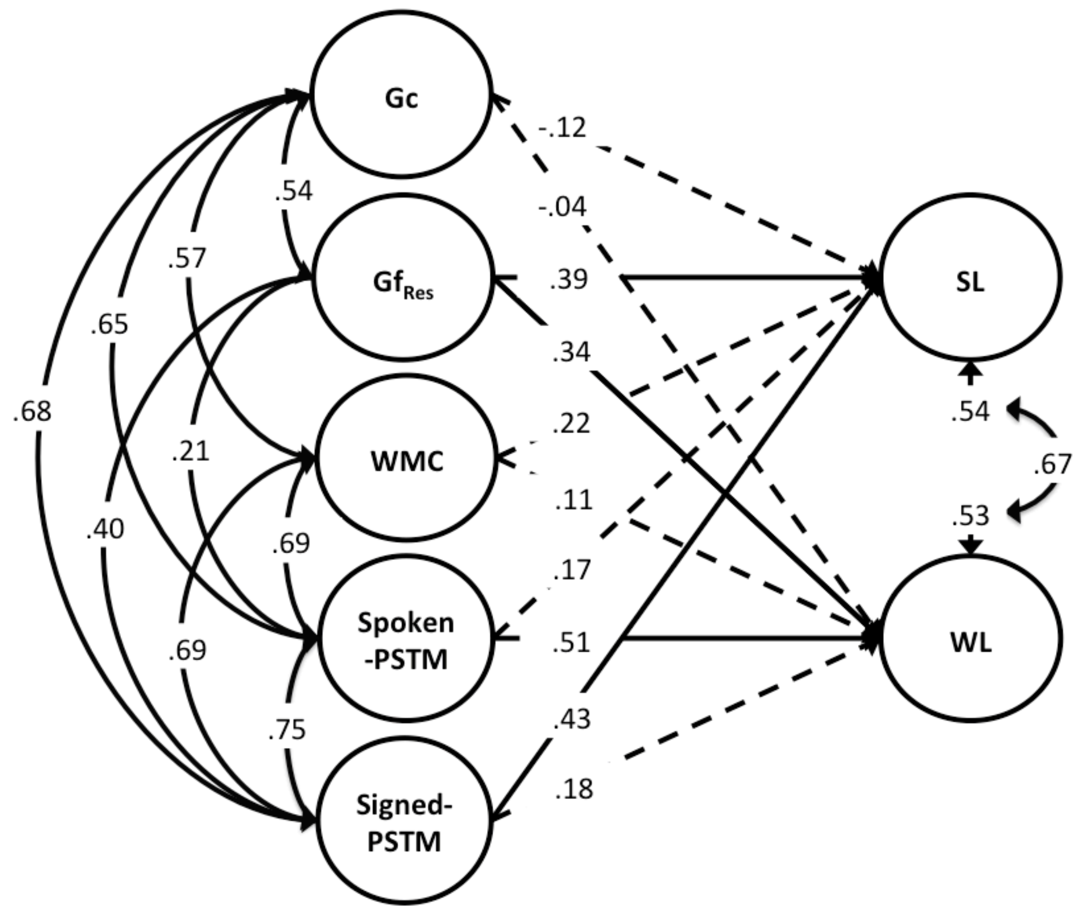


Figure 13. Model 4—Accounting for signed-PSTM

4.2.2 Modeling the relationship between sign and word learning

Next, structural equation modeling was used to directly explore the relationship between sign learning and word learning. As was observed in Table 6, the correlation between the sign learning and word learning factors was quite strong, suggesting that a general lexical learning factor underlies performance on all lexical learning tasks used in this study. To investigate this possibility, a one-factor model was designated by loading all lexical learning tasks onto a single factor; next, this model was compared with a baseline model consisting of separate but correlated sign and word learning factors (see Figures 14A and 14B).

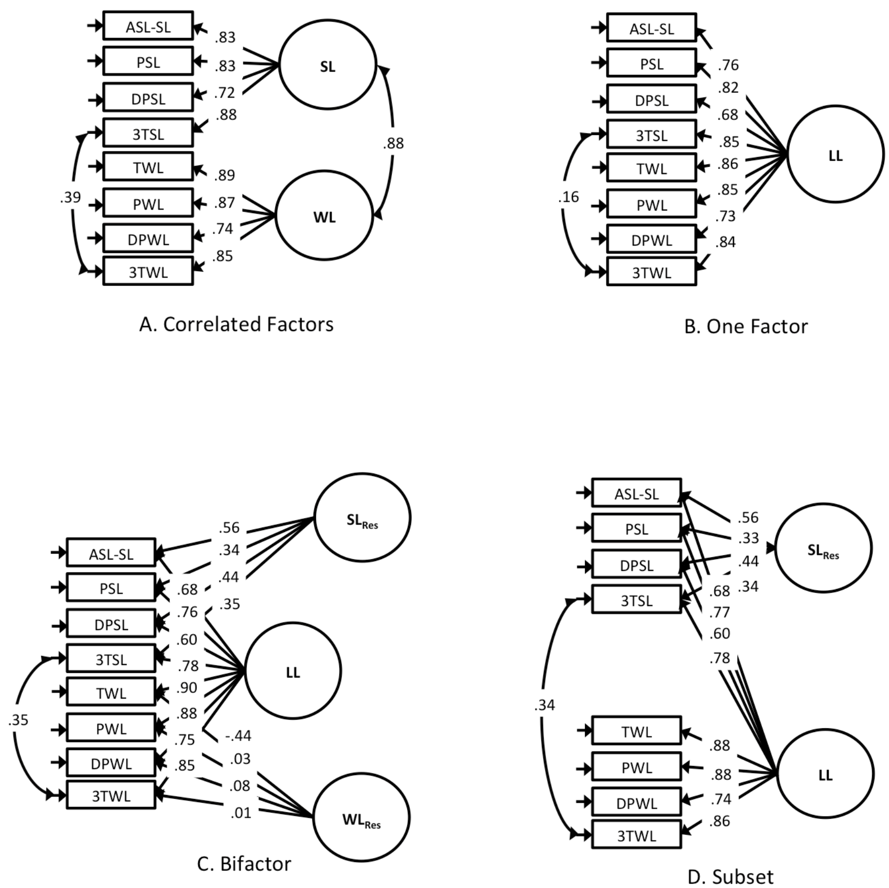


Figure 14. Exploratory models investigating the relationship between word learning and sign learning. Note: SL_{Res} = residualized sign learning factor; WL_{Res} = residualized word learning factor; LL = lexical learning factor

Mardia's normalized estimate for this and all subsequent models within this series of analyses was 4.31, indicating multivariate normality. As can be seen in Table 8, the one factor (OF) model had poor fit while the correlated factors model (CF) had adequate fit. These results suggest that sign learning and word learning are not completely independent factors nor are they fully determined by a single general lexical factor. Next, to test the possibility that a single factor contributed to individual differences on all lexical learning tasks but that specific factors *also* account for variance in performance, a bifactor model was designated by loading all tasks onto a single lexical learning factor and loading sign learning and word learning tasks onto residualized (or s*pecific*) factors (Figure 14C).

Table 8. Fit statistics for exploratory models

Model	X^2	*df*	CFI	SRMR	RMSEA (90% CI)	AIC
CF	39.06	18	.99	.027	.071 (.040, .101)	3.06
OF	117.37	19	.93	.047	.148 (.123, .174)	79.37
BF	20.23	11	.99	.018	.060 (.011, .100)	-1.77
SS	21.57	15	1.0	.019	.043 (.000, .081)	-8.43

Note: CF = correlated factors; OF = one-factor; BF = bifactor; SS = subset; AIC = Akaike Information Criterion

The bifactor (BF) model demonstrated good fit and fit the data better than the correlated factors model, as indicated by the lower Akaike Information Criterion (AIC) value (see Table 8). Importantly, however, an inspection of the path coefficients revealed a misspecification in the model: the path coefficients between the word learning tasks and the residualized word learning factor were insignificant or, in the case of the Turkish word learning task, moderately large and *negative.* What this suggested was that the word learning specific factor was redundant.

In the final model assessed, all lexical learning tasks were loaded onto a general factor and only the sign learning tasks were loaded onto a specific factor—because the tasks defining the sign learning factor are a subset of the tasks defining the general lexical learning factor, this bifactor model was labeled Model SS for subset (see Figure 14D). The model fit the data well (see Table 8) and, as indicated by the AIC value, accounted for the data better than all other models.

4.2.3 Investigating the lexical learning and sign specific factor

Finally, in order to investigate the general lexical learning factor and the specific sign learning factor, these variables were regressed on the predictor variables (Figure 15). The model fit well, X^2 (246) = 377.25, p < .001, CFI = .965, SRMR = .043, RMSEA = .048 (90% CI = .038,

.057), with predictors accounting for 72% of variance in the lexical learning factor and 40% in the specific sign learning factor. As can be seen in Figure 15, the lexical learning factor was significantly predicted by the residualized fluid intelligence factor and spoken-PSTM, while the specific sign learning factor was only significantly predicted by signed-PSTM. Note, the path coefficient between the spoken-PSTM and specific sign learning factor, though equal to the coefficient between the residualized fluid intelligence and general lexical learning factors, was not significant—an outcome due to differences in standard errors.

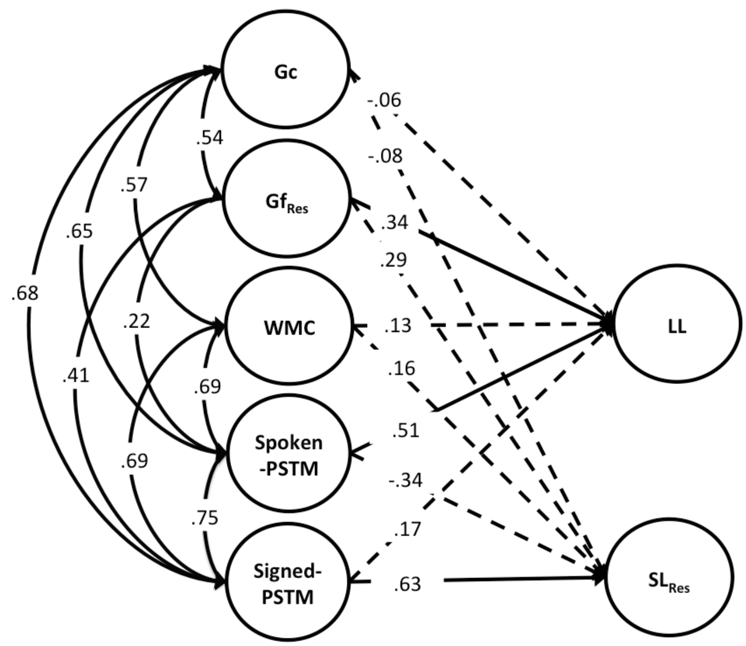


Figure 15. Model investigating the general lexical learning factor and specific sign learning factor. Note: LL = lexical learning factor; SL_{Res} = residualized (specific) sign learning factor.

5 Discussion

This study had two goals: first, to extend individual differences research in L2 word learning to sign learning and second, to examine the relationship between the two aforementioned constructs. Overall, the results of this study indicate that, amongst adult non-signers, associative L2 word learning and sign learning rely on similar processes, which can be partially accounted for by fluid intelligence, particularly *disengagement* (Engle, 2018; J. D. Martin et al., 2019; Shipstead et al., 2016), and modality-specific PSTM. Accordingly, the two constructs are highly related. Exploratory analyses revealed that, in fact, individual differences in lexical learning can be accounted for by a general lexical learning factor, however, sign learning engages additional sign specific processes. Regressing the general and specific factors on the predictors revealed that the lexical learning factor was significantly predicted by the residualized fluid intelligence factor and spoken-PSTM while the specific sign learning factor was only predicted by signed-PSTM, other factors held constant. Below, we elaborate on the results of this study, addressing the results of our confirmatory analyses before turning to our exploratory analyses.

5.1 Predicting Sign and Word Learning

5.1.1 Crystallized intelligence, fluid intelligence, and relation-construction

We expected that crystallized and fluid intelligence would be predictive of both sign learning and word learning because, regardless of the language modality, associative learning is partly determined by the quality and quantity of relationships constructed between items (Kyllonen et al., 1991). It was presumed that greater crystallized intelligence would enable a greater number of high-quality relationships; fluid intelligence, on the other hand, would enable the induction of appropriate relations and support the disposal of inappropriate ones.

Of the two intelligence constructs, only fluid intelligence accounted for a significant proportion of variance in sign learning and word learning over and above the other. Assuming that crystallized and fluid intelligence are generally involved in generating associations, then it is possible that the lack of an independent relationship between crystallized intelligence and the lexical learning variables was due to item characteristics, presentation order, and/or the amount of time given to study items.

Consider the relationships that could be formed between pairs in these two items: *electricity-banana* and *muz-banana*. (Before continuing, it may be instructive for the reader to actually attempt to construct relationships between each pair of words.) For the first pair, it should be fairly easy to generate associations for *both* words and to identify relationships. For example, when thinking about electricity, the following may come to mind: yellow, the symbol for a lightning bolt, Thomas Edison (the namesake of an electrical company), a light bulb, and Benjamin Franklin. For banana: yellow, mushy, fruit, breakfast, and mealy. From here, relationships can be formed linking electricity and banana, perhaps as an image of Thomas Edison holding a glowing yellow banana (as if it were a light bulb) or as a sentence: “Thomas Edison loved mushy bananas.” The Turkish word for banana, *muz,* however is unlikely to conjure up any associations independent of those that relate it to banana. So, for example, once one sees that *muz* is paired with *banana*, then one can observe that *muz* sounds somewhat similar to *mushy,* a feature of bananas.

For the first pair of words, *electricity-banana*, one drew upon crystallized intelligence to generate mediators; for the second pair, the role of crystallized intelligence was limited to one term in the pair—the familiar English word. The wealth of information present in the first case can facilitate the construction of a number of unique relationships that link *electricity* and *banana* together. Though it is certainly possible to generate more relationships between *muz* and *banana* than what was illustrated above (*mushy*), it is likely that, all other things being equal, the quantity and quality of relationships that can be generated between a familiar word and a highly unfamiliar lexical form will be less than that which can be generated for two familiar words.

The role that crystallized intelligence played in this study was likely further diminished by the fact that the unfamiliar lexical form was always presented first and the familiar word was presented second and only briefly (≤ 2 sec). By presenting the unfamiliar lexical form first, participants were limited in the associations they could generate before they saw the familiar word. Once they saw the familiar word, they only had a brief moment to attempt to form a relationship. This likely limited the role of crystallized intelligence.

In fact, it may be that processing speed acted as a suppressor variable—had processing speed been accounted for, crystallized intelligence *may* have been a significant predictor of lexical learning. To explain, studies have found that processing speed is related to associative learning (e.g., Kyllonen & Tirre, 1988; Kyllonen et al., 1991; Park et al., 1996; Salthouse, 1994; Salthouse & Dunlosky, 1995). In particular, Kyllonen et al. (1991) observed that when study time was brief (500 ms), fast processors tended to outperform slow processors independent of their verbal knowledge (a marker of crystallized intelligence). As study time increased (up to 8000 ms), the effect of verbal knowledge on lexical learning tended to increase while the effect of processing speed attenuated. Kyllonen and colleagues interpreted these results as indicating that, when study time was brief, fast processors were able to produce a greater number of relations compared to slow processors, however, as study time increased, individuals with high verbal ability were able to use the time to continue elaborating while individuals with low verbal ability were less able to do so. It would be worthwhile for future studies to include measures of processing speed and/or to manipulate the item characteristics, presentation order, and the amount of time given to study items.

5.1.2 Working memory capacity and phonological short-term memory

According to theory, both WMC and STM tasks assess executive attention, however, WMC is a domain-general construct that assesses executive attention to a greater degree than STM and STM additionally draws on domain-specific perceptual and motor processes (Cowan, 2008; Kane et al., 2004). In this study, it was presumed that the WMC factor would be a "purer" (i.e., more reliable) estimate of executive attention than PSTM, however, signed- and spoken-PSTM would account for modality-specific processes in sign and lexical learning. Thus we expected that WMC would be predictive of both sign and word learning, however, the PSTM factors would only be predictive within modality.

The results indicated that WMC did not account for a significant proportion of variance in either sign or word learning above that which was accounted for by the PSTM factors. This was somewhat surprising given that WMC is an excellent predictor of a range of tasks (Engle, 2002) and, when compared to STM, is often the superior predictor (Daneman & Merikle, 1996; Engle et al., 1999; Li, 2015; Linck et al., 2014). In hindsight, a likely explanation for the outcome observed here is that, in this study, fluid intelligence was included alongside WMC as a predictor. WMC and fluid intelligence are highly related factors (Foster et al., 2015; Kyllonen & Christal, 1990; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002) and so it is likely that, when fluid intelligence is not included as a predictor, WMC acts as a proxy for fluid intelligence (R. Engle, personal communication, August 15th, 2018). By including fluid intelligence alongside WMC, the role of WMC was largely reduced to maintaining information in a highly active state, a function accomplished by the PSTM factors, which *additionally* accounted for domain-specific processes. Thus the WMC factor was redundant.

5.1.3 The specificity and generality of phonological short-term memory

As expected, the full model (Figure 13) revealed that after accounting for other relevant variables, PSTM was only predictive of lexical learning within modality, indicating a certain degree of domain-specificity. This outcome supports Gathercole's (2006) theory that PSTM is related to lexical learning in part because of similarities in perceptual and motor processes.

The PSTM constructs also revealed a significant degree of domain-generality: the PSTM factors were strongly related to WMC and apparently accounted for the same portion of variance in lexical learning accounted for by WMC (see discussion above). Interestingly, of the two PSTM factors, signed-PSTM was the most general, exhibiting slightly larger correlations with most other factors and a substantially larger correlation with fluid intelligence. This is in line with prior work demonstrating that, compared to auditory-verbal abilities, visuospatial abilities tend to exhibit greater correlations with intelligence, particularly fluid intelligence (Groeger, Field, & Hammond, 1999; Kane et al., 2004; Lohman, 1996; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). In line with Miyake et al. (2001, p. 632), we posit that the relationship is due to familiarity, or the lack thereof.

Adult non-signers are quite adept at using speech-motor processes to aid in rehearsing auditory-verbal information; they are, however, unlikely to be skilled in rehearsing signs. Participants' experience with memorizing spoken language material (e.g., phone numbers, randomly generated passwords) may have biased them towards using a specific strategy during spoken-PSTM tasks, namely articulatory rehearsal. As such, there were likely few individual differences due to strategy use or *adaptivity* (Schunn & Reder, 2001). The novelty of signed-PSTM tasks, however, may have spurred variation in strategy use and it may be these differences that explain why signed-PSTM was more highly correlated with fluid intelligence and, in particular, with the residualized version of fluid intelligence.

In accordance with the idea that the residualized fluid intelligence factor reflects individuals' ability to disengage from outdated information, individuals with greater fluid intelligence may have optimized their performance by testing and discarding a number of strategies (Frankenmolen et al., 2017; Schunn & Reder, 2001). In fact, anecdotally, it was observed that some participants overtly rehearsed signs throughout the study, others initiated overt rehearsal at some point during the battery of signed-PSTM tasks, and some of those participants who used overt rehearsal seemingly abandoned the strategy during or between signed-PSTM tasks.

There are of course other possible explanations for the relationship between signed-PSTM and fluid intelligence. For example, the ability to disengage from outdated material may be related to the ability to mitigate proactive interference (Engle, 2018; Shipstead et al., 2016)—although attempts were made to reduce item similarity, the sign stimuli may not have been distinct enough compared to the spoken language material. Indeed, there is evidence that sign similarity has a greater detrimental effect on learning of novel signs in hearing non-signers compared to proficient signers (Siple et al., 1982). Clearly, more research is needed to investigate whether the magnitude of the correlations exhibited between the PSTM factors and fluid intelligence generalize and, if so, the cause.

5.2 Exploring the Relationship Between Sign and Word Learning

The second aim of this study was to investigate the relationship between L2 word learning and sign learning in non-signers. The two constructs were found to be highly correlated but not identical. Exploratory analyses revealed that all lexical learning tasks loaded onto a general factor, however, sign learning tasks loaded onto an additional specific factor. Furthermore, it was observed that spoken-PSTM and fluid intelligence were highly predictive of the general lexical learning factor, while the specific sign learning factor was only significantly predicted by signed-PSTM.

What these results suggest is that in hearing non-signers, all associative lexical learning tasks rely on similar processes, including fluid intelligence and spoken-PSTM. We have already specified the relationship between fluid intelligence and lexical learning and so we will not speak on this issue any further. With regard to spoken-PSTM, it appears that, no matter the modality, hearing non-signers rely on spoken language processes. This reliance on speech processes is likely due to the generation and rehearsal of labels and sentences to aid as cues as well the fact that half of the material being studied (i.e., the "translations") were English words. Sign learning, however, made additional demands and so a specific factor was needed to account for these processes, which, in hearing non-signers, are likely perceptuomotor rather than linguistic (Martinez & Singleton, 2018; Siple et al., 1982).

Future studies should investigate the conditions that affect the relationship between sign learning and word learning. For example, in this study, participants were instructed to use elaborative constructions to support learning and this may have increased the role of fluid intelligence and, consequently, the correlation between sign learning and word learning. It is possible that instructing participants to use rote rehearsal would increase the role of domain-specific processes and therefore lower the relationship between sign learning and word learning. These constructs can also be investigated in individuals with varying degrees of experience with a signed language (e.g., college students enrolled in their first semester of a sign language course compared to those enrolled in their third semester). On the one hand, increased experience with signs will enable individuals to effectively use rehearsal strategies which should increase domain-specific factors and therefore reduce the correlation between sign learning and word learning; on the other hand, experience may result in participants processing the signs linguistically (Newman-Norlund et al., 2006; J. T. Williams et al., 2016b), just as they do words, possibly increasing domain-generality and the correlation between sign learning and word learning. Quasi-experiments in conjunction with experimental, imaging, and computational studies can aid in explicating the relationship between sign and word learning.

5.3 Limitations

We have already discussed some limitations (e.g., not including a processing speed factor in our model) but here we point out a few more issues. First, our decision to model relationships amongst variables as we did was based on our own views but they could be modeled differently and fit the data just as well or better (Tomarken & Waller, 2003). For example, we modeled fluid intelligence as a predictor but it may be more accurate to state that the relationship is bidirectional.

A second issue has to do with the sample size. We expected fairly large relationships between the latent variables and so, based on that and our finite resources, we chose a sample size that allowed us to detect moderately sized effects. A larger sample, however, would result in more accurate estimates.

Third, we chose to estimate our lexical learning factors using associative tasks in which the English word was selected but there are many other ways that one could test lexical learning (e.g., provide the English word and ask participants to recall the target item). Whether the relationships observed here generalize to other lexical learning activities is an important question.

Finally, it is important to note that the pseudowords used in this study all followed English phonotactics while the pseudosigns did not necessarily adhere to the phonotactics of any particular language. The concern is that this resulted in a confound as individuals in this study all had experience with English but did not have experience with ASL or any other sign language. We chose to use English pseudowords because we were concerned that participants would focus attention on unfamiliar features and this would either disrupt normal word learning processes (e.g., rehearsal) or would allow individuals to use unfamiliar features as cues (e.g., a word produced with a click would be distinguished from all others). We did not, however, have this same concern with sign stimuli because, to the non-signer, all sign features should be relatively unfamiliar *except* in the case of iconic signs (Ortega & Morgan, 2015; Ortega, Ozyurek, & Peeters, 2019), which we attempted to avoid. While this potential confound is a valid concern, we note that we did use Turkish in one of our word learning tasks and it showed similar relationships as our other word learning tasks. Furthermore, if experience with English confounded the results, then it is all the more impressive that the sign- and spoken-language factors were so highly correlated. Lastly, it is also interesting to note that crystallized intelligence was not a significant predictor of either lexical learning factors—one would have expected greater process overlap between crystallized intelligence and word learning, as crystallized intelligence as estimated here and elsewhere is largely a spoken language variable (Carroll, 1993; Kan, Kievit, Dolan, & der Maas, 2011). Still, as noted above, the concern that experience with English confounded the results is valid and should be explored.

5.4 Summary and Implications

The results of this study corroborate and extend prior research on L2 lexical learning in spoken languages. Specifically, it was found that fluid intelligence and modality-specific PSTM were predictive of sign learning just as they have been found to be predictive of word learning (Baddeley et al., 1998; Kyllonen & Tirre, 1988). Interestingly, two predictors that were assumed to be important to lexical learning, crystallized intelligence and WMC, were not statistically significant in this study. It is suggested that the effect of crystallized intelligence may have been suppressed by processing speed due to the brief study period used here; individual differences in WMC on the other hand, were likely accounted for by the PSTM and fluid intelligence factors and therefore WMC was redundant. It was also observed that word learning and sign learning are highly correlated but partially distinct. Subsequent analyses revealed that all tasks loaded onto a general lexical learning factor but sign learning tasks additionally loaded onto a specific factor. As such, this study provides insight into the cognitive processes that are common to associative lexical learning regardless of language modality and those that are unique to either signed or spoken languages. Future studies should continue to investigate the relationship between word learning and sign learning as well as other aspects of L2 learning, such as grammar learning.

Acknowledgements

This research was supported by National Science Foundation award #1651115. We 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. Additionally, we would like to thank Drs. Michael Bunting, Susan Embretson, Randy Engle, and Dan Spieler.

REFERENCES

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working Memory and Intelligence: The Same or Different Constructs? *Psychological Bulletin, 131*(1), 30-60. doi:10.1037/0033-2909.131.1.30

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology, 18*(4), 362-365. doi:10.1080/14640746608400055

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

Bavelier, D., Corina, D., Jezzard, P., Clark, V., Karni, A., Lalwani, A., . . . Neville, H. J. (1998). Hemispheric specialization for English and ASL: Left invariance-right variability. *Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience, 9*(7), 1537-1542. doi:10.1097/00001756-199805110-00054

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

Bellugi, U., Klima, E. S., & Siple, P. (1974). Remembering in signs. *Cognition, 3*(2), 93-125. doi:10.1016/0010-0277(74)90015-8

Bentler, P. M. (2001). *EQS 6 structural equations program manual*. Encino, CA: Multivariate software.

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: a new self-report imagery questionnaire. *Applied Cognitive Psychology, 20*(2), 239-263. doi:10.1002/acp.1182

Boutla, M., Supalla, T., Newport, E. L., & Bavelier, D. (2004). Short-term memory span: Insights from sign language. *Nature Neuroscience, 7*(9), 997-1002. doi:10.1038/nn1298

Brentari, D. (1998). *A prosodic model of sign language phonology*. Massachusetts: MIT Press.

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.

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

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence, 15*(2), 229-246. doi:10.1016/0160-2896(91)90032-9

Cardin, V., Orfanidou, E., Kästner, L., Rönnberg, J., Woll, B., Capek, C. M., & Rudner, M. (2016). Monitoring Different Phonological Parameters of Sign Language Engages the Same Cortical Language Network but Distinctive Perceptual Ones. *Journal Of Cognitive Neuroscience, 28*(1), 20. doi:10.1162/jocn_a_00872

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*(3), 404.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*: Cambridge University Press.

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

Colom, R., Shih, P. C., Flores-Mendoza, C., & Quiroga, M. Á. (2006). The real relationship between short-term memory and working memory. *Memory, 14*(7), 804-813. doi:10.1080/09658210600680020

Conrad, R. (1970). Short-term memory processes in the deaf. *British Journal Of Psychology (London, England: 1953), 61*(2), 179-195.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*(4), 429-432. doi:10.1111/j.2044-8295.1964.tb00928.x

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, C. J. (1964). Some relationships between paired-associates learning and foreign-language aptitude. *Journal of Educational Psychology, 55*(3), 132-138. doi:10.1037/h0044223

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

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review, 24*(4), 1158-1170. doi:10.3758/s13423-016-1191-6

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422-433. doi:10.3758/bf03214546

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin, 145*(5), 508-535. doi:http://dx.doi.org/10.1037/bul0000192

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77*(5), 615-622. doi:10.1037/0021-9010.77.5.615

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

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367-413. doi:10.1111/j.1467-9922.2010.00613.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.

Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science, 11*(1), 19-23. doi:10.1111/1467-8721.00160

Engle, R. W. (2018). Working Memory and Executive Attention: A Revisit. *Perspectives on Psychological Science, 13*(2), 190-193. doi:10.1177/1745691617720478

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309-331. doi:10.1037/0096-3445.128.3.309

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

Frankenmolen, N. L., Altgassen, M., Kessels, R., de Waal, M. M., Hindriksen, J.-A., Verhoeven, B., . . . Oosterman, J. M. (2017). Intelligence moderates the benefits of strategy instructions on memory performance: an adult-lifespan examination. *Aging, Neuropsychology, and Cognition, 24*(1), 45-61. doi:10.1080/13825585.2016.1171289

Gardner, R. C., & Lambert, W. E. (1965). Language aptitude, intelligence, and second-language achievement. *Journal of Educational Psychology, 56*(4), 191-199. doi:10.1037/h0022400

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., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*(3), 336-360. doi:10.1016/0749-596X(90)90004-J

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

Glaze, J. A. (1928). The association value of non-sense syllables. *The Pedagogical Seminary and Journal of Genetic Psychology, 35*, 255-269. doi:10.1080/08856559.1928.10532156

Goldberg, D., Looney, D., & Lusin, N. (2015). Enrollments in Languages Other Than English in United States Institutions of Higher Education, Fall 2013. Retrieved June 15th, 2016, from Modern Language Association https://www.mla.org/content/download/31180/1452509/EMB_enrllmnts_nonEngl_2013.pdf

Gómez, M. J. L., Molina, T. B., Benítez, P. P., & de Torres, J. S. (2007). Predicting proficiency in signed language interpreting: A preliminary study. *Interpreting, 9*(1), 71-93.

Groeger, J. A., Field, D., & Hammond, S. M. (1999). Measuring Memory Span. *International Journal of Psychology, 34*(5-6), 359-363. doi:10.1080/002075999399693

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

Hasher, L., & Zacks, R. T. (1988). Working Memory, Comprehension, and Aging: A Review and a New View. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 22, pp. 193-225): Academic Press.

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher, A. Koriat, D. Gopher, & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application.* (pp. 653-675). Cambridge, MA, US: The MIT Press.

Hirshorn, E. A., Fernandez, N. M., & Bavelier, D. (2012). Routes to short-term memory indexing: Lessons from deaf native users of American Sign Language. *Cognitive Neuropsychology, 29*(1-2), 85-103. doi:10.1080/02643294.2012.704354

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty and M. H. Long (Eds.). *The handbook of second language acquisition (pp. 349-381), Malden, MA: Blackwell.*

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

Hundal, P. S., & Horn, J. L. (1977). On the relationships between short-term learning and fluid and crystallized intelligence. *Applied Psychological Measurement, 1*(1), 11-21. doi:10.1177/014662167700100104

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

Jenkins, J. J. (1985). Nonsense syllables: Comprehending the 'almost incomprehensible variation'. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(3), 455-460. doi:10.1037/0278-7393.11.3.455

Jones, D. M., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 1008-1018. doi:10.1037/0278-7393.21.4.1008

Jones, D. M., Hughes, R. W., & Macken, W. J. (2006). Perceptual organization masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory and Language, 54*(2), 265-281. doi:10.1016/j.jml.2005.10.006

Kan, K.-J., Kievit, R. A., Dolan, C., & der Maas, H. v. (2011). On the interpretation of the CHC factor Gc. *Intelligence, 39*(5), 292-302. doi:https://doi.org/10.1016/j.intell.2011.05.003

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(2), 336-358. doi:10.1037/0278-7393.26.2.336

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

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence, 37*(4), 374-382. doi:10.1016/j.intell.2009.03.004

Klima, E. S., & Bellugi, U. (1979). *The Signs of Language*. Cambridge, MA: Harvard University Press.

Kline, R. B. (2016). *Principles and practice of structural equation modeling (4th ed.)*. New York, NY, US: Guilford Press.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal, 73*(4), 440-464. doi:10.2307/326879

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*(4), 389-433. doi:10.1016/S0160-2896(05)80012-1

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

Kyllonen, P. C., & Woltz, D. J. (1989). Role of cognitive factors in the acquisition of cognitive skill. In R. Kanfer & R. Cudeck (Eds.), *Abilities, motivation, and methodology: The Minnesota Symposium on Learning and Individual Differences.* (pp. 239-280). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

La Pointe, L. B., & Engle, R. W. (1990). Simple and Complex Word Spans as Measures of Working Memory Capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(6), 1118-1133. doi:10.1037/0278-7393.16.6.1118

Li, S. (2015). The construct validity of language aptitude. *Studies in second language acquisition*, 1-42. doi:10.1017/S027226311500042X

Lilienthal, L., Tamez, E., Myerson, J., & Hale, S. (2013). Predicting performance on the Raven's Matrices: The roles of associative learning and retrieval efficiency. *Journal of Cognitive Psychology, 25*(6), 704-716. doi:10.1080/20445911.2013.791299

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21*(4), 861-883. doi:10.3758/s13423-013-0565-2

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

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (4th ed.). New York, NY: Lawrence Erlbaum Associates Publishers.

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.

Malaia, E., & Wilbur, R. B. (2018). Visual and linguistic components of short-term memory: Generalized Neural Model (GNM) for spoken and sign languages. *Cortex*. doi:10.1016/j.cortex.2018.05.020

Mann, W., Marshall, C., Mason, K., & Morgan, G. (2010). The acquisition of sign language: The impact of phonetic complexity on phonology. *Language Learning and Development, 6*(1), 60-86. doi:10.1080/15475440903245951

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*(3), 519-530.

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

Marshall, C., Mason, K., Rowley, K., Herman, R., Atkinson, J., Woll, B., & Morgan, G. (2015). Sentence repetition in deaf children with specific language impairment in British Sign Language. *Language Learning and Development, 11*(3), 237-251. doi:10.1080/15475441.2014.917557

Martin, J. D., Shipstead, Z., Harrison, T., Redick, T. S., Bunting, M., & Engle, R. W. (2019). The role of maintenance and disengagement in predicting reading comprehension and vocabulary learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: http://dx.doi.org/10.1037/xlm0000705

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

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

Mason, K., Rowley, K., Marshall, C. R., Atkinson, J. R., Herman, R., Woll, B., & Morgan, G. (2010). Identifying specific language impairment in deaf children acquiring British Sign Language: Implications for theory and practice. *British Journal of Developmental Psychology, 28*(1), 33-49. doi:10.1348/026151009X484190

Mayberry, R. I. (2010). Early language acquisition and adult language ability: What sign language reveals about the critical period for language. In M. Marshark & P. E. Spencer (Eds.), *The Oxford handbook of deaf studies, language, and education, Vol 2.* (pp. 281-291). New York, NY, US: Oxford University Press.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130*(4), 621-640. doi:10.1037/0096-3445.130.4.621

Nation, I. S. P. (1980). Strategies for receptive vocabulary learning. *Guidelines, 3*(1), 23.

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.

Noble, C. E. (1952). An analysis of meaning. *Psychological Review, 59*(6), 421-430. doi:10.1037/h0054087

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

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in second language acquisition, 29*(4), 557-582. doi:10.1017/S027226310707043X

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

Oberauer, K., Lewandowsky, S., Avh, E., Brown, G. D., Conway, A. R. A., Covan, N., . . . Hurlstone, M. (2018). Benchmarks for models of short term and working memory. *Psychological Bulletin*.

Ortega, G., & Morgan, G. (2015). Phonological development in hearing learners of a sign language: The influence of phonological parameters, sign complexity, and iconicity. *Language Learning, 65*(3), 660-688. doi:10.1111/lang.12123

Ortega, G., Ozyurek, A., & Peeters, D. (2019). Iconic gestures serve as manual cognates in hearing second language learners of a sign language: an ERP study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

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

Park, D. C., Smith, A. D., Lautenschlager, G., Earles, J. L., Frieske, D., Zwahr, M., & Gaines, C. L. (1996). Mediators of long-term memory performance across the life span. *Psychology and Aging, 11*(4), 621-637. doi:10.1037/0882-7974.11.4.621

Poizner, H., Klima, E. S., & Bellugi, U. (1987). *What the hands reveal about the brain*. Cambridge, MA, US: The MIT Press.

Pufahl, I., & Rhodes, N. C. (2011). Foreign Language Instruction in U.S. Schools: Results of a National Survey of Elementary and Secondary Schools. *Foreign Language Annals, 44*(2), 258-288. doi:10.1111/j.1944-9720.2011.01130.x

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

Rohrer, D., Taylor, K., Pashler, H., Wixted, J. T., & Cepeda, N. J. (2005). The Effect of Overlearning on Long-Term Retention. *Applied Cognitive Psychology, 19*(3), 361-374. doi:10.1002/acp.1083

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

Salthouse, T. A. (1994). Aging associations: Influence of speed on adult age differences in associative learning. *Journal of Experimental Psychology, 20*(6), 1486.

Salthouse, T. A., & Dunlosky, J. (1995). Analyses of adult age differences in associative learning. *Zeitschrift Für Psychologie Mit Zeitschrift Für Angewandte Psychologie, 203*(4), 351-360.

Sandler, W. (1989). *Phonological Representation of the Sign : Linearity and Nonlinearity in American Sign Language*. Berlin/Boston: De Gruyter Mouton.

Sasaki, M. (1993). Relationships among second language proficiency, foreign language aptitude, and intelligence: A structural equation modeling approach. *Language Learning, 43*(3), 313-344.

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

Schunn, C. D., & Reder, L. M. (2001). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology: General, 130*(1), 59-76. doi:10.1037/0096-3445.130.1.59

Seibert, L. C. (1930). An experiment on the relative efficiency of studying French vocabulary in associated pairs versus studying French vocabulary in context. *Journal of Educational Psychology, 21*(4), 297-314. doi:10.1037/h0070517

Shaw, S. (2011). Cognitive and motivational contributors to aptitude: A study of spoken and signed language interpreting students. *Interpreting, 13*(1), 70-84. doi:10.1075/intp.13.1.05sha

Shipstead, Z., & Engle, R. W. (2013). Interference within the focus of attention: Working memory tasks reflect more than temporary maintenance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 277-289. doi:10.1037/a0028467

Shipstead, Z., Harrison, T., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science, 11*(6), 771-799. doi:doi:10.1177/1745691616650647

Siple, P., Caccamise, F., & Brewer, L. (1982). Signs as pictures and signs as words: Effect of language knowledge on memory for new vocabulary. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(6), 619-625. doi:10.1037/0278-7393.8.6.619

Söderfeldt, B., Ingvar, M., Rönnberg, J., Eriksson, L., Serrander, M., & Stone-Elander, S. (1997). Signed and spoken language perception studied by positron emission tomography. *Neurology, 49*(1), 82-87. doi:10.1159/000053093

Sparks, R. L., Humbach, N., Patton, J., & Ganschow, L. (2011). Subcomponents of second-language aptitude and second-language proficiency. *Modern Language Journal, 95*(2), 253-273. doi:10.1111/j.1540-4781.2011.01176.x

Stone, C. A. (2017). Sign language interpreter aptitude: The trials and tribulations of a longitudinal study. *Translation & Interpreting, 9*(1), 72-87.

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability--and a little bit more. *Intelligence, 30*(3), 261-288. doi:10.1016/S0160-2896(01)00100-3

Tamez, E., Myerson, J., & Hale, S. (2008). Learning, working memory, and intelligence revisited. *Behavioural Processes, 78*(2), 240-245. doi:10.1016/j.beproc.2008.01.008

Thorndike, E. L. (1908). Memory for paired associates. *Psychological Review, 15*(2), 122-138. doi:10.1037/h0073570

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

Tomarken, A. J., & Waller, N. G. (2003). Potential Problems With "Well Fitting" Models. *Journal of Abnormal Psychology, 112*(4), 578-598. doi:10.1037/0021-843X.112.4.578

Tremblay, S., Saint-Aubin, J., & Jalbert, A. (2006). Rehearsal in serial memory for visual-spatial information: Evidence from eye movements. *Psychonomic Bulletin & Review, 13*(3), 452-457. doi:10.3758/BF03193869

Ullman, J. B. (2006). Structural Equation Modeling: Reviewing the Basics and Moving Forward. *Journal of Personality Assessment, 87*(1), 35-50. doi:10.1207/s15327752jpa8701_03

Underwood, B. J., Boruch, R. F., & Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General, 107*(4), 393-419. doi:10.1037/0096-3445.107.4.393

Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica, 134*(1), 16-28. doi:10.1016/j.actpsy.2009.11.010

Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin, 145*(1), 79-139. doi:10.1037/bul0000176

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: The role of controlled search. *Memory & Cognition, 41*(2), 242-254. doi:10.3758/s13421-012-0261-x

Unsworth, N., & Engle, R. W. (2007). On the Division of Short-Term and Working Memory: An Examination of Simple and Complex Span and Their Relation to Higher Order Abilities. *Psychological Bulletin, 133*(6), 1038-1066. doi:10.1037/0033-2909.133.6.1038

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2012). Working memory capacity and retrieval limitations from long-term memory: An examination of differences in accessibility. *The Quarterly Journal of Experimental Psychology, 65*(12), 2397-2410. doi:10.1080/17470218.2012.690438

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

Verhagen, J., & Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology, 141*, 65-82. doi:10.1016/j.jecp.2015.06.015

Ward, G., Avons, S. E., & Melling, L. (2005). Serial position curves in short-term memory: Functional equivalence across modalities. *Memory, 13*(3-4), 308-317. doi:10.1080/09658210344000279

Westland, C. J. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications, 9*(6), 476-487. doi:http://dx.doi.org/10.1016/j.elerap.2010.07.003

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology, 4*, 433. doi:10.3389/fpsyg.2013.00433

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

Williams, J. T., Darcy, I., & Newman, S. D. (2016a). The beneficial role of L1 spoken language skills on initial L2 sign language learning: Cognitive and linguistic predictors of M2L2 acquisition. *Studies in second language acquisition*, 1-18. doi:10.1017/S0272263116000322

Williams, J. T., Darcy, I., & Newman, S. D. (2016b). Modality-specific processing precedes amodal linguistic processing during L2 sign language acquisition: A longitudinal study. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 75*, 56-67. doi:10.1016/j.cortex.2015.11.015

Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review, 8*(1), 44-57. doi:10.3758/BF03196138

Wilson, M., & Emmorey, K. (1997). A visuospatial 'phonological loop' in working memory: Evidence from American Sign Language. *Memory & Cognition, 25*(3), 313-320. doi:10.3758/BF03211287

Wu, Y. C., & Coulson, S. (2014). A psychometric measure of working memory capacity for configured body movement. *PLOS ONE, 9*(1).

Yang, L. (1997). Tracking the acquisition of L2 vocabulary: The Keki language experiment. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (pp. 125-156). Cambridge: Cambridge University Press.