Construct Validity of Procedural Memory Tasks Used in Adult-Learned Language

Joshua Buffington (bffngtn2@uic.edu)
Department of Psychology, 1007 W. Harrison St.
Chicago, IL 60607 USA

Kara Morgan-Short (karams@uic.edu)
Department of Hispanic and Italian Studies, and Department of Psychology, 601 S. Morgan St.
Chicago, IL 60607 USA

Abstract

Research has examined the role of domain-general cognitive factors in second language (L2) acquisition, with emerging evidence implicating a role for procedural memory, a long-term memory system (e.g., Morgan-Short et al., 2014). Strong conclusions regarding the role of procedural memory are hindered by the lack of knowledge regarding the reliability and validity of procedural memory assessments. In this study, participants completed three assessments of procedural memory that have previously been used to study L2 learning, along with assessments of declarative memory, working memory, and an artificial L2 learning task. Results indicated that the procedural memory assessments generally showed evidence of reliability and discriminant validity, but, somewhat surprisingly, evidence for convergent validity was lacking. Finally, one procedural memory assessment showed predictive validity for the L2 learning task. Implications for future research on the role of procedural memory in L2 acquisition will be considered in light of these results.

Keywords: procedural memory; second language acquisition; reliability; construct validity

Introduction

Cognitive-based research in second language (L2) acquisition addresses how people learn languages by examining what mechanisms are involved in the learning process. In regard to domain-general mechanisms, there has been a substantial amount of research on the role of attention, awareness, and working memory in L2 acquisition (for review, see Dörnyei, 2006), but research on long-term memory and L2 acquisition is only emerging and merits further research (Hamrick, Lum, & Ullman, 2018). The current study focuses on one type of long-term memory called procedural memory that has been posited to play a role in learning the grammatical rules of language (DeKeyser, 2015; Paradis, 2009; Ullman, 2015).

Procedural memory is a type of implicit memory that supports the acquisition of cognitive and motor skills, as well as habits (Eichenbaum, 2011; Ullman, 2004; 2015). It may be contrasted with other memory systems such as declarative memory that support memory for facts and personal, episodic experiences (Eichenbaum, 2011; Ullman, 2004; 2015). Procedural memory may be described by a number of neurocognitive characteristics: (a) Learning is "implicit," meaning that it does not involve conscious awareness (Ullman, 2004; 2015). (b) Procedural memory is not facilitated by attention, and indeed attention may interfere with learning in procedural memory (Foerde, Knowlton, & Poldrack, 2006). (c) The development of knowledge occurs gradually and improves over multiple learning trials (Ullman, 2004; 2015). (d) Knowledge is typically encapsulated, meaning that it is unavailable for use by other memory systems and generally inflexible with respect to the contexts in which it can be applied (Ullman, 2004). (e) Procedural memory neuroanatomy involves a fronto-striatal circuit in which information is relayed from the cortex to the striatum (part of the basal ganglia), then to the thalamus, then back to frontal cortex (Eichenbaum, 2011).

Three theories predict a role for procedural memory in L2 acquisition: the Skill Acquisition Model (DeKeyser, 2015), Ullman's Declarative/Procedural model (Ullman, 2004; 2015), and Paradis' claims regarding declarative and procedural determinants of L2 (Paradis, 2009). These theories differ with respect to specific predictions about how procedural memory plays a role in L2 acquisition, but they all view procedural memory as potentially involved in the fluent production and comprehension of grammatical structures in a second language, at least at higher levels of proficiency. Empirical evidence suggests that procedural memory contributes to L2 acquisition largely in the manner predicted by these theories (e.g., Antoniou, Ettlinger, & Wong, 2016; Faretta-Stutenberg & Morgan-Short, 2018; Hamrick, 2015; Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong, 2014; see Buffington & Morgan-Short, in press; Hamrick et al., 2018 for review).

The studies examining procedural memory in L2 have used four procedural memory tasks: the Alternating Serial Reaction Task (ASRT), the Serial Reaction Task (SRT), the Weather Prediction Task (WPT), and the Tower of London (TOL).[1] In the ASRT, used by Faretta-Stutenberg and Morgan-Short (2018) among others, participants respond to a filled-in circle whose location alternates in a second-order pattern. Previous research shows that participants gradually improve on this task without demonstrating explicit

[1] Here we focus on the Alternating Serial Reaction Task rather than on the Serial Reaction Task as the two tasks are very similar, and the Alternating Serial Reaction Task has been used more often in previous L2 empirical work.

awareness of the pattern (Howard & Howard, 1997), and other work provides indirect evidence for the use of procedural memory neural circuits on the ASRT (Fletcher et al., 2005), with related sequence learning tasks such as the SRT showing direct evidence of the engagement of procedural memory neural circuits (e.g., Rauch et al., 1997). In the WPT, used by Morgan-Short and colleagues (2014) among others, participants predict fictional weather outcomes based on probabilistic cues. Previous research shows that a dual-task version, which involves a secondary tone-counting task, does not involve declarative knowledge about cue-outcome associations and is positively correlated with activity in a procedural memory neural structure (Foerde et al., 2006). In the TOL, used by Antoniou and colleagues (2016) among others, participants are instructed to match a goal configuration of circles by moving the circles around on pegs. Though developed to assess executive functions (Kaller, Unterrainer, & Stahl, 2012), evidence for the use of procedural memory on this task comes from behavioral research showing implicit, gradual improvement in accuracy and time to complete the task (Ouellet, Beauchamp, Owen, & Doyon, 2004), as well as neuroimaging and neuropsychological research implicating a role for procedural memory neural structures in task performance (Beauchamp, Dagher, Aston, & Doyon, 2003; Owen et al., 1992). In sum, the implicit, gradual acquisition of knowledge and activation of procedural memory neural circuits during learning on these three tasks supports the claim that they assess procedural memory learning abilities.

Though previous research suggests that participants use procedural memory when learning these tasks, none of the tasks have been validated as psychometric tools to predict the role of procedural memory abilities in L2 acquisition. Morgan-Short et al. (2014) provide data showing that the WPT and TOL are positively related to each other, but the WPT also trended to a positive association with a test of declarative memory, which provides reason to question the discriminant validity of the WPT. These data motivate a more thorough examination of the psychometric properties of tasks used to assess procedural memory in previous work. Such an examination will permit stronger conclusions in future work regarding the role of procedural memory in L2 acquisition.

In order to answer these questions, the present study includes all three assessments in a within-subjects, counter-balanced design, along with assessments of declarative memory, working memory, and L2 learning. Though preliminary due to the small sample size, this study will provide insight into the following questions: (1) Reliability: Do procedural memory assessments demonstrate internal consistency? (2) Convergent validity: Do different procedural memory assessments correlate with each other? (3) Discriminant validity: Do procedural memory assessments *not* correlate with assessments of other cognitive abilities, such as declarative and working memory? (4) Predictive validity: Which procedural memory assessments correlate with L2 learning ability?

Method

Participants

Participants (N = 31, 19 female, age 18-24 years, mean = 19.7 years) were recruited through the psychology subject pool at the University of Illinois at Chicago. All participants received course credit for completing the study. There were no language background or other selection requirements.

Materials and Procedure

Procedural Memory Tasks. Participants completed three assessments of procedural memory. The first assessment, the ASRT, was based on the same task from Howard and Howard (1997) and presented participants with a sequence learning task in which an item of the sequence consisted of a circle in a row of four circles being filled in by a dog's head. The sequence involved an eight-element alternating sequence where patterned trials alternated with random trials. As such, participants might see a repeating sequence such as 3r1r4r2r, where the numbers correspond to the location of the dog's head in the row of circles and "r" represents a random location from one to four. The task instructed participants to press a key corresponding to the location of the dog's head as quickly and accurately as possible. Participants would press either the *z*, *c*, *b*, or *m* keys on a QWERTY keyboard (using their left and right middle and index fingers, respectively), with *z* corresponding to the leftmost circle position and *m* to the rightmost position. The task consisted of 20 blocks, with 85 trials per block (5 random trials followed by 80 alternating patterned and random trials). Learning was assessed by comparing reaction times on patterned vs. random trials, with a greater difference reflecting greater procedural learning on the task, following Faretta-Stutenberg and Morgan-Short (2018). Lastly, reliability was assessed by calculating the Spearman-Brown split-half coefficient.

The second assessment of procedural memory was the dual-task version of the WPT (Foerde et al., 2006). The WPT assesses knowledge of probabilistic weather outcomes associated with combinations of cue cards. The cue cards combine with each other and each combination is associated with a certain probability of sunshine or rain. For example, a combination of a card with circles and a card with squares may be associated with an 80% chance of sunshine. Participants were instructed to predict the weather based on a cue combination, and then the weather outcome appeared on the screen. In this dual-task version, participants were also tasked with keeping track of the number of high tones that occurred during each trial. The high tones were pseudo-randomly interspersed with low tones, but participants were instructed to only count the high tones. This secondary task has been shown to impede the use of declarative memory on the weather prediction component of the task (Foerde et al., 2006). Thus, each trial consisted of making a weather prediction while counting the number of high tones. The task included a total of 320 dual-task trials divided into 8 blocks, with 40 trials per block. Learning was assessed by examining accuracy of weather prediction on the final dual-task block, where chance performance was 50%, following Faretta-Stutenberg and Morgan-Short (2018). Reliability was assessed by calculating Cronbach's alpha.

The final assessment of procedural memory was the TOL (Kaller, Unterrainer, & Stahl, 2012). In this task, participants matched a goal configuration of colored circles that rest on pegs. In producing the goal configuration, participants could only move the topmost circle on each peg, and when moved the circle would fall to the lowest possible peg position. Participants were instructed to plan their sequence of moves before beginning the first move, and to do their best in completing the goal configuration in the stated number of moves, which began at three moves and increased to six moves by the end of the task. Per Antoniou and colleagues (2016), participants repeated this task immediately after completing it, and procedural learning was assessed by examining the average total time to match a goal configuration on the second administration of the task, normalized relative to the other participants, as a measure of improvement. Reliability was assessed by calculating the Spearman-Brown split-half coefficient.

Declarative and Working Memory Tasks. Participants completed two assessments of declarative memory, a verbal learning segment of the Modern Language Aptitude Test (MLAT, Carroll & Sapon, 1959) and the Continuous Visual Memory Test (CVMT, Trahan & Larrabee, 1988). The MLAT, a paired-associates learning task, assessed participants' learning of English translations of 24 pseudo-Kurdish words. Participants studied the word-pair list for four minutes, and then proceeded to a five-option multiple choice assessment that instructed them to select the correct English translation given one of the pseudo-Kurdish words. They had four minutes to complete this assessment. Accuracy on the assessment was measured as the dependent variable of word learning, and reliability was indicated by calculating Cronbach's alpha.

The CVMT assessed recognition of abstract shapes. In this task, participants saw a series of abstract images and responded whether they had previously seen the image. Seven images were repeated, whereas 63 images were only presented once. Performance on the CVMT was measured by computing d' scores and reliability was assessed by calculating Cronbach's alpha. Although the MLAT and CVMT do not assess exactly the same processes, because the MLAT is a paired-associates task of verbal learning and the CVMT is a recognition test of visual learning, both can be described as declarative memory assessments because they involve memory for factual information and/or episodic experiences (Eichenbaum, 2011).

Participants completed three shortened working memory assessments, taken from Oswald, McAbee, Redick, and Hambrick (2015): operation span (OSpan), reading span (RSpan), and symmetry span (SSpan). Oswald et al. present evidence for highly similar psychometric properties for the shortened vs. full-length assessments. In each working memory assessment, participants were asked to make judgments about a series of items (e.g., decide if an arithmetic operation is true, decide if a sentence makes sense) and then recall a list of elements (e.g., letters, locations in a matrix) so the two essential elements of working memory, processing and storage, were involved in each task. Performance on each working memory assessment was measured as the partial-credit score, which is the proportion of correctly recalled elements in each trial. In order to be included in analysis, participants had to score above 80% on the processing component of the working memory task. To measure working memory, a composite score was created by normalizing scores on each task and then averaging them. Reliability was not calculated because Oswald et al. (2015) provide data in support of the reliability of these assessments.

Artificial L2 Task. The artificial L2 in this study comes from Ettlinger and colleagues (2014), who showed that learning on this L2 task was associated with procedural memory. The language consists of 30 noun stems denoting different animals and two affixes (one denoting plural and the other diminutive) that can combine with nouns. In addition, there are two types of word formation rules, a simple rule involving application of a pattern of morphemes (L2 pattern), and a complex rule involving vowel changes that can be learned via analogy with similar words (L2 analogistic). As an example of patterned word formation, [pag] means dog, [ka-] is diminutive, and [-il] is the plural morpheme, so [ka-pag-il] denotes "little dogs." An example of analogistic word formation involving vowel changes is [ka-maz-el], meaning "little cows," which consists of the underlying morphemes [mez] (cow), [ka-], and [-il]. In the language, nouns with [-e-] stems trigger the analogistic word formation rule. Here we focus on the L2 pattern rule because previous research has shown procedural memory to be associated with the L2 pattern, but not the L2 analogistic, rule (Antoniou et al., 2016; Ettlinger et al., 2014). During learning, participants saw twelve nouns in all four possible forms (singular, diminutive, plural, and diminutive plural) and each form was repeated four times, resulting in 192 implicit exposures with words presented auditorily and paired with the corresponding image. Following this acquisition phase, L2 grammar learning was assessed by presenting novel nouns and then asking participants to select the correct form of the word in a two-option forced-choice design (e.g., given [tib] "tiger" choose the correct form of “little tigers” given the correct [katibil] and the incorrect [katibel] forms). Accuracy on the assessment phase was recorded as the dependent measure of L2 grammar learning. Because words were only analogistic in the diminutive plural form, which also included words formed with the L2 pattern rule, only questions involving the diminutive plural were included in analysis (per Ettlinger et al., 2014). Reliability was measured by calculating Cronbach’s alpha.

Procedure. Participants completed the study over two 2-hour sessions scheduled on separate days. The order of tasks in each session was partially counterbalanced to avoid fatigue effects, particularly because the procedural memory assessments are longer assessments (each takes between 20-35 minutes to complete). In Session 1, participants provided informed consent and then completed the WPT, ASRT, CVMT, and MLAT. The order of the procedural memory assessments was counterbalanced and the declarative memory assessments were also counterbalanced across participants. One of the procedural memory tasks always occurred as the first task of Session 1, and procedural memory tasks alternated with declarative memory tasks. Session 2 consisted of the other procedural memory assessment (TOL), working memory assessments, and the artificial L2 task. The order of the TOL and the working memory assessments was counterbalanced. Per Oswald et al. (2015) the order within the working memory assessments was fixed, as follows: OSpan, SSpan, RSpan. The artificial L2 task was always the final task of Session 2.

Results

Before analysis, the data was cleaned by casewise deletion of participants who failed to complete one or more of the cognitive tasks, resulting in the deletion of four participants for a final total of 27 participants included in the analyses. Learning on all tasks was considered by the 95% confidence intervals being above chance performance, which showed the following (Table 1): (a) For the procedural memory tasks, learning was evidenced for the ASRT, but not for the WPT. The TOL scores were normalized so the absolute numbers are not meaningful. (b) Learning was observed for both declarative memory tasks, MLAT and CVMT. (c) The working memory scores were normalized so the absolute values are not meaningful, but Oswald et al. (2015) provide convincing evidence that these tasks accurately and reliably measure working memory capacity. (d) For the L2 tasks, learning was observed for the L2 pattern rule but not the L2 analogistic rule. The L2 analogistic rule actually showed significantly below-chance learning, suggesting that participants were incorrectly applying the L2 pattern rule for analogistic cases.

To address the first research question, DO PROCEDURAL MEMORY ASSESSMENTS DEMONSTRATE RELIABILITY?, a reliability coefficient above .70 was considered “acceptable” reliability (Lance, Butts, & Michels, 2006). This yielded acceptable reliability for the WPT and ASRT, but low reliability for the TOL (Table 1). Table 1 also includes reliability values for the declarative memory and L2 tasks. Notably, both L2 tasks showed low reliability, with particularly low reliability for the L2 analogistic rule.

Table 1: Descriptive statistics and reliability

	M	SD	95% CIs	Reliability
WPT	55.76	16.83	49.10, 62.41	.79[a]
ASRT	5.51	7.01	2.73, 8.28*	.99[b]
TOL	.48	.30	.37, .60	.54[b]
MLAT	14.41	5.79	12.12, 16.70*	.87[a]
CVMT	1.56	0.65	1.31, 1.82*	.90[a]
WM	0.00	0.77	-0.31, 0.29	NA
L2 Pattern	69.07	20.22	61.07, 77.07*	.65[a]
L2 Analogistic	34.20	23.47	24.91, 43.48	.32[a]

Note: 95% CI = 95% confidence interval; *above-chance learning effect; [a]Cronbach’s alpha; [b]Spearman-Brown split-half coefficient

To address the second research question regarding convergent validity, DO DIFFERENT PROCEDURAL MEMORY ASSESSMENTS CORRELATE WITH EACH OTHER?, Spearman correlation coefficients were calculated to examine correlations among the procedural memory assessments (Table 2). Significant positive correlations would be expected if there is convergent validity. Cohen’s (1992) guidelines for interpreting the effect size of the correlation coefficient were used, where r = .10 is a small effect size, r = .30 is a medium effect size, and r = .50 is a large effect size. The WPT evidenced a medium negative correlation with the ASRT, which trended towards significance at p = .05. No other significant correlations were observed. Taken together, these results do not provide evidence for convergent validity because none of the tasks showed statistically significant positive correlations with each other.

To address the third research question regarding discriminant validity, DO PROCEDURAL MEMORY ASSESSMENTS NOT CORRELATE WITH ASSESSMENTS OF OTHER COGNITIVE ABILITIES?, Spearman correlations were examined for each procedural memory assessment against the MLAT, CVMT, and WM composite score (Table 2). Nonsignificant or negative correlations would be expected if there is discriminant validity. No significant correlations were observed, but two trends are notable: (a) Both the ASRT and TOL showed trending negative correlations with WM (p = .07 and p = .10, respectively), indicating that superior ASRT or TOL performance may be associated with inferior WM capacity. (b) The correlation between the WPT and CVMT was positive and approximately medium in size, although not significant. This may have to do with the visual-spatial processing overlap in both the WPT and CVMT, but future research should confirm this

Table 2: Validity of procedural memory assessments

	WPT	ASRT	TOL	MLAT	CVMT	WM
WPT	-					
ASRT	-.37^	-				
TOL	.12	-.03	-			
MLAT	.06	-.29	-.19	-		
CVMT	.23	.05	-.06	.38^	-	
WM	.05	-.36^	-.33	.23	.01	-
L2 Pattern	.47*	.04	-.12	.23	.44*	.08
L2 Analogistic	-.18	-.08	.00	-.28	-.09	-.07

Note: Spearman correlations; * $p < .05$; ^ $p < .10$

speculation.[2] Taken together, the nonsignificant, and in some cases, trending towards negative correlations suggest that the procedural memory assessments show some discriminant validity from declarative and working memory.

To answer the fourth research question regarding predictive validity, WHICH PROCEDURAL MEMORY ASSESSMENTS ARE POSITIVELY CORRELATED WITH L2 GRAMMATICAL LEARNING ABILITY?, Spearman correlations between each procedural memory assessment and accuracy on the L2 pattern word formation rule were examined (Table 2). Significant positive correlations would be expected if there is predictive validity. Results indicated that only performance on the WPT correlated positively with accuracy on the L2 pattern rule, showing a significant medium (almost large) positive association.

Discussion

The present study examined the reliability and validity of procedural memory assessments that have been used to predict outcomes in L2 acquisition. Regarding reliability, it was discovered that the WPT and ASRT showed acceptable levels of reliability, but the TOL did not. Additionally, both L2 measures had low reliability, indicating that future research should improve the reliability of the L2 and TOL tasks or choose more reliable tasks as replacements. There was evidence for discriminant validity, but no evidence for convergent validity and only the WPT showed predictive validity with the L2 pattern rule.

Regarding the unexpected absence of convergent and predictive validity for the procedural memory tasks, there are a few considerations to mention. First, much of the previous empirical work on the role of procedural memory in L2 acquisition has used composite scores from multiple tasks to assess procedural memory (e.g., Faretta-Stutenberg & Morgan-Short, 2018; Morgan-Short et al., 2014). As such, it is reasonable to ask whether using composite measures of procedural memory would make a difference with respect to the observed pattern of findings. Second, it was noted in the Introduction that current theory on procedural memory regards performance on procedural tasks as encapsulated, in the sense that learning one task in procedural memory does not necessarily transfer to performance on another task that also requires procedural memory. Although this is traditionally regarded as encapsulation of *performance* on these tasks, and not encapsulation of the *ability* to learn a task in procedural memory, it may be that the learning abilities in procedural memory are themselves independent from each other. For example, it may be that learning serial reaction tasks (e.g., the ASRT) is encapsulated, or separated from, learning probabilistic classification tasks (e.g., the WPT). Under this view, (a) convergent validity would not be expected, and (b) predictive validity might differ for the different procedural memory tasks. Both of these predictions would be consistent with the results from this study. Third, a somewhat different approach is to consider the procedural memory tasks used in this study as drawing on complex combinations of different types of memory. As in the previous suggestion, this view would obviate any expectation of convergent validity and would imply differing levels of predictive validity because the tasks would all use slightly different combinations of memory systems. As noted in the Introduction, the Tower of London was originally developed as a planning task to assess executive functions, so using it to assess procedural memory admits the task's potential for drawing on multiple cognitive systems. Again, the results of the current study would be consistent with a model of these procedural memory tasks that viewed them as each drawing on different combinations of memory systems, and thus not being strongly associated with each other.

Future research should conduct a larger study to enable more sophisticated analyses, such as latent variable analysis, in order to elucidate the underlying constructs that these procedural memory assessments are measuring and seek out the best task or set of tasks to measure the role of procedural memory learning abilities in L2 acquisition. Researchers should also continue to carefully motivate the choice of a particular procedural memory task based on supporting evidence from previous literature. The current findings

[2] Morgan-Short et al. (2014) observed a similar pattern between the WPT and CVMT ($r = .48$, $p < .10$ in their study). A larger sample size may confirm the significance of this correlation.

indicate that the measurement of procedural memory is a nontrivial issue and more research is needed to place claims regarding the role of procedural memory in L2 acquisition on solid methodological ground.

Acknowledgments

We thank Deszo Nemeth and Karolina Janacsek for sharing the ASRT task with us, as well as members of the Cognition of Second Language Acquisition Laboratory, Cognitive Brown Bag at the University of Illinois at Chicago, and members of the Psychonomic Society for valuable feedback.

References

Antoniou, M., Ettlinger, M., & Wong, P. C. M. (2016). Complexity, training paradigm design, and the contribution of memory subsystems to grammar learning. *PLOS One, 11*(7), e0158812.

Beauchamp, M., Dagher, A., Aston, J., & Doyon, J. (2003). Dynamic functional changes associated with cognitive skill learning of an adapted version of the Tower of London task. *NeuroImage, 20*(3), 1649-1660.

Buffington, J., & Morgan-Short, K. (in press). Declarative and procedural memory as individual differences in second language aptitude. In Z. E. Wen, P. Skehan, A. Biedron, S. Li & R. Sparks (Eds.), *Language aptitude: Recent advances and emerging trends*. Routledge.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test*. San Antonio, TX: Psychological Corporation.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed.). NY: Routledge.

Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review, 19*(1), 42-68.

Eichenbaum, H. (2011). *The cognitive neuroscience of memory: An introduction*. New York: Oxford University Press.

Ettlinger, M., Bradlow, A. R., & Wong, P. C. M. (2014). Variability in the learning of complex morphophonology. *Applied Psycholinguistics, 35*(4), 807-831.

Faretta-Stutenberg, M., & Morgan-Short, K. (2018). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research, 34*(1), 67-101.

Fletcher, P. C., Zafiris, O., Frith, C. D., Honey, R. A. E., Corlett, P. R., Zilles, K., & Fink, G. R. (2005). On the benefits of not trying: Brain activity and connectivity reflecting the interactions of explicit and implicit sequence learning. *Cerebral Cortex, 15*(7), 1002-1015.

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences of the United States of America, 103*(31), 11778-11783.

Hamrick, P. (2015). Declarative and procedural memory abilities as individual differences in incidental language learning. *Learning and Individual Differences, 44*, 9-15.

Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences*. Advance online publication.

Howard, J. H., & Howard, D. V. (1997). Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and Aging, 12*(4), 634-656.

Kaller, C. P., Unterrainer, J. M., & Stahl, C. (2012). Assessing planning ability with the Tower of London task: Psychometric properties of a structurally balanced problem set. *Psychological Assessment, 24*(1), 46-53.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*(2), 202-220.

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K., Carpenter, H., & Wong, P. C. M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition., 17*(1), 56-72.

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods, 47*(4), 1343-1355.

Ouellet, M., Beauchamp, M. H., Owen, A. M., & Doyon, J. (2004). Acquiring a cognitive skill with a new repeating version of the Tower of London task. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale, 58*(4), 272-288.

Owen, A.M., James, M., Leigh, P.N., Summers, B.A., Marsden, C.D., Quinn, N. P., . . . Robbins, T.W. (1992). Fronto-striatal cognitive deficits at different stages of Parkinson's disease. *Brain, 115*(6), 1727-1751.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Philadelphia, PA: John Benjamins Publishing Company.

Rauch, S. L., Whalen, P. J., Savage, C. R., Curran, T., Kendrick, A., Brown, H. D., . . . Rosen, B. R. (1997). Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. *Human Brain Mapping, 5*(2), 124-132.

Trahan, D. E., & Larrabee, G. J. (1988). *Continuous visual memory test*. Odessa, FL: Psychological Assessment Resources.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*, 231-270.

Ullman, M. T. (2015). The declarative/procedural model: A neurobiologically-motivated theory of first and second language. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed.). NY: Routledge.