Title
The Cognitive Underpinnings of Inductive Grammar Learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors
Martinez, David
Tseng, Alison
Karuzis, Valarie
et al.

Publication Date
2019

Peer reviewed

The Cognitive Underpinnings of Inductive Grammar Learning

David Martinez (dmartin5@umd.edu)
Alison Tseng (atseng1@umd.edu)
Valarie Karuzis (vkaruzis@umd.edu)
Meredith Mislevy-Hughes (mmislevy@umd.edu)
Nick B. Pandža (npandza@umd.edu)
Gregory J. H. Colflesh (colflesh@gmail.com)
Polly O'Rourke (porourke@umd.edu)
University of Maryland, College Park, 7005 52nd Ave, College Park, MD 20742

Abstract

The acquisition of the grammar of a second language requires a variety of cognitive mechanisms, including inductive reasoning. In the current study, we examine the cognitive underpinnings of grammar learning with an explicit-inductive (rule search) learning task, designed to capture more of the complexity associated with grammar learning than purely deductive tasks. Research in language aptitude has shown that working memory capacity (WMC) is a key predictor of grammar learning outcomes. Inductive reasoning and grammatical sensitivity are other established aptitude factors. The goal of the present study was to determine the degree to which relevant variables predict learning on an explicit-inductive grammar learning task. Our results indicate that both WMC and inductive reasoning ability predict learning over three days of grammar training.

Keywords: L2 learning; L2 aptitude; working memory capacity; inductive reasoning; individual differences

Introduction

The acquisition of second language (L2) grammar is extremely challenging for adult learners. One of the reasons for this difficulty is the heavy and diverse processing demands associated with learning grammatical rules as well as applying them during comprehension and production of L2 utterances (Doughty & Long, 2003). Insights into the precise cognitive underpinnings of grammar learning come from the field of language aptitude. Importantly, working memory capacity (WMC) has emerged as a key predictor of L2 grammar learning ability (Linck, Osthus, Koeth & Bunting, 2014; Miyake & Friedman, 1998). WMC is defined as the ability to maintain attention on a limited amount of information, even in the face of interference (Engle, 2002, 2018), and underpins many aspects of higher cognition and goal-directed behavior. Another predictor is inductive reasoning ability, the ability to extrapolate rules and patterns from specific examples. While both WMC and inductive reasoning are predictors of grammar learning outcomes, there is a lack of research examining whether the two account for independent portions of variance in learning. In the current study, we examined the cognitive underpinnings of grammar learning using an explicit-inductive learning task. In this task, participants were presented with L2 phrases and asked to figure out the grammatical rules, and then tested on those rules. The goal was to examine the degree to which relevant variables predict grammar learning.

Explicit-Inductive Grammar Learning

In explicit-inductive (or "rule-search") grammar learning tasks, learners are presented with a number of L2 examples (sentences or phrases) exhibiting target grammatical structures in both a foreign language and the individuals' native language and are asked to figure out the rule(s) for subsequent testing. These tasks differ from deductive tasks in which rules are explicitly taught (DeKeyser, 2003). They also differ from artificial grammar learning tasks (also referred to as statistical learning tasks) in that, in artificial grammar learning tasks, rules are acquired without conscious awareness (i.e., implicitly) and there is no meaning ascribed to the material under study (Misyak & Christiansen, 2012). Though, it is very likely that in providing numerous exemplars in explicit-inductive grammar learning tasks, individuals not only infer rules but likely implicitly acquire statistical regularities as well. Thus, explicit-inductive grammar learning tasks likely involve *both* explicit and implicit learning processes (DeKeyser, 1995). Given that both types of learning are known to be involved in grammar acquisition (DeKeyser, 2003), these tasks may better capture the cognitive complexity of grammar learning.

Working Memory Capacity

Individual differences in WMC are strongly predictive of performance on a range of tasks assessing cognitive abilities and processes (Engle, Tuholski, Laughlin, & Conway, 1999; Kyllonen & Christal, 1990) including L1 processing (Daneman & Merikle, 1996) and L2 learning (Linck et al., 2014). Indeed, in a meta-analysis synthesizing the results of 79 studies with a combined sample size of over 3,000 participants, Linck et al. (2014) found that WMC tasks are positively associated with L2 outcomes. Moreover, Tagarelli, Borges-Mota and Rebuschat (2011) found that WMC predicted performance on an explicit-inductive grammar learning task.

Relations observed between WMC and other cognitive tasks are typically explained as owing to the fact that complex cognition requires sustained attention on the task at hand, often while performing various operations, which themselves produce interference (Daneman & Carpenter, 1980;

Daneman & Merikle, 1996). This emphasis on controlled attention has led some to theorize that WMC plays a greater role in L2 learning under explicit, rather than implicit, learning conditions (e.g., Tagarelli, Mota, & Rebuschat, 2011). Indeed, Tagarelli, Mota, and Rebuschat (2011) found that WMC was predictive of learning under explicit, but not implicit, learning conditions.

Inductive Reasoning

Another predictor of L2 learning that figures prominently is inductive reasoning (Gardner & Lambert, 1965; Sparks, Humbach, Patton, & Ganschow, 2011). In inductive reasoning, one infers general principles from specific observations. For example, an adult interested in learning another language for use during a trip may begin by learning "survival phrases" such as "I am American" and "I am sorry." In doing so, one may *infer* grammatical rules and the meaning of certain words in the L2, which can then be used to construct new words and sentences (though of course the accuracy of the constructions will be dependent on the premises, e.g., not all verbs in English can be changed from present to past tense by affixing an *-ed*). Like WMC, inductive reasoning ability has been found to predict grammar learning under explicit, but not, implicit conditions (Gebauer & Mackintosh, 2007).

Relationship between WMC and Inductive Reasoning

An issue arises, however, when one notes that WMC is highly correlated with inductive reasoning ability (Engle et al., 1999; Kyllonen & Christal, 1990). In the individual differences literature, inductive reasoning tasks are often used as indicators of fluid intelligence (Marshalek, Lohman, & Snow, 1983; Wilhelm, 2005). According to a recently proposed theory (see Shipstead, Harrison, & Engle, 2016), WMC and fluid intelligence/inductive reasoning are highly correlated because both rely on attention control; however, while WMC tasks primarily assess the ability to *maintain* attention, fluid intelligence/inductive reasoning tasks additionally assess the ability to *disengage* attention. Consider that in WMC tasks, the goal is to maintain to-be-remembered information (e.g., sets of letters) in mind exactly as they were presented; in inferential tasks, however, the goal is to produce a novel solution, entailing some kind of transformation or restructuring of inputs as multiple solutions or hypotheses are investigated (Oberauer, Süß, Wilhelm, & Sander, 2007). During the reasoning process, one has to maintain relevant pieces of information in mind, implicating WMC, but at other times, one has to abandon an incorrect solution and begin anew, requiring one to disengage attention from one problem representation for another.

The issue is that there is little research investigating whether WMC and inductive reasoning ability *independently* account for variance in L2 learning. To investigate this issue, we included a measure of inductive reasoning and a measure of WMC as predictors in the present study. Given that the outcome variable is an explicit-inductive grammar learning task, we expect inductive reasoning to be predictive of learning, however, a WMC task should account for variance over and above an inductive task, as WMC is a well-established predictor of language learning (Linck, Osthus, Koeth & Bunting, 2014; Miyake & Friedman, 1998).

Grammatical Sensitivity

In addition to WMC and inductive reasoning, a measure of grammatical sensitivity was also included in this study—the Words in Sentences (WIS) subtest from the Modern Language Aptitude Test (MLAT), developed by Carroll and Sapon (1959). According to Carroll (1964), grammatical sensitivity is the "ability to recognize the grammatical functions of words in sentences (p. 95)". Studies have shown the WIS to be a predictor of L2 learning (Li, 2015), particularly under explicit learning conditions (Li, 2014); however, there is also research and theorizing that grammatical sensitivity depends on inductive reasoning (Li, 2015; Sasaki, 1993). Thus, including this measure as a predictor allows us to investigate whether grammatical sensitivity influences novel grammar learning over and above inductive reasoning and working memory.

The Present Study

With the above in mind, this study was undertaken to assess the relative contributions of WMC, inductive reasoning, and grammatical sensitivity on one aspect of L2 learning, grammar learning. For this study, we developed an explicit-inductive (i.e., rule-search) grammar learning task in which individuals were tasked with learning syntactic rules in an L2. During learning, participants were exposed to a number of L2 phrases and their English translations and attempted to infer rules for arranging words in the L2. Superficially, the task is similar to what was described earlier when one learns "survival" phrases and induces rules, however, (and as will be clarified below) this task obviates the need to memorize phrases and thus should be a relatively *pure* measure of grammatical (rule) induction.

Method

Participants

A total of 34 individuals participated in the study; however, three did not complete the entire study, leaving 31 with complete data. All participants were recruited from the university and surrounding community and were compensated for their time. No participant reported experience with Indonesian or related languages.

Procedure

All participants completed a total of three sessions in a room with up to six other participants. Each session contained a grammar learning task. In addition to the grammar learning task, in Session 1 participants also completed a demographics and language history questionnaire, administered before the grammar learning task; in Session 2 participants completed

Letter Sets, an Antisaccade task, and a Speeded Lexical Decision task, administered, in that order, after the grammar learning task; and in Session 3 participants completed the Remember and Count task, another Speeded Lexical Decision task, and the Words in Sentences, administered, in that order, after the grammar learning task. Each session took approximately 60 minutes to complete. All tasks were completed on desktop computers. Below, we offer descriptions of the tasks included in this study.

Instruments

Explicit-Inductive Grammar Learning Task We chose an explicit-inductive task as our criterion measure because research indicates that during the early stages of L2 learning, adults tend to engage control processes to learn grammatical rules, while at later stages, they tend to rely on implicit learning processes (Hamrick, Lum, & Ullman, 2018). Thus an explicit-inductive task likely captures processes similar to those engaged throughout the learning process.

The grammar learning task consisted of three phases: learning, recall, and recognition. During the learning phase, participants were presented with example Indonesian phrases and their English translations, ordered from short/simple phrases to long/complex phrases. Participants therefore had to infer the "simple" rules and then take mental note of how these "simple" rules combined to construct complex phrases. Because participants did not know Indonesian, the Indonesian words and their English translations were color-coded, such that translation equivalents were the same color; Indonesian words without direct translations (e.g., classifiers) were presented in black font with no background color. See Figure 1.

Figure 1. Three example grammar learning items.

During the recall phase, participants were asked to translate English phrases into Indonesian by selecting words from a word bank and placing them in the correct sequence. The word bank included the words needed for the translation as well as all function words. Where possible, English translations were provided (see Figure 2). During the recognition phase, participants saw Indonesian noun phrases and indicated whether they were grammatical or not (see Figure 3).

Participants were *never* given feedback in either the recall or recognition phases, but were given a score of their overall recognition phase performance at the end of the day.

Table 1. Syntactic Structures in the Grammar Learning Task.

	Structure	Example English Phrase
1	Demonstrative noun	that uncle
2	Number w/classifier	two apes
3	Single adjective	moody zebra
4	Demonstrative + single adjective	that bold scientist
5	Double adjective	new, red school
6	Possessive	my rabbit
7	Possessive + single adjective	my hungry uncle
8	Number/classifier + single adjective	two small warehouses
9	Number/classifier + single adjective + possessive	my two friendly fish
10	Number/classifier + double adjective + possessive	my two new, crowded stores
11	Triple adjective	fancy, young, skilled lawyer
12	Noun + single adjective + pre-intensifier	very crowded arena
13	Noun + single adjective + post-intensifier	very skilled teacher
14	Noun + number/classifier + single adjective + pre-intensifier	two very clever bears
15	Noun + number/classifier + single adjective + post-intensifier	two very chilly cinemas
16	Noun + possessive + number/classifier + single adjective + pre-intensifier	my two very expensive palaces
17	Noun + possessive + number/classifier + single adjective + post-intensifier	my two very tired bears

Figure 2. Example grammar recall item.

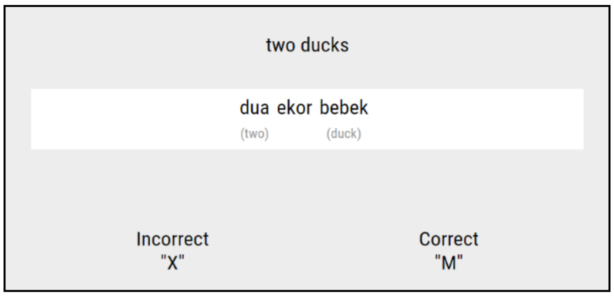

Figure 3. Example grammar recognition item. The correct answer is "M", correct.

Across sessions, the same 111 noun phrases were used in the learning phase. However, because our interest was in *grammar* learning, noun phrases used in the recall and recognition phases were never repeated across sessions, resulting in 123 noun phrases for the recall phase (41/session) and 312 noun phrases for the recognition phase (104/session).

Words in Sentences (WIS) In the WIS, each item consisted of two or more English sentences. One word in the first sentence was printed in uppercase letters. Four or five words in the remaining sentences were underlined and were labeled with corresponding letter answer options (Figure 3). Participants indicated which of these underlined words served the same grammatical function as the word in uppercase letters in the first sentence.

Chidi said THAT Eleanor is incorrigible.

In our (A) session, that (B) lady said that (C) she knew that (D) actor on the (E) television show.

Figure 3. A sample item from the *words in sentences* test modified from https://lltf.net/mlat-sample-items/mlat-part-iv/. The correct answer is C.

Remember and Count (RAC) WMC was assessed by the RAC task (Hughes et al., 2016; O'Rourke et al., 2017), a visuospatial complex span task. In the RAC task, participants first see a sequence of triangles of different colors presented in a sequence in different quadrants. Next, they see an image of dark and light blue circles and squares; participants are to count and report the number of dark blue circles in the image. Finally, in the critical portion of the task, they must recall the sequence of triangles by indicating the color, the location, and the order of each triangle in the sequence. The number of triangles in a sequence varied between three and five, with four trials of set size 3, nine trials at set size 4, and eight trials at set size 5, for a total of 21 trials. Each trial was scored as a proportion of correctly recalled triangles; thus, participants could achieve a maximum of 1 point per trial.

Letter Sets (LSET) Inductive reasoning was assessed by the LSET task (Doughty, Campbell, Bunting, Bowles, & Haarmann, 2007). In each item, participants are presented with five sets of four letters. Four of the sets are arranged such that they follow the same rule while one does not; participants are to determine which set of letters does not follow the same rule as the others. There were a total of 15 items.

Results

Correlations amongst the predictors and descriptive statistics are provided in Table 2. Figure 4 depicts average learning curves for both the recall and recognition grammar measures.

Table 2. Predictor Correlations and Descriptive Statistics

	WIS	LSET	RAC
WIS			
LSET	0.20		
RAC	0.48*	0.48*	
$\bar{X}$	.42	.76	.53
SD	.13	.13	.18
Skew	-.02	-.10	-.91
Kurtosis	-.54	.16	.19

Note: * *p* <.05; LSET = letter sets; RAC = remember and count; WIS = words in sentences.

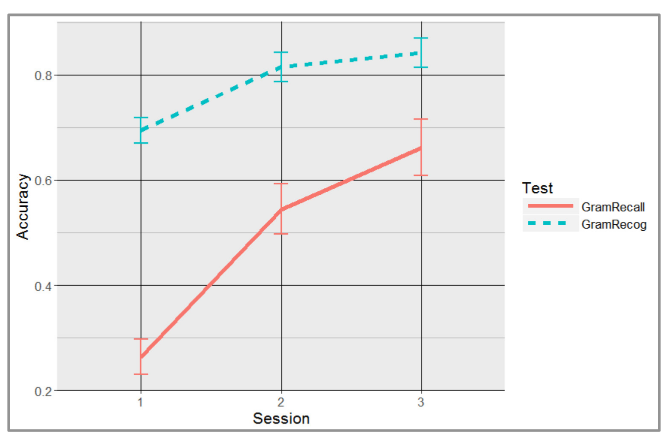

Figure 4. Average Learning Curves. Error bars: ± 1 *SE*.

Next, the recall and recognition data were each submitted to mixed-effects logistic regression modeling using a forward-testing procedure for random slopes and a backward elimination procedure for fixed effects to arrive at the model of best fit using likelihood ratio tests. This procedure allows us to find the most maximal model supported by the data, balancing Type I error and power (Matuschek et al., 2015). For each analysis, the first model included items and participants as random intercepts, session as a fixed effect, WIS, LSET, and RAC as covariates, and each Session x Covariate interaction. In each model, covariate tasks were mean centered.

Table 3 displays the final recall model. There was a significant effect of session, indicating that performance significantly improved across sessions ($b = 1.67$, $SE = .25$, $z = 7.97$, $p < .001$). There was also a significant effect of Session x RAC ($b = .58$, $SE = .21$, $z = 2.74$, $p = .006$), indicating that participants with higher RAC scores showed greater gains in recall accuracy over sessions. No other covariates or interactions were significant.

Table 3. Final Recall Model

Fixed Effects	*b*	Odds	*SE*	*p*
Intercept	-3.15	0.04	0.33	<.001*
Session	1.67	5.31	0.25	<.001*
RAC	0.06	1.06	0.16	.825
Session x RAC	0.58	1.79	0.21	.006*
Random Effects	**Var**	***SD***	**Corr**	
Intercept | Item	1.90	1.38		
Intercept | Part.	1.26	1.12		
Session | Part.	1.14	1.07	-.45	

Table 4. Final Recognition Model

Fixed Effects	*b*	Odds	*SE*	*p*
Intercept	0.24	1.27	0.15	.116
Session	0.89	2.43	0.10	<.001*
WIS	0.10	1.11	0.14	.472
LSET	-0.14	0.87	0.14	.304
RAC	0.05	1.05	0.16	.744
Session x LSET	0.23	1.26	0.10	.027*
Session x RAC	0.27	1.31	0.11	.014*

Random Effects	Var	*SD*	Corr		
Intercept | Item	.77	.88			
Session | Item	.07	.27	-.26		
LSET | Item	.04	.21	.00	.68	
WIS | Item	.09	.30	.45	-.93	-.81
Intercept | Part.	.27	.52			
Session | Part.	.20	.44	.12		

Table 4 displays the final model for the recognition analysis. Session was once again significant ($b = .89$, $SE = .10$, $z = 8.83$, $p < .001$), as was Session x LSET ($b = .23$, $SE = .10$, $z = 2.21$, $p = .027$) and Session x RAC ($b = .27$, $SE = .11$, $z = 2.46$, $p = .014$), indicating that as performance on these measures increased, individuals showed greater gains in accuracy over sessions. No other covariate or interaction was predictive.

Discussion

The primary aim of the study was to investigate the cognitive underpinnings of explicit-inductive grammar learning. In our grammar learning task, participants attempted to learn a subset of Indonesian syntax by inferring the rules of the language from a number of exemplars. Based on prior studies and theory, we chose indicators of grammatical sensitivity, WMC, and inductive reasoning as our predictors. Aware of significant relationships amongst the predictors, we were also interested in investigating whether the predictors uniquely accounted for variance in grammar learning and if so, to what degree. Our analyses indicated that the WMC measure, RAC, and the inductive reasoning measure, LSET, were significantly related to grammar learning, however the grammatical sensitivity measure, the WIS, was not. Moreover, logistic mixed-effects modeling indicated that for our recall measure, performance on our WMC measure interacted with session, such that individuals who performed better on RAC showed greater gains in accuracy over sessions. A similar result was obtained for the recognition measure; however, additional variance in learning performance was accounted for by an interaction between the inductive reasoning measure, LSET, and session.

Overall, the results of the study suggest that WMC and inductive reasoning facilitate grammar learning. The fact the predictors interacted with learning session is likely due to the fact that grammatical learning builds on previous learning. For example, in English, it would be difficult for one to generate, “my two very fancy goats” without also being able to correctly generate “my two goats” or “my fancy goat.” Individuals with greater WMC and inductive reasoning ability were likely more able to learn rules, build upon them, and reinforce their own learning as they learned more complex rules, increasing their learning rate. Individuals with lower abilities, however, may have found it difficult to learn even the simpler rules and therefore struggled to see recurring patterns in more complex sentences, possibly interfering with (rather than reinforcing) learning; thus, their learning rate was slower compared to higher-ability individuals.

While the Session x RAC interaction was a significant predictor of both the recall and recognition measures, it is important to note that the Session x LSET interaction only accounted for a significant proportion of variance in recognition performance. This pattern of results confirms that WMC was generally involved in learning, however, the role of inductive reasoning is somewhat ambiguous. One possibility is that individuals with greater inductive reasoning ability were able to infer more rules but not necessarily retain accurate representations in long-term memory (a function supported by WMC; Unsworth & Engle, 2007) and thus they were unable to accurately retrieve rules during the recall test. When tested using a recognition paradigm, however, high ability individuals were able to use cues to "fill in" or *redintegrate* their partial representations, and thus were more likely to correctly choose the grammatical phrase. Future research should continue investigating the role of inductive intelligence in explicit-inductive grammar learning.

Despite the fact that the grammatical sensitivity measure, WIS, did not account for variance in the learning tasks above and beyond the other predictors, the results of this study should not be interpreted as suggesting that grammatical sensitivity does not play a role in L2 learning. As noted above, prior research indicates that grammatical sensitivity is related to L2 learning and, in fact, the coefficients observed between the WIS and the grammar learning measures are similar to what have been found in the literature (Li, 2015). The null result observed here may have been due to sample size or characteristics (e.g., a restricted range in performance). Still, to the extent that the estimates are accurate, it is interesting to note that the role of grammatical sensitivity in grammar learning appears to be smaller than that of WMC and inductive reasoning. This may be because grammatical sensitivity is more a measure of English grammatical knowledge than learning (Carroll, 1993).

With the above limitations in mind, this study corroborates prior research indicating that WMC is a robust predictor of L2 learning and, more specifically, L2 grammar learning (Linck, Osthus, Koeth & Bunting, 2014; Miyake & Friedman, 1998). Moreover, while it may be somewhat intuitive that inductive reasoning is predictive of explicit-inductive grammar learning, we found that inductive reasoning accounts for at least one measure of grammar learning above and beyond WMC. Considering the large number of individuals that engage in L2 learning and the significance of knowing an L2, researchers should continue investigating the cognitive components of L2 learning.

References

Carroll, J. B. (1964). The Prediction of Success in Intensive Foreign Language Training. *Reprint from Training Research and Education, Chapter 4, p 87-136*: University of Pittsburgh Press, 1962.

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

Carroll, J. B., & Sapon, S. M. (1959). Modern language aptitude test. *New York: The Psychological Corp.*

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466. doi:http://dx.doi.org/10.1016/S0022-5371(80)90312-6

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

DeCaro, M. S., Van Stockum, C. A., & Wieth, M. B. (2016). When working memory capacity hinders insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 39–49. doi:10.1037/xlm0000152

DeKeyser, R. (2003). "Implicit and Explicit Learning." In C. J. Doughty & M. H. Long (Eds.),*The handbook of second language acquisition* (Vol. 27) (pp. 313-348). Malden, MA: Blackwell.

DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in second language acquisition*, *17*(3), 379-410.

Doughty, C.J., Campbell, S.G., Bunting, M.F., Bowles, A., & Haarmann, H. (2007). *The development of the High-Level Language Aptitude Battery*. College Park, MD: University of Maryland Center for Advanced Study of Language.

Doughty, C. J., & Long, M. H. (Eds.). (2008). *The handbook of second language acquisition* (Vol. 27). John Wiley & Sons.

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

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

Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 34-54.

Hamrick, P., Lum, J. A., & 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*, 115(7), 1487-1492.

Hughes, M. M., Karuzis, V. P., Kim, S., O'Rourke, P., Sumer, A., Liter, A., … Campbell, S. G. (2016). *Assessing aptitude for USAF cyber warfare operations training: Interim results from field testing*. College Park, MD: University of Maryland Center for Advanced Study of Language.

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

Li, S. (2014). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, *36*(3), 385-408.

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

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

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.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2015). Balancing type I error and power in linear mixed models. arXiv preprint arXiv:1511.01864. *Journal of Memory and Language*, accepted pending minor revisions.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302-331.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy, L. E. Bourne, Jr., A. F. Healy, & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention.* (pp. 339-364). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Oberauer, K., Süß, H. M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. *Variation in working memory*, 49-75.

O'Rourke, P., Karuzis, V. P., Kim, S., Tseng, A., Pandža, N. B., Young, J. M., … Campbell, S. G. (2017). *Assessing aptitude for USAF cyber warfare operations training: Test specifications for USAF-CATA*. College Park, MD: University of Maryland Center for Advanced Study of Language.

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition, 19*(2), 223-247. Retrieved from http://www.jstor.org/stable/44488684

Peterson, C. R., & Al-Haik, A. R. (1976). The development of the Defense Language Aptitude Battery (DLAB). *Educational and psychological measurement, 36*(2), 369-380. doi:10.1177/001316447603600216

Sasaki, M. (1993), Relationships Among Second Language Proficiency, Foreign Language Aptitude, and Intelligence: A Structural Equation Modeling Approach. *Language Learning,* 43: 313-344. doi:10.1111/j.1467-1770.1993.tb00617.x

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

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

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

Tagarelli, K. M., Mota, M. B., & Rebuschat, P. (2011, January). The role of working memory in implicit and explicit language learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological review*, *114*(1), 104.

Wilhelm, O. (2005). Measuring reasoning ability. In O.W. Wilhelm O.W. & R.W. Engle (Eds.), *Handbook of understanding and measuring intelligence*, (pp. 373-392). Thousand Oaks: Sage.