Possibilities and Caveats of Implicit Language Aptitude Measurements

Yinjie Tang*, Shaoqian Luo

School of Foreign Languages and Literature, Beijing Normal University, China

Abstract

This paper probes into implicit language aptitude (LA), a set of cognitive abilities that showcase learners' unconscious, automatic second language (L2) attainment. Implicit LA is becoming an important part of aptitude research, and advances the understanding of L2 learning mechanism. This paper critiques the construct of implicit LA and proposes that implicit learning ability and implicit memory ability are two key components of implicit LA. We then explore the measurements of implicit LA. We raise a number of caveats derived from the validation of these measurements, and discuss how to circumvent these issues. One apparent reason for these caveats is the lack of empirical research that focuses exclusively on implicit LA, and thus, aptitude researchers are encouraged not only to include implicit LA measurements in their test batteries, but also to address the connection between the construct and its measurements.

Keywords: *Language Aptitude, Implicit Learning, Implicit Measurements*

Introduction

Exploring factors that contribute to the success of second language (L2) learning has become a key issue in the realm of Second Language Acquisition (SLA). Factors endowed with some degree of predictive power, include LA, motivation (Kim & Kim, 2021) and language anxiety

(MacIntyre & Wang, 2022), are thus attract researchers' attention. Among these factors labeled as individual differences (IDs), LA has been proven to be the most predictive variable for L2 achievement (Li, 2016; Li, 2021; Wen et al., 2019). This variable can be perceived as one's language learning talents (Carroll, 1990), by which one can learn foreign languages faster and easier (Li & Luo, 2019). LA can be used to predict and explain L2 learners' foreign language learning attainment given proper instruction (Wen et al., 2022), and can provide information to government agencies or institutes to select candidates that are qualified for high-demanding linguistic programs (Linck et al., 2013).

Serving as the cornerstone of LA research, the Carrollian four-factor aptitude construct (Carroll, 1962) has exerted a far-reaching influence on many contemporary LA tests since the 1960s. However, this construct has been proven to be more strongly correlated with explicit than implicit learning conditions (Li, 2017), which is manifested by various subtests in well-established LA batteries, including the Modern Language Aptitude Test (MLAT; for a detailed discussion, see Skehan, 2022). This limitation could be compensated by the inclusion of implicit LA, a set of cognitive abilities that manifest implicit L2 learning. Implicit LA is defined as "a cluster of cognitive abilities" (Li & DeKeyser, 2021) that entails unconscious processing of linguistic elements (Godfroid & Kim, 2021). It is worth noting that implicit learning is the ultimate goal of language learning (Zhang & Chen, 2020). If an L2 learner is detected to have a high level of implicit LA, s/he may encode, maintain, and retrieve (Squire, 1987) linguistic patterns more effortlessly (Granena, 2020), which undoubtedly saves time by allowing him/her to process L2 material automatically.

In this paper, though the focus is on the measurement, or "tools" of implicit LA, we start by examining this construct, which is fundamental in language testing (Luo et al., 2021; Messick, 1989). We first provide a detailed illustration of some representative measurements, particularly the Serial Reaction Time (SRT) task and the LLAMA_D, two widely used tasks in the implicit LA paradigm. Then, we address some caveats concerning the validation of implicit LA measurements and provide suggestions to alleviate potential problems.

The Construct of Implicit LA

Unlike traditional LA that exhibits a fixed construct involving phonetic coding, language analytic ability, and rote memory (Li, 2022), implicit LA is unstable in its nature (Li & DeKeyser, 2021). This might be attributed to the fact that research on implicit LA is just beginning and much of its rubric remains unknown (Li & Zhao, 2021). However, studies have shown that some attributes of implicit LA are known and provide a blueprint for further research. First, implicit LA, although not yet found to be a unitary construct, separates from its explicit counterpart (Li & Qian, 2021). That is, learners' performance on the two LAs has been found to be uncoorelated (Li & Qian,

2021). Second, implicit LA can be found in adult language learners and remains static as one ages (Long, 2017; Ullman & Lovelett, 2018), challenging the traditional hypothesis that implicit leaning functions only in children's language learning process (Bley-Vroman, 2009). Third, implicit LA might make a difference with a certain degree of predictive validity in naturalistic learning conditions, where the L2 is the primary language in the learners' surroundings (Li & Qian, 2021, Suzuki & DeKeyser, 2017).

Based on the attributes of implicit LA, three tentative constructs are worth mentioning. The first construct proposed by Li and DeKeyser (2021) is comprised of three components: sensitivity to frequency and conditional probability, priming, and selective attention (Li & DeKeyser, 2021, p. 482). The first component alludes to one's cognitive ability to learn patterns and regularities from omnipresent linguistic input. The second component, priming, is the ability to be tacitly influenced by recent events. The third component, selective attention, is the tendency that allows people to "pick up information that is behaviorally relevant and automatically ignore vast amounts of irrelevant information" (Jiang & Chun, 2001).

Although this three-component construct makes it easier for researchers to address the theme of implicit LA, it has two possible drawbacks. First, there seems to be a dissonance between components and corresponding measurements, which is also a universal problem in the current research agenda. That is to say, most promising measurements work on the sensitivity to frequency and conditional probability, including the SRT task, artificial grammar, and traditional measurements of procedural memory such as the Tower of London task (TOL; e.g., Kaller et al., 2016). For the latter two components, only the priming task (Li & Qian, 2021) firmly relates to the component of priming. There is no solid measurement for selective attention. Second, although selective attention is included in implicit LA theory, it is postulated as a prerequisite, rather than a component (Granena, 2020; Perruchet, 2008). Also, an enquiry of selective attention may bring challenges to measurement design, since it induces distractors that serve as "secondary tasks" (Li & DeKeyser, 2021, p. 486) inserted in implicit LA measurements. The test takers' performance, therefore, would encounter a loss of accuracy (Shanks et al., 2005).

The second tentative construct proposed by Granena (2020) divides implicit LA into implicit inductive learning and implicit memory (p. 14). This construct is based on three stages of memory: encoding, maintenance, and retrieval (Squire, 1987). Implicit inductive learning ability is similar to the first component of Li & DeKeyser's (2021) design, referring to learning patterns and regularities unconsciously. Inductive memory ability deals with the unconscious maintenance and retrieval of information, somewhat analogous to "procedural memory", a long-term memory system which entails implicit processing of patterns (see also Buffington & Morgan-Short, 2019; Goldstein, 2011).

This approach might be relatively persuasive in the following two ways. First, the

arrangement echoes with Skehan's (2019) proposal that links LA components to processing stages, and regards different components as parts of a unified whole (i.e., language acquisition, memory, etc.). This is a clear way for building a construct since it places various components on a continuum that is related to the L2 learning process. Second, this design has been supported by some empirical evidence (Granena, 2013a, 2019). Utilizing LLAMA, Granena (2013a) investigated the LA profiles of 186 adult foreign language learners with different L1 backgrounds. A set of exploratory principal component analyses (PCAs) was conducted and reported that LLAMA_B, E, and F loaded on the same factor, while LLAMA_D loaded on a different factor, which was named as "implicit LA". This finding was further studied in Granena (2019), who investigated 135 college-level learners' performance on LA and speaking proficiency. The predictor variable, LA, was measured by 8 cognitive tests from LLAMA and Hi-Lab; the criterion variable, speaking proficiency, was measured by an oral picture describing task in which complexity, accuracy, and fluency were taken into account. The findings revealed that the SRT and LLAMA_D loaded on two different factors. She then justified the names of the components "implicit learning" and "implicit memory" based on the properties of the two measurements. the SRT is a task for measuring automatic sequence learning (Kaufman et al., 2010; Nissen & Bullemer, 1987), and the LLAMA_D is a familiarity-based recognition task that sheds light on implicit memory (Wang & Yonelinas, 2012).

The third construct, proposed by Godfroid and Kim (2021), includes implicit learning, statistical learning, and procedural memory (p. 3). However, they amended the term "implicit LA" or "implicit aptitude" in accordance with statistical learning and introduced the term, "implicit-statistical aptitude". The merit of this design is the linkage between implicit learning and statistical learning, the latter referring to the abilities to learn patterns and regularities from the environment (Christiansen, 2019). By situating implicit aptitude and statistical aptitude at the same level, this design extends the application of LA battery and involves measurements in statistical learning that originates from cognitive psychology (Christiansen, 2019), thus helps to advance the understanding of the L2 learning mechanism (see also Rebuschat & Monaghan, 2019). However, this design may create confusion in the following ways. First, it may contradict itself by viewing procedural memory as a "neural substrate" (Godfroid & Kim, 2021, p. 21) of implicit knowledge. That is to say, procedural memory seems to become a subordinate concept or functional unit of all implicit LA types, rather than a component. Second, the juxtaposition of implicit learning and statistical learning might be questioned because the two concepts overlap under the paradigm of implicit LA. By using the term "implicit LA", statistical learning is often naturally warranted (Christiansen, 2019; Frost et al., 2019)

Despite the different classifications for the aforementioned constructs of implicit LA, they share some commonalities. First, they place emphasis on the automatic, unconscious input of

linguistic materials. Second, they confirm and situate implicit memory ability as an indispensable component of implicit LA by referring to the distinction between declarative and procedural memory (Buffinton & Morgan-Short, 2019; Ullman & Lovelett, 2018). Although Li and DeKeyser (2021) did not treat implicit memory ability as one component, they used priming instead, and priming can reflect implicit memory to some extent (Granena, 2020; Gupta & Cohen, 2002). Third, implicit LA pays attention to domain-general cognitive abilities, while traditional LA focuses on components that are more related to language learning itself (i.e., domain-specific abilities). Thus the advent of implicit LA calls for the integration between domain-general and domain-specific abilities, which has been hypothesized for a long period of time (Sparks et al., 2019). Hence, implicit LA can be roughly viewed as the combination of implicit learning (including statistical learning) and implicit memory (including procedural memory and priming), as well as the combination of domain-general and domain-specific cognitive abilities, which make this term consistent with Skehan's (2022) expectations about the future development of LA study.

Possible Measurements of Implicit LA
Although the construct of implicit LA has yet to be confirmed, there are a number of studies that have proposed measurements (e.g., Godfroid & Kim, 2021; Granena, 2020; Granena & Yilmaz, 2019; Li, 2020; Li & Zhao, 2021). Note that these measurements may not significantly correlate with each other and thus cause difficulty in categorizing them (Siegelman & Frost, 2015). Meanwhile, these measurements can be used interchangeably (Granena, 2020). For Instance, the Alternative SRT (ASRT) task could be used to measure both implicit learning (Faretta-Stutenberg & Morgen-Short, 2018) and implicit memory (Buffington et al., 2021). Hence, it is better now to traverse the major measurements of implicit LA before discussing whether it is necessary to make categorizations.

 Artificial Grammar Learning. Proposed by Reber (1967), this task is a cornerstone in implicit learning (Guo, 2003), although it is rarely used in the current implicit learning paradigm. In this task (Reber, 1967, 1993), participants are required to memorize unpronounceable and meaningless letter strings that are randomly generated, and to tell whether the strings in the testing phase correspond to the underlying grammatical rule exhibited in the study phase. From the perspective of current claims in implicit LA, this measurement may not fully exhibit implicit learning ability (Kaufman et al., 2010). It seems like a deliberate and conscious process, rather than an implicit one, because the participant is told in advance the existence of the underlying grammar rules. But still, it may have some implications for associative learning and implicit knowledge (Pothos, 2007).

 SRT. SRT is a nonverbal task that measures participants' sequence learning ability (Granena,

2019; Nissen & Bullemer, 1987). For its emphasis on the unconscious learning of patterns and regularities, this task is introduced as a robust measurement of implicit LA (Hamrick, 2015; Kaufman et al, 2010; Suzuki & DeKeyser, 2015; Walker et al., 2020). In the SRT task, participants respond (e.g., press a key) as quickly as possible to one of the four positions activated on the screen, with the stimuli being shining dots, smiling faces, or something else. The occurrence of stimuli follows either a regular sequence (accounting for 85% of all the stimuli, for example) or a control sequence (accounting for the remaining 15%). This task is scored by measuring the difference between the time lag of participants' mean reaction time (RT) for the regular and control sequence ($RT_{control}$ - $RT_{regular}$). A larger value indicates a better sequence learning ability, which is considered as an implicit learning ability (Li & Qian, 2021). In terms of stimuli-presentation modes, SRT has deterministic, probabilistic and alternating versions. In a deterministic SRT task, all except for the last one (or the penultimate) are regular sequences. For instance, there are six sequences in total, with sequence 1-5 being the mere repetition of "1-3-2-4-..." and sequence 6 being "2-3-1-4-...". Participants' performance is measured by RT_6-RT_5. In a probabilistic SRT task, the control sequence is evenly scattered and placed in a random order that mingles with the regular sequence, following a second-order condition (Kaufman et al., 2010; Li & Qian, 2021), which means that for both regular sequence and control sequence, participants can never predict what comes next depending on the prior position only. Rather, the position of every stimulus is determined by two previous stimuli. For instance, if the first stimulus appears in the position 1, then the participant can never judge what comes next, but if the participant gets two consecutive positions like "2-1", the next position is determined as 4. It would take at least 12 digits (Kaufman et al., 2010) to shape this sequence rather than 4 digits in the case of first-order condition, which considerably enhances the complicity and implicitness. By and large this arrangement would prevent participants from conscious learning which invokes explicit LA (Suzuki & DeKeryser, 2015). In an ASRT task, the control sequence is interspersed between the regular sequence, and generated in an utterly random manner (Howard & Howard, 1997; Tagarelli et al., 2016), which makes the pattern become something like "3-r-1-r-2-r-4-r-3-r-1-r-..." (r means a random position). Like probabilistic SRT, ASRT also makes it harder for test takers to consciously detect regularities.

Priming. Priming is defined as the learners' tendency to reapply a linguistic structure "due to a previous encounter with the structure" (Li & Qian, 2021, p. 1). There are two main types of priming: semantic priming and syntactic priming. An influential measure of semantic priming derives from the High-Level Language Aptitude Battery (Hi-Lab; Linck et al., 2013), where participants are required to detect synonymous words based on the pair of prime and target. In the prime phase, participants look at two words and another five words. One of the former two words is synonymous with three words in the latter five, while the other is synonymous with the

remaining two words. Participants then judge as quickly as possible which word of the two has more synonyms. In the target phase, participants are only shown a pair of words and asked to decide whether they have similar meanings. This pair can consist of words appearing in the primed phase or new words. In syntactic priming, participants also undergo a prime and a target phase. In the prime, they are asked to listen to a sentence with a certain linguistic structure, and to repeat the sentence. In the target, they are presented with a picture and describe the picture with their own words. Successful priming would be induced if sentences they say in the target phase share the same structure as the sentences presented in the prime phase. No matter what kind of priming task is used, it is typically implicit-biased (Doughty, 2019; Li & Qian, 2021) in that a) participants are unaware of what they are going to learn (especially syntactic priming, where they are told beforehand that the whole activity is a memory task), and b) they may not find any feasible strategy during the process of learning because it is rather time-limited.

LLAMA_D. LLAMA_D (Meara, 2005) is an auditory recognition task comprising a learning phase and a testing phase. Participants initially learn the novel language by listening to ten sounds in an artificial language, then the testing phase plays sounds and asks participants to judge whether these sounds appeared in the learning phase. To some extent, the task does involve implicit processing not only because the sounds are unfamiliar to all the participants, but also because of the unconscious exposure (Granena, 2019) the task provides in its learning phase. A similarity that LLAMA_D shares with syntactic priming is that they both involve auditory input and thus can predict participants' L2 learning skills. However, LLAMA_D is somewhat problematic for two reasons. First, in terms of test administration, the traditional LLAMA_D induces explicit processing (Suzuki, 2021) by letting participants know that they will hear sounds that do not occur in the learning phase. This pretest instruction makes them alert and pay full attention to the new sounds, leading to the explicit retrieval of acquired information. Second, in spite of Granena's (2019) evidence, LLAMA_D is, in its essence, still unstable concerning its preference of explicit/implicit processing. Suzuki (2021) reexamined the construct validity of LLAMA_D and stated that it involves both declarative-explicit memory and proceduralization, that is, the ability to shift from the declarative memory to the procedural memory (Skehan, 2019; Ullman, 2015). In sum, LLAMA_D, albeit being a putative measurement of implicit LA, may turn out to be an indicator of explicit LA, or both, or neither.

Therefore, LLAMA_D will need some modifications in order to measure implicit LA to the utmost extent. First, before the test begins, participants are only told to listen to the sounds in the learning phase carefully, without being aware that some new sounds will appear in the testing phase (Saito, 2017). Second, participants are allowed to adjust the volume of the recording before the test (Suzuki, 2021). By doing so they can feel less anxious and their attention can be diverted, causing the readiness for implicit learning. Third, confidence ratings can be applied to

the test phase. As soon as participants make a judgement in the testing phase, they have to evaluate their answer by choosing "confident/not confident". This step can not only detract the participants from possible conscious learning, but also make test administrator know whether or not implicit learning is dominant, in which case the confidence rating is not correlated with the accuracy (Scott & Dienes, 2010). Suzuki (2021) added confidence ratings in his modified LLAMA_D, and found a positive correlation between confidence ratings and accuracy. However, regarding the low reliability of this version ($r = .20$), there is a need to further refine the LLAMA_D and reapply the confidence check. Fourth, main RT is recorded for each correct answer and viewed as another factor that causes implicit learning. RT, reflecting proceduralization (Skehan, 2019), is a frequently used way to measure implicit learning (Nissen & Bullemer, 1987). Fifth, LLAMA_D can have additional test items to compensate for its lack of internal consistency reliability (Bachman, 2004). In the traditional version as well as Suzuki's (2021) modified version, there are 30 items in the testing phase, which could be extended to some degree (e.g., 40 items in LLAMA v3, see Rogers et al., 2022).

Other measurements. There are some other implicit LA measures, such as Weather Prediction (Faretta-Stutenberg & Morgan-Short, 2018; Knowlton et al., 1996; Morgan-Short et al., 2014; Pili-Moss et al., 2019), TOL (Kaller et al., 2016; Shallice, 1982), and Sugar Production (Ullman & Lovelett, 2018), which can be integrated into a broader type termed as "process control" (Li & Zhao, 2021) that measures learners' procedural memory (Buffington & Morgan-Short, 2019; Faretta-Stutenberg & Morgan-Short, 2018). In the paradigm of statistical learning, there are the Visual Statistical Learning Task (VSL; Glicksohn & Cohen, 2013) and Audio Statistical Learning Task (ASL; Siegelman et al., 2018). Similar to SRT, both measure learners' ability to recognize patterns in an ostensible random setting. Meanwhile, both require the discernment between familiar and unfamiliar items, which is somewhat analogous to LLAMA_D (For a detailed description of VSL and ASL, see Godfroid & Kim, 2021). However, they failed to produce significant correlations with L2 attainment represented by grammar learning (Godforid & Kim, 2021).

Caveats of Implicit LA Measurements

The caveats of implicit LA measurements mainly lie in the difficulties with construct validation. It is clear that projects of language testing do not allow bypassing examination of their validity, which is the prerequisite for the test to gather satisfying evidence in sampling test takers' performance and thus to convince stakeholders (Bachman & Damböck, 2018; Brown & Abeywickrama, 2010; Hughes et al., 1988; Messick, 1989). As do typical language tests, measurements of implicit LA must also comply with test validation procedures in order to assess L2 learners' unconscious learning ability.

According to Li and Zhao (2021), the validation of implicit language measurements, or other measurements of IDs, should involve the following five types of evidence: a) reliability, b) content validity, c) convergent validity, d) divergent validity, and e) predictive validity (For descriptions, see Table 1). A catalogue of caveats concerning implicit LA can be accomplished by considering the information derived from each type.

First, Implicit LA tests are usually less reliable (see Table 2). That is to say, scores on these tests are more likely to be inconsistent (Granena, 2020). The lack of reliability might become an inherent drawback of implicit LA measurements because unconscious processing is much more difficult to measure than explicit learning capacities. To avoid explicit learning, a task should be performed rapidly (Epstein, 2008), with the examined being unaware of what is exactly going to be measured (e.g., Li & Qian, 2021), which contradicts the principles that are conducive to the reliability of a test (Ward et al., 2013; Woltz, 2003). One way to alleviate this problem is to adjust the threshold for implicit LA measurements since they are not *anticipated* to become as reliable as explicit LA measurements (e.g., .7 for explicit LA and .5 for implicit LA; see Yi, 2018; Robinson, 2002). However, this is just a compromise solution. Although the standard of reliability can be "tailored" (Li, 2022, p. 22) according to the properties of various LA measurements, the low reliability does undermine the interpretation of other types of evidence, and directly accounts for the divergence among neighbouring tasks that are supposedly related (Dang et al., 2020). Thus, The reliability of these tests should be improved as much as possible, and a common way to achieve this is to add test items (Bachman, 2004). For instance, the reliability of LLAMA_D would be improved by adding the length of exposure during the studying phase and the items (e.g., new sounds) during the testing phase (see Rogers et al., 2022; Suzuki, 2021).

Table 1

Dimensions of LA Test Validation (Adapted from Li & Zhao, 2021)

Dimensions	Description
Reliability	The consistency of test takers' scores
Content validity	The relevance and representativeness for the target domain
Convergent validity	The relationship among the components within the observed test or paradigm
Divergent (Discriminant) validity	The relationship between the observed test and other tests that have different components
Predictive validity	The predictive power of the test towards the criterion variable (e.g., L2 proficiency)

Table 2

Major Studies Concerning Reliability of Implicit LA Measurements

Empirical Study	Measurement	Reliability (Internal consistency)
Kaufman et al. (2010)	Probabilistic SRT	r (split-half) = .44
Granena (2013b)	Probabilistic SRT	r (split-half) = .44
Li & Qian (2021)	Probabilistic SRT	$\alpha > .90$
Godfroid & Kim (2021)	Alternating SRT	ICC = .96
Granena (2019)	Deterministic SRT	r (split-half) = .79
Granena (2013b)	LLAMA_D	$\alpha = .75$
Granena (2019)	LLAMA_D	$\alpha = .50$
Bokander & Bylund (2020)	LLAMA_D	$\alpha = .54$
Suzuki (2021)	Modified LLAMA_D	$\alpha = .20$
Li & Qian (2021)	Priming	$\alpha = .51$

The second caveat lies with content validity. Due to limited empirical enquiry, both the relevance and representativeness of implicit LA for language learning is in doubt. For relevance, one major measurement of implicit LA, the SRT, originates from psychology and detects domain-general cognitive abilities (see Kaufman et al., 2010). Measurements like Weather Prediction and TOL are also unrelated with language learning *per se* (see Buffington et al., 2021). For representativeness, the target domain, or the theoretical background of implicit LA is hard to discern since the measurements involve both domain-general (i.e., cross-disciplinary) and domain-specific (i.e., SLA) cognitive abilities (Li & DeKeyser, 2021), and implicit LA can be justified through several paradigms (e.g., Declarative/Procedural Model, see Ullman & Loverlett, 2018; Statistic Learning, see Christiansen, 2019). A possible solution is to address the linkage between the domain-general cognitive abilities and language learning. For example, patterns and regularities not only occur in the daily routine, but also serve as typical phenomena in natural languages comprising sounds and lexis (Ellis, 2005). Skehan (2022) also noted that non-linguistic measurements are inseparable for future LA batteries. Therefore, to better validate implicit LA measurements, human cognition in general should be taken into account.

The third caveat concerns the convergent and divergent validity. The divergence of implicit LA from explicit LA has been largely confirmed by quantitative methods like factor analysis and structure equation modeling (e.g., Granena, 2019; Li & Qian, 2021), yet there is only scant evidence for the convergent validity of implicit LA measurements, which means that they are not only weakly correlated with explicit LA, but also distinct from each other. For instance, a

negligible correlation was exhibited between SRT and LLAMA_D ($r = .06$) (see Li & Qian, 2021), and SRT was also found to load on a different factor from measurements of procedural memory (e.g., Tower of London) although both are purported to measure implicit LA (Godfroid & Kim, 2021).

Suggestions for examining the convergent and divergent validity in implicit LA are presented twofold. First, although this construct seems to lack convergence, the number of relevant studies that focus exclusively on implicit LA is also scarce, which may not suffice for a final judgement, especially for the relation between SRT and LLAMA_D. Therefore, more studies are needed for further clarification. Second, the lack of convergent validity can be attributed to the fact that some measurements are not fully "implicit". For instance, deterministic SRT is more likely to arouse conscious learning of patterns compared with probabilistic SRT and ASRT (Howard & Howard, 1997; Kaufman et al., 2010), and further amendments would also apply to LLAMA_D to highlight its role in implicit memory (Suzuki, 2021; Perruchet, 2021).

The fourth caveat lies in predictive validity. Due to a relatively haphazard choice of predictor and criterion variables, the predictive power of implicit LA as a whole remains unclear, and current research is too limited to support an exclusive research synthesis or meta analysis (for explicit LA, see Li, 2016). However, it has been largely shown that the two "default" (Li & Qian, 2021) measurements of implicit LA, SRT and LLAMA_D, show a certain degree of prediction. Kaufman et al. (2010) determined that implicit learning measured by probabilistic SRT can predict French and German course grade ($r = .27$ and $.29$ respectively, outperforming general intelligence). Using Hi-Lab, Linck et al. (2013) stated that SRT is a robust predictor of participants' listening and reading skills (estimated coefficient $\beta = 0.46$ for either skill in full-model analysis). LLAMA_D has been found to correlate significantly, albeit weakly, with lexical complexity in speaking ($r = .17$; Granena, 2019). In Saito et al.'s (2019) study, LLAMA_D turned out to be the most striking predictor of speech proficiency ($r = .44$) for late L2 learners. Considering the potential improvement of LLAMA_D, the further increment of its predictive power, especially in terms of speech proficiency (Suzuki, 2021), might be anticipated.

All of the aforementioned caveats could result in the inconsistency between the construct and the measurements of implicit LA. Specifically, the lack of reliability and convergent validity may cause serious harm to the aptitude battery in that the results fail to explain anything, or at least prevent the tasks from measuring what they are intended to measure. To address the tension, it is necessary to *confirm* the components of implicit LA, and categorize the measurements accordingly. Besides, the modifications of LLAMA_D (see the previous section) imply that some current measurements of implicit LA could also be improved to induce incidental learning to the fullest extent possible.

Conclusion

To sum up, we have probed into fundamental concepts and controversies of implicit LA. We noted that the construct of implicit LA has not been clarified, but implicit learning and implicit memory ability might be two promising components. The measurements, regardless of various paradigms from which they are drawn, would more often than not shed light on unconscious learning (input) from the learners' surroundings, unconscious retrieval from learners' memory, or both. Based on the caveats of validating and using implicit LA measurements, we noted the gap between the LA construct and the measuring instruments that would impair the application of the aptitude batteries. However, by clarifying the components and revising the testing measurements, implicit LA might greatly enhance the scope of LA.

References

Bachman, L. F. (2004). *Statistical analysis for language assessment*. Cambridge University Press.

Bachman, L. F., & Damböck, B. (2018). *Language assessment for classroom teachers*. Oxford University Press.

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition*, *31*, 175-198. https://doi.org/10.1017/S0272263109090275

Bokander, L., & Bylund, E. (2020). Probing the internal validity of the LLAMA language aptitude tests. *Language Learning*, *70*, 11-47. https://doi.org/10.1111/lang.12368

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). Pearson Education.

Buffington, J., & Morgan-Short, K. (2019). Declarative and procedural memory as individual differences in second language aptitude. In Z. Wen, P. Skehan, A. Biedron, S. Li, & R. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 215-237). Routledge.

Buffington, J., Demos, A., & Morgan-Short, K. (2021). The reliability and validity of procedural memory assessments used in second language acquisition research. *Studies in Second Language Acquisition*, *43*(3), 635-662. https://doi.org/10.1017/S0272263121000127

Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education* (pp. 87-136). University of Pittsburgh Press.

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11-29). Prentice-Hall.

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, *11*, 468-481. https://doi.org/10.1111/tops.12332

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated?. *Trends in cognitive sciences*, *24*(4), 267-269. https://doi.org/10.1016/j.tics.2020.01.007

Doughty, C. J. (2019). Cognitive language aptitude. *Language Learning*, *69*, 101-126. https://doi.org/10.1111/lang.12322

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*, 305-352. https://doi.org/10.1017/S027226310505014X

Epstein, S. (2008). Intuition from the perspective of cognitive–experiential self-theory. In H. Plessner, C. Betsch, & T. Betsch (Eds.), *Intuition in judgment and decision making* (pp. 23-37). Erlbaum.

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*, 67-101. https://doi.org/10.1177/0267658316684903

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*, 1128-1153. https://doi.org/10.1037/bul0000210

Glicksohn, A., & Cohen, A. (2013). The role of cross-modal associations in statistical learning. *Psychonomic Bulletin & Review*, *20*, 1161-1169. https://doi.org/10.3758/s13423-013-0458-4

Godfroid, A., & Kim, K. (2021). The contributions of implicit-statistical learning aptitude to implicit second-language knowledge. *Studies in Second Language Acquisition*, *43*(3), 606-634. https://doi.org/10.1017/S0272263121000085

Goldstein, E. (2011). *Cognitive psychology: connecting mind, research, and everyday experience*. Cengage Learning.

Granena, G. (2013a). Cognitive aptitudes for L2 learning and the LLAMA language aptitude test. In G. Granena, & Long, M. H. (Ed.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105-130). John Benjamins.

Granena, G. (2013b). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, *63*, 665–703. https://doi.org/10.1111/lang.12018

Granena, G. (2019). Cognitive aptitudes and L2 speaking proficiency: Links between LLAMA and Hi-LAB. *Studies in Second Language Acquisition*, *41*, 313-336.

Granena, G. (2020). *Implicit language aptitude*. Cambridge University Press. https://doi.org/10.1017/S0272263118000256

Granena, G., & Yilmaz, Y. (2019). Language aptitude profiles and the effectiveness of implicit and explicit corrective feedback. In R. Leow (ed.), *The Routledge handbook of second language research in classroom learning* (pp. 438-451). Routledge.

Guo, X. (2003). A review of implicit learning. *Journal of East China Normal University (Educational Sciences)*, *22*(1), 49-55. https://doi.org/10.16382/j.cnki.1000-5560.2004.01.008

Gupta, P., & Cohen, N. J. (2002). Theoretical and computational analysis of skill learning, repetition priming, and procedural memory. *Psychological Review*, *109*, 401-448. https://doi.org/10.1037/0033-295X.109.2.401

Hamrick, P. (2015). Declarative and procedural memory abilities as individual differences in incidental language learning. *Learning and Individual Differences*, *44*, 9-15. https://doi.org/10.1016/j.lindif.2015.10.003

Howard Jr, 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. https://doi.org/10.1037/0882-7974.12.4.634

Hughes, A., Porter, D., & Weir, C. J. (1988). *Validating the ELTS Test: A Critical Review*. The British Council and the University of Cambridge Local Examination Syndicate.

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology Section A*, *54*, 1105-1124. https://doi.org/10.1080/713756001

Kaller, C. P., Debelak, R., Köstering, L., Egle, J., Rahm, B., Wild, P. S., Blettner, M., Beutel, M. E., & Unterrainer, J. M. (2016). Assessing planning ability across the adult life span: Population-representative and age-adjusted reliability estimates for the Tower of London (TOL-F). *Archives of Clinical Neuropsychology*, *31*, 148-164. https://doi.org/10.1093/arclin/acv088

Kaufman, S., DeYoung, C., Gray, J., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321-340. https://doi.org/10.1016/j.cognition.2010.05.011

Kim, T. Y., & Kim, Y. (2021). Structural relationship between L2 learning motivation and resilience and their impact on motivated behavior and L2 proficiency. *Journal of Psycholinguistic Research*, *50*(2), 417-436. https://doi.org/10.1007/s10936-020-09721-8

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399-1402. https://doi.org/10.1126/science.273.5280.1399

Li, L., & Luo, S. S. (2019). Development and preliminary validation of a foreign language aptitude test for Chinese learners of foreign languages. In Z. E. Wen, P. Skehan, A. Biedroń, S. Li, & R. L. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 33-55). Routledge.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, *38*(4), 801-842. https://doi.org/10.1017/S027226311500042X

Li, S. (2017). The effects of cognitive aptitudes on the process and product of L2 interaction. In L. Gurzynski-Weiss (Ed.), *Expanding individual difference research in the interaction approach: Investigating learners, instructors, and other interlocutors* (pp. 41-70). John Benjamins.

Li, S. (2020). Implicit language aptitude: A new construct in second language acquisition [Conference Session]. The 6th Forum on Second Language Acquisition in China, Shanghai, China.

Li, S. (2021). Individual differences in second language acquisition: Theory, research and practice [Poster Presentation]. Shanghai, China.

Li, S. (2022). Explicit and implicit language aptitudes. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge Handbook of Second Language Acquisition and Individual Differences* (pp. 37-53). Routledge. https://doi.org/10.4324/9781003270546-15

Li, S., & DeKeyser, R. (2021). Implicit Language Aptitude: Conceptualizing the construct, validating the measures, and examining the evidence: Introduction to the special issue. *Studies in Second Language Acquisition*, *43*(3), 473-497. https://doi.org/10.1017/S0272263121000024

Li, S., & Qian, J. (2021). Exploring syntactic priming as a measure of implicit language aptitude. *Studies in Second Language Acquisition*, *43*(3), 574-605. https://doi.org/10.1017/S0272263120000698

Li, S., & Zhao, H. (2021). The methodology of the research on language aptitude: A systematic review. *Annual Review of Applied Linguistics*, *41*, 25–54. https://doi.org/10.1017/S0267190520000136

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., Smith, B. K., Bunting, M. F., & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language learning*, *63*(3), 530-566. https://doi.org/10.1111/lang.12011

Long, M. (2017). Instructed second language acquisition (ISLA): Geopolitics, methodological issues, and some major research questions. *Instructed Second Language Acquisition*, *1*, 7-44. https://doi.org/10.1558/isla.33314

Luo, S., Zhang, Y., & Zhao, H. (2021). *Practical language testing and assessment*. Foreign Language Teaching and Research Press.

MacIntyre, P. D., & Wang, L. (2022). Anxiety. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge Handbook of Second Language Acquisition and Individual Differences* (pp. 175-189). Routledge. https://doi.org/10.4324/9781003270546-15

Meara, P. (2005). *LLAMA language aptitude tests*. Lognostics.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). American Council on Education & Macmillan. https://doi.org/10.1002/j.2330-8516.1987.tb00244.x

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong, P. C. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, *17*, 56-72. https://doi.org/10.1017/S1366728912000715

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1-32. https://doi.org/10.1016/0010-0285(87)90002-8

Perruchet, P. (2008). Implicit learning. In J. Byrne (Ed.),*Cognitive psychology of memory. Vol. 2 of learning and memory: A comprehensive reference* (pp. 597-621). Elsevier.

Perruchet, P. (2021). Why is the componential construct of implicit language aptitude so difficult to capture?: A commentary on the special issue. *Studies in Second Language Acquisition*, *43*(3), 677-691. https://doi.org/10.1017/S027226312100019X

Pili-Moss, D., & Brill–Schuetz, K. A., Faretta-Stutenberg, M., & Morgan-Short, K. (2019). Contributions of declarative and procedural memory to accuracy and automatization during second language practice. *Bilingualism: Language and Cognition*, *23*, 1-13. https://doi.org/10.1017/S1366728919000543

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological bulletin*, *133*(2), 227. https://doi.org/10.1037/0033-2909.133.2.227

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855-863. https://doi.org/10.1016/S0022-5371(67)80149-X

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford University Press.

Rebuschat, P., & Monaghan, P. (2019). Editors' introduction: Aligning implicit learning and statistical learning: Two approaches, one phenomenon. *Topics in Cognitive Science*, *11*, 459-467. https://doi.org/10.1111/tops.12438

Robinson, P. (2002). Individual differences in intelligence, aptitude and working memory during adult incidental second language learning: A replication and extension of Reber, Walkenfeld, and Hernstadt (1991). In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211-266). John Benjamins.

Rogers, V., Meara, P., & Rogers, B. (2022). Testing Language Aptitude: LLAMA evolution and refinement. In Z. E. Wen, P. Skehan., & R. Sparks (Eds.), *Language aptitude: Theory and Practice* (pp. 63-90). Cambridge University Press.

Saito, K. (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning*, *67*, 665-693. https://doi.org/10.1111/lang.12244

Saito, K., Sun, H., & Tierney, A. (2019). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and neurophysiological measures. *Bilingualism: Language and Cognition*, *22*(5), 1123-1140. https://doi.org/10.1017/S1366728918000895

Scott, R. B., & Dienes, Z. (2010). Knowledge applied to new domains: The unconscious succeeds where the conscious fails. *Consciousness and Cognition: An International Journal*, *19*, 391-398. https://doi.org/10.1016/j.concog.2009.11.009

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London*, *298*, 199-209. https://doi.org/10.1098/rstb.1982.0082

Shanks, D., Rowland, L., & Ranger, M. (2005). Attentional load and implicit sequence learning. *Psychological Research*, *69*, 369-382. https://doi.org/10.1007/s00426-004-0211-8

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105-120. https://doi.org/10.1016/j.jml.2015.02.001

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198-213. https://doi.org/10.1016/j.cognition.2018.04.011

Skehan, P. (2019). Language aptitude implicates language and cognitive skills. In Z. E. Wen, P. Skehan, A. Biedroń, S. Li, & R. L. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 56-77). Routledge.

Skehan, P. (2022). Testing language aptitude: A commentary on batteries and reanalysis of constructs. In Z. E. Wen, P. Skehan., & R. Sparks (Eds.), *Language aptitude: Theory and Practice* (pp. 234-271). Cambridge University Press.

Sparks, R. L., Patton, J., & Luebbers, J. (2019). Language aptitude: Insights from U.S. high school students. In Z. E. Wen, P. Skehan, A. Biedroń, S. Li, & R. L. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 123-146). Routledge.

Squire, L. R. (1987). *Memory and brain*. Oxford University Press.

Suzuki, Y. (2021). Probing the construct validity of LLAMA_D as a measure of implicit learning aptitude: Incidental instructions, confidence ratings, and reaction time. *Studies in Second Language Acquisition*, *43*(3), 663-676. https://doi.org/10.1017/S0272263120000704

Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning*, *65*, 860-895. https://doi.org/10.1111/lang.12138

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747-790. https://doi.org/10.1111/lang.12241

Tagarelli, K. M., Ruiz, S., Vega, J. L. M., & Rebuschat, P. (2016). Variability in second language learning: The roles of individual differences, learning conditions, and linguistic complexity. *Studies in Second Language Acquisition*, *38*(2), 293-316. https://doi.org/10.1017/S0272263116000036

Ullman, M. T. (2015). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten, & J. Williams (Eds.), *Theories of second language acquisition: An introduction* (pp. 135-158). Lawrence Erlbaum Associates, Inc.

Ullman, M., & Lovelett, J. (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research*, *34*, 39-65. https://doi.org/10.1177/0267658316675195

Walker, N., Monaghan, P., Schoetensack, C., & Rebuschat, P. (2020). Distinctions in the acquisition of vocabulary and grammar: An individual differences approach. *Language Learning*, *70*, 221-254. https://doi.org/10.1111/lang.12395

Wang, W. C., & Yonelinas, A. P. (2012). Familiarity is related to conceptual implicit memory: An examination of individual differences. *Psychonomic Bulletin & Review*, *19*, 1154-1164. https://doi.org/10.3758/s13423-012-0298-7

Ward, E., Berry, C., & Shanks, D. (2013). Age effects on explicit and implicit memory. *Frontiers in Psychology*, *4*, 1-11. https://doi.org/10.3389/fpsyg.2013.00639

Wen, Z. E., Skehan, P., Biedroń, A., Li, S., & Sparks, R. L. (2019). *Language aptitude: Advancing theory, testing, research and practice*. Routledge.

Wen, Z. E., Skehan, P., Sparks, R. L. (2022). Language aptitude research: From testing to theory and practice. In Z. E. Wen, P. Skehan., & R. Sparks (Eds.), *Language aptitude: Theory and Practice* (pp. 1-21). Cambridge University Press.

Woltz, D. J. (2003). Implicit cognitive processes as aptitudes for learning. *Educational Psychologist*, *38*, 95-104. https://doi.org/10.1207/S15326985EP3802_4

Yi, W. (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*, *40*(4), 831-856. https://doi.org/10.1017/S0272263118000141

Zhang, R., & Chen, Y. (2020). Exploring the relationship between individual difference and implicit L2 knowledge. *Modern Foreign Language*, *43*(3), 377-388.

Ethics Declarations

Competing Interests

No, there are no conflicting interests.

Rights and Permissions