Critical Commentary

EPILOGUE

EVIDENCE FOR THE VALIDITY OF IMPLICIT APTITUDE AND THE NEED FOR CONSTRUCT VALIDATION

Robert DeKeyser

University of Maryland

Shaofeng Li 🄳 *

Florida State University

Abstract

In this commentary, we summarize the findings of the seven included studies that examined implicit language aptitude from various perspectives and highlight issues to be resolved in the validation of this new construct in second language research. We start by providing an overview of the contributions of the studies. We then identify the lack of convergent validity of the measures of implicit aptitude reported in the included studies and problematize the equally varied nature of the measurement of implicit knowledge—the outcome variable of aptitude research—and related concepts. In particular, by drawing on empirical evidence and theoretical claims, we attempt to clarify the relationships between implicit and explicit knowledge, implicit and explicit learning, and implicit and explicit instruction. Next, we draw attention to the interactions reported by the included studies between aptitude and outcome measures and between aptitude and instruction type, emphasizing the value and importance of interactional research. We conclude by making recommendations for future research.

The guest editors for this special issue decided to invite contributors tackling aptitude for implicit language learning ("implicit aptitude" for the sake of brevity) from a variety of angles. Although implicit aptitude is a relatively new concept that has been conceptualized and operationalized in a variety of ways and that often overlaps with other concepts, enough research has accumulated now to take stock of what we have, especially what we need, to make this area of research move forward. The studies contributed the following evidence for the validity of the construct of implicit aptitude. First, implicit aptitude is separate from cognitive abilities in the explicit paradigm. The studies show that measures hypothesized to measure implicit aptitude such as serial reaction time (SRT), syntactic

* Correspondence concerning this article should be addressed to Shaofeng Li, Florida State University, Tallahassee, Florida 32306. E-mail: sli9@fsu.edu

priming, or frequency-following response (FFR) are uncorrelated with measures of explicit aptitude (Li & Qian; Yilmaz & Granena), declarative memory (Buffington et al.), working memory (Fu & Li), or music memory (Sun et al.). Second, measures of implicit aptitude lack convergent validity—they are uncorrelated or fail to load on the same factor—a point to be revisited in later sections. This is a striking feature of implicit aptitude, which contrasts with explicit aptitude, whose measures are typically significantly correlated or load on the same factor. This finding suggests that there are likely different routes through which implicit learning occurs and there is no overlap between them. Third, in naturalistic or immersion settings, implicit aptitude is more likely to be correlated with the L2 attainment of learners with longer residence in the host country (Godfroid & Kim) than those with shorter residence (Sun et al.). Fourth, in foreign language settings where learners have limited exposure to the target language and where learners receive heavy doses of form-focused instruction (which results in explicit knowledge), implicit aptitude has low predictive validity (Li & Qian). However, it may show significant associations with L2 attainment if implicit aptitude is measured through coefficient of variance (CV) and the outcome is operationalized as fluency—a proxy of implicit knowledge whose mechanism matches CV in that they are both based on time measures (Suzuki). Fifth, the experimental studies examining the interaction between aptitude type and treatment type show that the role of implicit aptitude varies as a function of learning conditions. Specifically, implicit aptitude is implicated in implicit feedback rather than explicit feedback (Yilmaz & Granena) and in immediate feedback (after instruction) instead of delayed feedback (Fu & Li). It appears that implicit aptitude is more likely to be involved in conditions where implicit learning is encouraged (Yilmaz & Granena) or where learners have a solid base of explicit knowledge (obtained through instruction and reinforced through immediate feedback) that is proceduralized through communicative practice (Fu & Li).

In addition to evidence on the validity of a new construct in SLA, the studies of the special issue contribute to the field by their originality—they examine new perspectives and introduce new measures; robustness—they are based on relatively large samples and use rigorous methods; and variety—they investigate the topic using different approaches. The contributions to this special issue demonstrate the variety of approaches very clearly. For instance, some authors' contributions deal with the declarative/procedural distinction rather than the explicit/implicit distinction, but given the amount of conceptual overlap between the two dichotomies and the similar measures of the two pairs of variables, we decided this was within the purview of this issue. The authors of various contributions also lay bare the lack of convergent and divergent validity of some measures, be they considered explicit/implicit or declarative/procedural, and this is true for aptitude measures as well as measures of knowledge.

On the aptitude side, the lack of convergent and divergent validity is particularly obvious in the articles by Godfroid and Kim; Buffington, Demos, and Morgan-Short; and Li and Qian. In Godfroid and Kim, the measures of "implicit-statistical learning aptitude" do not cluster together as expected, but instead show three clearly separate factors, which Godfroid and Kim call motor sequence learning, procedural memory, and statistical learning. This is only tentative, of course, as two of the factors are represented by only one test, so it is hard to link them to any construct. In Buffington et al., the same lack of expected patterning is found, this time for measures of declarative and procedural

memory: The three measures of procedural memory—the dual-task Weather Prediction Task, the Alternating Serial Reaction Time Task, and the Tower of London—did not show convergent validity (they were uncorrelated), and the Weather Prediction Task did not even show divergent validity with the measures of declarative memory (it loaded under the factor of declarative memory). Similarly, in Li and Qian's study, syntactic priming did not converge with other putative measures of implicit aptitude such as SRT and LLAMA_D; LLAMA_D even loaded on the factor of explicit aptitude.

The need for construct validation is not unique to the measurement of implicit aptitude; it is also applicable to the operationalization of implicit knowledge and other related concepts. Over the last few decades, there has been a large amount of theorizing about implicit versus explicit instruction, implicit versus explicit learning, and implicit versus explicit knowledge. We have learned that either form of instruction does not necessarily lead to a homologous form of learning, and that a given form of learning does not necessarily determine the ultimate form(s) of knowledge. Our understanding of what "implicit" means has narrowed, and the term is now mostly used in a much more restricted sense than before. Where learning is concerned, a simple definition is that implicit learning is learning without awareness of what is being learned (DeKeyser, 1995), not just that that the learning happens incidentally. Similarly, when it comes to instruction, for most authors implicit instruction does not just mean that there is no systematic teaching of rules, but that there is little or no mentioning of formal regularities. In a strong version of task-based language teaching, it even means that instruction should not be organized around forms, whether they are mentioned or not (Long, 2015) (see Ellis et al., 2020; Li et al., 2016 for a different view). Finally, having implicit knowledge means having knowledge that can be used without paying attention to it; the same individual can of course also have explicit knowledge in parallel. The latter means that for the measurement of implicit knowledge, access to any explicit knowledge that may exist must be prevented. While time pressure can make such access harder, it does not preclude it because highly automatized explicit knowledge can be accessed very quickly. Therefore, great care must be taken to make sure that the learner's attention is completely taken away from form, for example, through word-monitoring tasks or self-paced reading (Jiang, 2012; Suzuki & DeKeyser, 2015).

On the outcome side, the measures used in the contributions to this issue vary widely. Some are explicit, such as tests of untimed grammaticality judgment and metalinguistic knowledge, while others are implicit, including measures based on accuracy (elicited imitation and oral production) and reaction time (word monitoring and self-paced reading). While Godfroid and Kim argue that accuracy-based tests are better measures of implicit knowledge, Suzuki and DeKeyser (2015) have a different view. However, as Li and Qian point out, the nature of the measured knowledge depends partly on the instructional setting or the kind of instruction learners received in the local context. For example, in foreign language contexts where learners receive intensive form-oriented instruction and where the chances of developing implicit knowledge are small, the bulk of learners' L2 knowledge is likely explicit, regardless of the test format. We would like to point out that both explicit and implicit measures are necessary in aptitude research and that it is crucial to consider the nature of the outcome measure when interpreting the findings. The declarative/procedural distinction is even harder to make where outcome measures are concerned. While it is safe to assume that the knowledge used on a

metalinguistic text is declarative in nature, it is harder to know to what extent open-ended production tests draw on declarative or procedural knowledge without either using neuroimaging or computer modeling of accuracy and reaction time data.

Finally, regardless of the type of instruction or the nature of initial learning, the nature of the knowledge used on the outcome test depends on the format of the test and the relative level of the test-taker's implicit and explicit knowledge (or declarative and procedural knowledge). Particularly important for our purposes here is that the test-takers' knowledge may be quite different from the knowledge they originally acquired by drawing on their aptitudes, especially in cases in which considerable time has elapsed between initial learning and outcome testing (Suzuki, 2017). Godfroid and Kim's article in this issue, as well as Suzuki (2017), illustrate how difficult it can be, therefore, to interpret the relationship between aptitudes and final outcomes. In Suzuki and DeKeyser (2017) there was some evidence that any implicit knowledge at the end may have developed out of earlier automatized explicit knowledge, acquired explicitly, because aptitude for explicit learning clearly predicted (automatized) explicit knowledge, which in turn predicted implicit knowledge, while the direct link between implicit aptitude and eventual implicit knowledge was much weaker. This suggests that explicit knowledge became automatized and that this automatized explicit knowledge eventually helped to develop implicit knowledge. This interpretation is corroborated by Suzuki and DeKeyser (2015), which showed that for learners with long residence in the L2 environment implicit aptitude was a good predictor for knowledge measured with a real-time comprehension task (a word-monitoring task). Godfroid and Kim showed that their two-factor solution (explicit knowledge and implicit knowledge) and three-factor solution (explicit knowledge, automatized explicit knowledge, and implicit knowledge) both showed good fit, but the two-factor model showed a better fit. In cases like this, when more than one model shows good fit, the interpretation can go both ways, depending on whether one prefers criteria like parsimony, theoretical expectation, or specific evaluation measures.

The contributors to this issue have provided considerable methodological detail and have courageously documented how tricky the conceptual as well as methodological issues in this area of research can be. They have also paid particular attention to the interaction of aptitude with both treatments and outcome formats. Six of the empirical articles look at the relationship between aptitudes and learning outcomes, and five find that different (measures of) aptitude(s) differentially predict success either with different types of instruction (Yilmaz & Granena; Fu & Li) or for different outcome measures (Li & Qian). The exact aptitude–treatment interactions or aptitude–testing interactions vary widely from contribution to contribution though. Yilmaz and Granena found that implicit aptitude was implicated in implicit feedback (for gender agreement) and explicit aptitude was involved in explicit feedback (for differential object marking). Fu and Li shows an interaction of feedback timing with aptitude: Implicit aptitude predicted the effectiveness of immediate feedback while explicit aptitude was associated with the effects of delayed feedback. These results are encouraging, especially as aptitude–treatment interactions are of special relevance to education (DeKeyser, 2019a; Li, 2017, 2018), but of course the results need to be replicated, especially given the difference in aptitude measures used and the uncertain validity of some of the measures.

Interactions between aptitudes and outcome measures were found in three of the contributions to this study. Li and Qian show an unexpected pattern where measures of

explicit aptitude predicted all aspects of proficiency, while implicit aptitude only predicted metalinguistic knowledge (negatively). Godfroid and Kim found that implicit aptitude predicted accuracy but not reaction time (both on timed tests). Suzuki digs even deeper by showing that different aspects of LLAMA D, which has been a measure of implicit aptitude, showed different predictive validity: Only the coefficient of variation for old items was predictive of only the mid-clause duration of learners' speech production. Only Sun and Saito did not document any interactions: They found that only explicit, not implicit, aptitude predicted the acquisition of both segmental and nonsegmental prosody in early stages of L2 learning in an immersion setting. However, in their study, the outcome measure, which required learners to recognize sounds in isolated words, likely tapped explicit knowledge. Implicit aptitude may have demonstrated predictive validity had a measure of implicit phonetic knowledge been utilized.

Given the varied and complicated findings of the research that the present issue included and the uncertainty of the validity of the instruments most commonly used as measures of implicit aptitude, one cannot look past the urgent need for more research mapping the relationships between the different measures. This is not surprising, as the field of implicit aptitude for SLA is very young. Perhaps we all need to be a bit more patient and wait for better construct validity before making claims about predictive validity or how to interpret it. In the meantime, those who want to investigate the predictive validity of implicit aptitude at this point should do what various contributions to this issue have done: Use a variety of measures of aptitude to mitigate the risk of drawing conclusions using a single measure of doubtful validity and thereby misinterpreting or overgeneralizing the findings. Given how often even measures of working memory show little convergent validity and/or have very different predictive validity for the same outcome measures (DeKeyser, 2019b; Wen & Li, 2019), in a field with so much more research history than aptitude for implicit learning, we should take heart and forge ahead. Understanding the role of implicit and explicit aptitude in second language learning is important for understanding the role of implicit and explicit learning, a central issue in second language acquisition research and applied linguistics, and the contributions to this issue, together with the comments from Perruchet, provide rich documentation of how research involving implicit aptitude can avoid a variety of pitfalls, and what methodological options are most likely to advance research in this area.

REFERENCES

DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, *17*, 379–410.

DeKeyser, R. (2019a). Aptitude-treatment interaction in second language learning. Introduction to the special issue. *Journal of Second Language Studies*, *2*, 165–168.

DeKeyser, R. (2019b). The future of aptitude research. In Z. Wen, P. Skehan, A. Biedron, S. Li, & R. L. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research, and practice* (pp. 317–329). Routledge.

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University.

Jiang, N. (2012). *Conducting reaction time research in second language studies*. Routledge.

Li, S. (2017). Cognitive differences and ISLA. In S. Loewen & M. Sato (Eds.), *Handbook of instructed second language learning* (pp. 396–417). Routledge.

Li, S. (2018). Language aptitude. In A. Burns & J. Richards (Eds.), *The Cambridge guide to learning English as a second language* (pp. 63–72). Cambridge University.

Li, S., Ellis, R., & Zhu, Y. (2016). Task-based versus task-supported language instruction: An experimental study. *Annual Review of Applied Linguistics*, *36*, 205–229.

Long, M. (2015). *Task-based language learning*. Wiley-Blackwell.

Suzuki, Y. (2017). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics*, *38*, 1229–1261.

Suzuki, Y., & DeKeyser, R. (2015). Does elicited imitation measure implicit knowledge? Evidence from the word-monitoring task. *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. *Language Learning*, *67*, 747–779. https://doi.org/10.1111/lang.12241.

Wen, Z., & Li, S. (2019). Working memory in L2 learning and processing. In J. Schwieter & A. Benati (Eds.), *The Cambridge handbook of language learning* (pp. 365–389). Cambridge University.