2
EXPLICIT AND IMPLICIT LANGUAGE APTITUDES

Shaofeng Li

Background

Language aptitude (alternatively “aptitude”) is a set of cognitive abilities that are essential for the learning of a second language (L2). In its traditional conceptualization, aptitude consists of three components: phonetic coding, language analytic ability, and rote memory, which roughly correspond with the learning of pronunciation, grammar, and vocabulary, respectively (Carroll, 1981). Phonetic coding refers to the ability to recognize sounds, learn sound–symbol associations, and mimicry. Language analytic ability represents the ability to 1) recognize the grammatical functions of sentence elements, and 2) extrapolate linguistic regularities from linguistic materials. Rote memory, also referred to as associative memory, is the ability to memorize the associations between words and their meanings. Aptitude has been found to be separable from other cognitive and affective individual difference variables and is strongly predictive of second language development (Li, 2016). It is considered a relatively stable trait that cannot be easily changed. However, there has been evidence that longer language learning experience may lead to higher aptitude (Granena & Long, 2013) and that it increases with age and grade level until a certain point (about age 11) before it starts to stabilize (Roehr-Brackin & Tellier, 2019). Aptitude has been found to be more strongly correlated with the effects of explicit instruction, which encourages conscious learning, such as rule explanations, than the effects of implicit instruction that encourages incidental, unconscious learning through approaches, such as task-based language teaching (Li, 2015). To a certain degree, this finding motivated the recent interest in research on implicit language aptitude.

The notion of implicit aptitude is a recent innovation in aptitude research. It refers to the ability to conduct unconscious tallying and computation of linguistic tokens to extract linguistic regularities that govern the distribution and contingency between linguistic units (Li & DeKeyser, 2021). It underlies and contributes to the process and outcome of implicit learning. In psychology, where implicit learning made its debut (Reber, 1993), the primary interest is in whether implicit learning occurs, not whether implicit learning is an ability that is predictive of learning outcomes. In the field of second language acquisition, research on language aptitude has predominantly revolved around traditional aptitude, which moderates explicit learning and can therefore be called explicit aptitude. Research on implicit language aptitude commenced with attempts to identify cognitive abilities for high-level proficiency (Linck et al., 2013) and map the relationship between different aptitudes and maturation effects (Granena & Long, 2013). Subsequently, the scope of inquiry was extended to identify the nature of L2 knowledge by exploring its associations with cognitive aptitudes (Suzuki & DeKeyser, 2015, 2017), investigating the interaction between implicit and explicit...
aptitude on the one hand and implicit and explicit instruction on the other (Yilmaz & Granena, 2019), and examining the predictive power of implicit aptitude for first language (L1) and advanced L2 speakers’ sensitivity to the frequency and probability of English collocations (Yi, 2018). Despite increasing interest in implicit aptitude, the research is in its infancy and the validity of the construct has yet to be fully established.

What theoretical models have been proposed to account for the role of language aptitude in language learning? Major theoretical models include the Carrollian model, the Fundamental Difference Hypothesis, the aptitude–treatment interaction (ATI) model, and the staged model. All the models concern traditional or explicit aptitude or were proposed with explicit aptitude in mind, although they have implications for implicit aptitude. To start with, the Carrollian model is named after John Carroll (1981), who developed the Modern Language Aptitude Test (MLAT), which has demonstrated the highest predictive validity among all aptitude tests (Li, 2015, 2016). Based on this model, aptitude is a determinant of learning success, and therefore there is no need to tailor instruction to cater to learners’ different aptitude levels. In this approach, aptitude is implicated in instruction that requires learners to engage in conscious, effortful processing of linguistic materials, hence its relevance to explicit learning/instruction instead of implicit learning/instruction.

The Fundamental Difference Hypothesis (Bley-Vroman, 1990) guided some age-related studies investigating the causes for maturation or age effects in second language acquisition (SLA), namely, that children are more successful language learners than adults—while child language acquisition is uniformly successful, adult language learning is subject to high variability. The theory states that the reason for the disparate outcomes of child and adult language acquisition is that children and adults learn differently. Whereas children have access to a domain-specific language acquisition device that allows them to acquire a language with ease and speed, adults have lost that kind of ability and must draw on domain-general, problem-solving abilities. Although Bley-Vroman (1990), who proposed this theory, did not specify what these abilities are, aptitude researchers such as DeKeyser (2000) and Abrahamsson and Hyltenstam (2008) interpreted domain-general abilities as language aptitude, which adult learners utilize to analyze linguistic input and learn a second language in a conscious manner. However, if these abilities are exclusively important for language learning, then they are domain specific. The ATI model views aptitude as a dynamic construct that interfaces with treatment or learning conditions. The theory holds that different learning conditions pose different processing demands on the learner and may draw on different sets or clusters of abilities (Robinson, 2011). Aptitude–treatment interaction can be defined as the differential associations of aptitudes with any instructional feature that is manipulated for the purpose of ascertaining whether it leads to variation in the way learners deploy their cognitive abilities. The ATI model proposed by Robinson (2002) posits different clusters of cognitive abilities, called aptitude complexes, for different learning conditions, such as a focus on form, incidental learning, etc. Robinson (2002) also states that adult learning is a conscious process and therefore the corollary is that there is no role for implicit aptitude. The staged model seeks to account for the interaction between aptitude and learning stages (Skehan, 2012). According to this model, language learning happens in four stages that implicate different aptitude components. The first stage is noticing, where learners receive and detect linguistic input, and this stage implicates phonetic coding and working memory. The second stage is patterning, where learners extrapolate rules and regularities based on noticed input. This stage requires learners to utilize their analytic ability to process and understand the structure of linguistic input. The third stage is controlling, where learners proceduralize learned knowledge, and proceduralization refers to the application of linguistic knowledge in language behaviors or skills, such as speaking and writing. Skehan did not specify what type of ability is necessary for proceduralization, and this gap can be conveniently filled by implicit aptitude, which is called procedural memory in the declarative-procedural model (Morgan-Short et al., 2014). Implicit aptitude may also fulfill its duties in the final stage of this model—lexicalization, where L2 knowledge is automatized and retrieved at a fast rate, similar to the way lexical items or unanalyzed chunks are activated.
Explicit and Implicit Language Aptitudes

in long-term memory. As can be seen, implicit aptitude can be easily fit into this model because of its importance at more advanced stages of learning, and it stands in contrast with explicit aptitude that is more important at the initial learning stages.

**Research**

**Evidence**

In general, the research has contributed three types of evidence for the construct of language aptitude: convergent validity, divergent validity, and predictive validity. Convergent validity refers to whether aptitude is correlated or behaves similarly with variables in the same paradigm. Divergent validity concerns whether it is uncorrelated with or dissociable from variables in other paradigms hypothesized to be different from the paradigm in question. Predictive validity pertains to whether it is correlated with learning behaviors, processes, or outcomes. In the following sections, the review of empirical evidence is structured based on these three types of evidence, and given the close connection between convergent and divergent validity, they are discussed in one section. In the course of the review, traditional aptitude refers to explicit aptitude, and implicit aptitude refers to the newly conceptualized type of language aptitude hypothesized to moderate implicit learning. Aptitude refers to the general concept of language aptitude that includes both explicit and implicit aptitude.

**Convergent and Divergent Validity**

The convergent and divergent validity concerns the connections of language aptitude with its covariates or other cognitive and affective variables. Given the explicit–implicit distinction in aptitude/instruction/knowledge, the two types of validity can be further discussed from this perspective. Thus, evidence for the two types of validity should show that explicit and implicit aptitudes are correlated with explicit and implicit knowledge/instruction, respectively, and accordingly they are uncorrelated with the knowledge and instruction in the opposite paradigm. Li (2016) meta-analyzed all the empirical research that examined traditional aptitude associations with other individual difference variables. The results showed that traditional aptitude is significantly correlated with, and yet dissociable from, intelligence and working memory, both of which are abilities for explicit learning. The results also showed that traditional aptitude is uncorrelated with motivation and negatively correlated with anxiety. These results suggest that as a set of abilities for conscious learning, traditional aptitude is correlated with other abilities for conscious learning and that it is an independent construct that is dissociable from variables in the same paradigm and different from variables in other paradigms. Further evidence for the convergent and divergent validity of traditional/explicit aptitude comes from Li’s meta-analyses (2015, 2017) that showed a stronger correlation between traditional aptitude and the effects of explicit instruction than implicit instruction.

What evidence is there for convergent and divergent validity of the newly conceptualized implicit aptitude? Implicit aptitude has been operationalized as sequence learning (Granena, 2013), which is measured by means of serial reaction time (SRT) tasks and LLAMA_D. In an SRT test, learners respond to a stimulus on a computer screen whose locations are based on a certain sequence, and it is hypothesized that repeated exposure to the sequence will make the learner unconsciously learn the sequence. Learning is represented by faster responses to the stimuli based on the sequence compared with responses to a control sequence that is less frequent in the stimuli. In LLAMA_D, the learner listens to some unfamiliar sound sequences and is subsequently asked to recognize them. Given that implicit aptitude has been measured using these two tests, our discussion of the two types of validity focuses on these two measures. For SRT, it has been found to be uncorrelated with, or separate from, abilities for explicit learning, such as traditional/explicit apti-
tude, phonological short-term memory, and executive working memory (Granena, 2019; Hamrick, 2015; Kaufman et al., 2010; Linck et al., 2013; Suzuki & DeKeyser, 2017; Yi, 2018). Besides, SRT was not predictive of the effects of explicit instruction, whereas phonological short-term memory was (Granena & Yilmaz, 2019; Yilmaz & Granena, 2019). These findings suggest that SRT behaves differently from measures of variables in the explicit domain and therefore has divergent validity. Regarding convergent validity, SRT has been found to be correlated with performance on measures hypothesized to measure implicit knowledge (Granena, 2013; Suzuki & DeKeyser, 2015), the effects of explicit instruction (Granena & Yilmaz, 2019; Yilmaz & Granena, 2019), and an experiential-intuitive cognitive style (Granena, 2016). SRT also loaded on the same factor with LLAMA_D, another measure of implicit aptitude (Granena, 2012).

Turning to LLAMA_D, in terms of divergent validity, it was found to be uncorrelated with the other subtests of the LLAMA battery: LLAMA_B (rote memory), LLAMA_E (phonetic coding), and LLAMA_F (grammatical sensitivity)—all measures of explicit aptitude (Saito et al., 2019). It also loaded on different factors from measures of explicit aptitude (Granena, 2012, 2019). However, there is also evidence for its lack of dissociation with variables in the explicit domain. For example, it was correlated with explicit aptitude (but not implicit aptitude) (Granena, 2019). One methodological variation that may cause the inconsistency in the validity of LLAMA_D is whether learners are informed that they will be tested on the items they heard during the study phase. In Granena’s (2019) studies, the prospect of being tested might have encouraged intentional/explicit learning. In Saito et al. (2019), learners were asked to listen to the auditory stimuli to check the volume, which may have encouraged implicit learning. For convergent validity, Granena (2012) found that LLAMA_D and SRT loaded on the same factor. Granena (2019) revealed that LLAMA_D loaded on the same factor as semantic priming, and this factor was called “implicit memory”, whose primary function is to retrieve information from long-term memory.

**Predictive Validity**

Research on the predictive validity of language aptitude can be divided into two broad categories: correlational and experimental. The former investigates the associations between aptitude (or aptitude components) and L2 proficiency, and the latter investigates whether aptitude and aptitude components mediate the processes and effectiveness of different types of instructional treatments in different ways. Correlational research can be further divided into two categories: studies investigating the relationships between aptitude and adult L2 learning, and studies investigating the interface between aptitude and age. Thus, the discussion of the literature in the following sections is structured around three themes: aptitude and L2 attainment, aptitude and age, and aptitude and treatment effects.

**Aptitude and L2 Attainment**

In a meta-analysis, Li (2016) aggregated the results of correlational research on the associations between traditional/explicit aptitude and L2 achievements. The results showed that the average correlation between overall aptitude measured by whole test batteries and overall proficiency represented by learners’ course grades or their scores on proficiency tests is \( r = 0.50 \), which means 25% of the variance of L2 learning is explained by aptitude. It is worth pointing out that the predictive power of aptitude is stronger than other individual difference factors that have been meta-analyzed, such as anxiety, \( r = -0.36 \) (Teimouri et al., 2019); motivation, \( r = 0.37 \) (Masgoret & Gardner, 2003); and working memory, \( r = 0.27 \) (Linck et al., 2013). Thus, traditional aptitude has by far been the most robust and consistent predictor of L2 attainment among individual difference factors that have been examined in the research. However, the predictive power of overall aptitude for different aspects of L2 learning is not equal. It is a stronger predictor of listening, reading, speaking, and grammar than vocabulary, and a non-significant predictor of writing (Li, 2016).
Since traditional aptitude consists of three components: phonetic coding, analytic ability, and rote memory, it is necessary to know how predictive each aptitude component is of learning outcomes. Li’s (2016) meta-analysis shows the following patterns. Phonetic coding is a stronger predictor of overall proficiency than the other two aptitude components, and this is probably because most studies included in the meta-analysis involved beginning foreign language learners who relied heavily on phonetic coding when processing linguistic input in the initial stages of L2 learning (Skehan, 2012). Phonetic coding is more predictive of vocabulary than other aspects of L2 learning, suggesting the importance of this cognitive ability in decoding and encoding novel sounds in vocabulary learning. Interestingly, phonetic coding seems more important than rote memory in vocabulary learning, which is counterintuitive because rote memory is the ability to memorize word–meaning associations and therefore the default mechanism for vocabulary learning. Phonetic coding is the least predictive of listening comprehension, which may suggest that L2 listeners rely less on bottom–up decoding skills and more on top–down skills using contextual and schematic information. Language analytic ability showed stronger predictive power for grammar learning and reading comprehension than other outcome measures. It is of no surprise that the ability to learn grammar is important for grammar learning but the finding that it is important for reading comprehension is significant. It suggests the critical role of grammar knowledge in L2 comprehension, echoing the findings reported by Jeon and Yamashita’s (2014) meta-analysis about factors that correlate with L2 reading comprehension. They found that among all correlates of reading comprehension, learners’ grammar knowledge was the strongest predictor, with an average correlation of 0.85, demonstrating the importance of morphosyntactic knowledge in understanding L2 written texts. The third component, rote memory, showed the weakest correlations with general L2 proficiency and with specific aspects of proficiency, such as L2 knowledge and L2 skills. This finding is significant in that it constitutes evidence against the commonly assumed importance of rote memory in learning a foreign language. However, rote memory, which is also called associative memory or declarative memory in cognitive psychology, may be more important in initial than advanced L2 learning.

Li’s meta-analyses (2015, 2016) also showed that traditional aptitude is more strongly correlated with high school learners’ L2 achievements than that of university learners. Li speculated that this may mean traditional aptitude is more likely to be drawn upon by beginning learners than more advanced learners, given that high school students are less advanced than university students. This finding confirms Carroll’s (1990) speculation that traditional aptitude, as represented in the MLAT, concerns the rate of learning a language “from scratch” (p. 24), and there is a need for research into abilities relevant in later stages of learning. Indeed, the relationship between aptitude and proficiency is a topic that has received very little attention in primary research. Li’s (2015) findings regarding high school and university learners are based on a meta-analysis, which in turn is based on studies carried out in different settings. In those studies, proficiency was not examined as an independent variable, and the learners were not matched otherwise. The only published study on aptitude and proficiency is conducted Curtin, Avner, and Smith (1983), who administered the Pimsleur Language Aptitude Battery (PLAB) (Pimsleur, Read, & Stansfield, 2004) to first- and fourth-year high school students. The study found stronger associations between aptitude (PLAB) scores and course grades among first-year students compared with fourth-year students.

A few studies have examined the associations between implicit aptitude and L2 attainment. These studies can be divided into two categories based on whether they examined instructed or naturalistic learners. In the former category, Kaufman et al. (2010) examined the associations between implicit aptitude and foreign language learning together with other academic achievements. In their study, L1 English students at a secondary school completed an SRT task as a measure of implicit aptitude and other tasks measuring their personality and cognitive abilities, such as working memory and intelligence. The dependent variables were students’ scores on subject-related national exams. Implicit aptitude was found to be a significant predictor of the students’
foreign language scores in German and French. Kaufman et al.’s study was the first to examine implicit learning as a predictor of L2 achievements, and SRT has become the default test of implicit aptitude in SLA thereafter. Granena (2019) reported that intermediate university L2 Spanish learners’ implicit aptitude tested via SRT was not a significant predictor of their speech performance. However, LLAMA_D, another measure of implicit aptitude, loaded onto the same factor as semantic priming, and this factor was significantly predictive of speech fluency. Linck et al. (2013) sought to identify aptitudes for high-level proficiency. They developed the Hi-LAB, which consisted of 11 tests measuring implicit aptitude (SRT), working memory, long-term memory, and auditory perceptual acuity. The dependent variable was advanced L2 Spanish learners’ reading and listening comprehension. Among the 11 tests, implicit aptitude, rote memory, and phonological short-term memory were found to be significant predictors of advanced L2 Spanish listening and reading comprehension. Saito et al. (2019) tracked a cohort of Japanese university students’ L2 English pronunciation development for two semesters and examined whether aptitude played a role in the learners’ pronunciation development. They found that rote memory and phonetic coding—two explicit learning abilities—were predictive of the learners’ fluency in the first semester, while LLAMA_D—a measure of implicit aptitude—was significantly predictive of their pronunciation accuracy in the second semester.

Three studies explored the role of implicit aptitude in naturalistic settings, all examining adult L2 learners who have received formal language instruction before coming to the country where the L2 is spoken. Suzuki and DeKeyser (2015, 2017) investigated the associations between L1 Chinese speakers’ implicit/explicit aptitude and implicit/explicit grammar knowledge in L2 Japanese. They found that SRT was a near-significant predictor of long-residence learners’ (who lived in the country longer than two years) implicit knowledge, but it was not predictive of short-residence learners’ (who lived in the country for less than two years) grammar knowledge (be it implicit or explicit). However, LLAMA_F, a test of language analytic ability, was predictive of short-residence learners’ automatized explicit knowledge. Yi (2018) reported that adult L2 English speakers who had lived in the US for an average of 25.8 months did not draw on either implicit (SRT) or explicit aptitude when making judgments about English collocations.

What do these correlational studies show about the role of implicit aptitude in L2 attainment? It needs to be clarified that the research is limited and heterogeneous in learner characteristics and proficiency measures. Therefore, any patterns and conclusions are preliminary, suggestive, and subject to further verification. However, while it is difficult to reach conclusions, it is possible and useful to synthesize the findings and provide some meaningful interpretations. First, these studies show that implicit aptitude is predictive of L2 proficiency and therefore has the potential of opening up a new, fruitful area of research in language aptitude and SLA in general. Implicit aptitude tested by SRT tasks has been found to be predictive of general L2 proficiency (with the caveat that more information is needed on the proficiency tests) (Kaufman et al., 2010), L2 grammar (Suzuki & DeKeyser, 2015), and listening and reading comprehension (Linck et al., 2013). Implicit aptitude tested by LLAMA_D is predictive of speech fluency (Granena, 2019) and accuracy (Saito et al., 2019). Second, it would seem that implicit aptitude is more likely to show significant associations with L2 learning in instructed rather than naturalistic settings. This is probably because the naturalistic learners in relevant studies were recruited from the local community and had heterogeneous backgrounds. In instructed settings, however, learners are more homogeneous in that they are students in foreign language classes with similar ages and educational backgrounds. The homogeneity may have made it easier to find significant results. Third, it would seem that SRT is more likely to be correlated with L2 grammar (Suzuki & DeKeyser, 2015) and LLAMA_D with oral proficiency (Granena, 2019; Saito et al., 2019). Fourth, explicit aptitude is more likely to be drawn upon by beginning instructed learners (Saito et al., 2019) and naturalistic learners with shorter residence in the host country (Suzuki & DeKeyser, 2017), while implicit aptitude is more likely to be drawn upon by advanced instructed learners (Linck et
Explicit and Implicit Language Aptitudes

al., 2013; Saito et al., 2019) and naturalistic learners who live in the country for a longer period (Suzuki & DeKeyser, 2015).

Aptitude and Age

A few age-related studies investigated traditional aptitude as a variable to gain insights into the mechanism through which age affects L2 attainment. These studies include DeKeyser (2000), Abrahamsson and Hyltenstam (2008), Granena and Long (2013), and Granena (2013). As discussed previously, the theoretical basis for the relevance of aptitude to age is Bley-Vroman’s (1990) Fundamental Difference Hypothesis, which states that children rely on domain-specific language learning abilities (language acquisition device or LAD) that are not subject to individual variation and that lead to uniform learning success; adults, however, rely on analytic abilities or explicit language aptitude. The research has shown mixed findings about the hypotheses. Some studies showed that traditional/explicit aptitude is only predictive of the ultimate attainment of learners who start to be exposed to the second language after the critical period, not of the attainment of those who started to learn the language from childhood (DeKeyser, 2000; Granena & Long, 2013). Other studies found aptitude to be important for both early and late bilinguals (Granena, 2013; Abrahamsson & Hyltenstam, 2008). The hypothesis, however, has been confirmed from another perspective: among learners identified as having native-like proficiency, early starters’ aptitude scores are varied but late starters’ scores are uniformly high (Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000). This suggests that early starters, who have the magic LAD, all achieved high-level success regardless of their aptitude but late starters must have exceptional aptitude in order to achieve native-like proficiency because they have lost the LAD.

Finally, Granena (2013) has investigated the role of implicit aptitude in early and late bilinguals’ ultimate L2 attainment. The study reported that SRT predicted late bilinguals’ implicit knowledge and LLAMA_D predicted early bilinguals’ explicit knowledge, but the significant results were found only for linguistic structures involving agreement between sentence elements, not for non-agreement structures. The author interpreted the results as suggesting that early and late bilinguals “rely on similar implicit learning mechanisms” (p. 694), which runs counter to Bley-Vroman’s claim that late starters or adult learners have lost the LAD—a domain-specific implicit learning ability that only early starters have. There are two possibilities here. One is that late starters or adults have not totally lost the LAD. The other is that they have lost the LAD, and the implicit learning ability measured via SRT and LLAMA_D is different from the LAD. In educational psychology, the ability measured by SRT is conceived as a domain-general ability, which means it is not exclusively important for language learning. LLAMA_D is language specific in that the test involved sound recognition. However, it is unlikely to know whether it is the LAD. In Granena’s study, the early learners’ SRT and LLAMA_D scores were substantially higher than that of the late learners, which seems to suggest that the language learning experience enhances implicit aptitude. However, because both groups of learners were adults at the time when the study was conducted, the data provides no clue for answering the question whether younger and older learners are different in their implicit aptitude. Furthermore, a surprising finding is that LLAMA_D is correlated with explicit knowledge but not with implicit knowledge, which casts doubt on its validity as a measure of implicit aptitude—a point to be raised again in later sections.

Aptitude and Treatment Effects

Different from correlational research that examines how aptitude correlates with the final learning outcome regardless of treatment type and learning condition, the research on aptitude–treatment interaction seeks to investigate whether different instructional treatments or learning conditions implicate different cognitive abilities. As with aptitude research in general, most research in this stream has investigated traditional/explicit aptitude and there have been only several studies on
implicit aptitude. The following review will discuss the research on traditional/explicit aptitude and then implicit aptitude.

Traditional/explicit aptitude. The research on traditional aptitude has examined its associations with explicit/implicit oral feedback, explicit/implicit computerized instruction, written feedback, deductive and inductive instruction, the timing of form-focused instruction, and the nature of the linguistic target. To start with, corrective feedback refers to responses to learners’ errors in their oral and written production in a second language. Oral feedback can be classified as explicit and implicit, depending on whether the learner’s attention is overtly drawn to errors. Li’s (2017) meta-analysis showed that traditional aptitude is more strongly correlated with the effects of explicit feedback than implicit feedback. The stronger associations of traditional aptitude with explicit instruction are reinforced by studies on the role of aptitude in computerized instruction. In these studies, the explicitness of instruction was examined as an independent variable. For example, in Robinson’s (1997, 2002) studies, learners in the explicit condition were asked to process or search for rules, while learners in the implicit conditions were simply asked to memorize some sentences or answer a few comprehension questions. In Carpenter (2008), learners in the explicit condition were presented with the metalinguistic information about an artificial language, while those in the implicit condition were exposed to some exemplars of the language. These studies all demonstrated a stronger effect for aptitude under explicit learning conditions.

The distinction between explicit and implicit does not apply to written corrective feedback, whose corrective force is easily noticeable, especially when it targets linguistic errors. Written feedback falls into three types: direct, indirect, and metalinguistic. Direct feedback provides the correct form, whereas indirect feedback only draws attention to errors such as through highlighting, bold type, or circling; metalinguistic feedback provides comments on the well-formedness of the learner’s written production. Research on the role of aptitude in written feedback all focused on language analytic ability—an aptitude component for learning the morphosyntactic aspects of an L2. The research showed that 1) analytic ability was more predictive of the effects of direct rather than metalinguistic feedback (Benson & DeKeyser, 2019; Stefanou & Revesz, 2015), 2) it had stronger associations with new writing tasks than with revisions (Shintani & Ellis, 2015), and 3) it was more correlated with the effects of the treatment involving rewriting than with the treatment that did not require learners to rewrite (Shintani & Ellis, 2015). Shintani and Ellis pointed out that the role of analytic ability is evident in written feedback when the instruction requires deeper cognitive processing.

One stream of research has centered on whether traditional aptitude is implicated differently in deductive and inductive instruction (Erlam, 2005; Hwu & Sun, 2012; Hwu, Pan, & Sun, 2014). In deductive instruction, learners are presented with rule explanations followed by practice activities where the rules are applied. In inductive instruction, learners are encouraged to formulate rules based on given linguistic materials. The studies demonstrated that high-aptitude learners benefited more from inductive instruction than deductive instruction, whereas the reverse was true for low-aptitude learners—they benefited more from deductive instruction than inductive instruction. These studies also showed that aptitude was more strongly correlated with the effects of inductive instruction that requires learners to utilize their own cognitive resources to extrapolate linguistic regularities than deductive instruction that provides external support in the form of rule explanation. These studies suggest that providing explicit grammar instruction may favor learners with low aptitudes but it may disadvantage learners with high aptitudes, and that the role of aptitude is evident when less external support is available. However, there are appreciable differences between the two studies. For example, in Erlam (2005), significant effects were found for phonetic coding, but not for analytic ability or working memory. In Hwu and Sun (2012) and Hwu et al. (2014), significant effects were found for memory for text, but not for analytic ability or rote memory. More research is needed to further examine the associations of aptitude with inductive and deductive instruction.
The claim that aptitude is important when less support is available is confirmed by Li et al. (2019), who investigated the interface between language analytic ability and the timing of form-focused instruction. The study involved 150 eighth-grade EFL learners divided into five treatment groups depending on whether and when they received form-focused instruction. One group received pre-task grammar instruction; a second group received within-task corrective feedback; a third group received both pre-task instruction and within-task feedback; a fourth group only received post-task feedback; and the fifth group only performed communicative tasks without receiving any form-focused instruction. Analytic ability was predictive of the effects of the group that only performed communicative tasks and the group that only received post-task feedback. The researchers argued that the results were due to the lack of form-focused instruction for the task-only group and less form-focused instruction for the post-task feedback group.

The role of aptitude is also constrained by the nature of the linguistic target. Li (2013a, b) reported that when metalinguistic feedback was provided, language analytic ability was not important for the learning of Chinese classifiers, an easy and transparent structure, because the feedback leveled off the role of aptitude. However, analytic ability was implicated in the learning of the Chinese perfective -le, a difficult and opaque structure, because the processing of the metalinguistic explanation of a difficult structure necessitates the use of analytic ability. Yalçın and Spada (2016) explored the association between analytic ability and the effects of form-focused instruction consisting of rule explanation followed by communicative practice in the learning of two linguistic structures in English: the passive voice and the past progressive. They found a significant correlation between aptitude and the learning of the passive voice but not the past progressive, and they argued that this is because the former structure is more difficult than the latter.

Implicit aptitude. Based on the same dataset, Granena and Yilmaz (2019) and Yilmaz and Granena (2019) investigated whether implicit aptitude tested via SRT was predictive of the effectiveness of implicit (recasts) and explicit feedback (explicit correction) in the learning of Spanish gender agreement and differential case marking by intermediate L2 Spanish learners at a US university. Learning gains were measured via a self-paced reading task, an oral production task, and a grammaticality judgment test. The studies showed that implicit aptitude was predictive of the effects of implicit feedback in the learning of gender agreement on the self-paced reading task and the grammaticality judgment test, but not on the oral production task. The studies further found that phonological short-term memory, an explicit learning ability, was predictive of the effects of explicit feedback in learning gender agreement. These two studies are the first to examine the role of implicit aptitude in different instructional treatments in SLA, and their findings contribute significantly to the advancement of the research on implicit aptitude. They confirmed the hypothesis that implicit aptitude is implicated in implicit instruction rather than explicit instruction. One unexpected finding of these two studies is that implicit aptitude is predictive of learners’ grammaticality judgment, which purports to measure explicit knowledge, but not oral production, which is regarded as a measure of implicit knowledge.

Several studies investigated the relationship between implicit aptitude measured by SRT and incidental learning, both involving the learning of an artificial language based on L1 words and L2 morphosyntax. The studies showed disparate results. In Hamrick (2015), 31 native speakers of English underwent a 20-minute treatment where they read sentences built with English words and three Persian structures, followed by a recognition test asking them to identify which sentences were new and which had been presented in the treatment. The learners took a delayed post-test two weeks later. They were also administered the LLAMA_B (rote memory) and an SRT. It was found that LLAMA_B was correlated with the learners’ immediate gains, while SRT was correlated with their delayed post-test scores. The researcher interpreted the results as suggesting that declarative memory facilitated initial L2 learning and procedural memory enhanced learning at later stages. Note that rote memory and sequence learning (SRT) are called declarative and procedural memory respectively in the declarative-procedural research paradigm. Using a similar
Shaofeng Li

design, Tagarelli et al. (2015) recruited 50 native speakers of English from the UK and the US. The participants received training on an artificial language based on English words and three German structures with different levels of difficulty. They were divided into two groups: incidental and instructed. In the incidental condition, they listened to sentences in the artificial language, judged whether the sentences were meaningful, and repeated each sentence after making a judgment. In the instructed condition, they were taught grammar rules followed by practice activities. Treatment effects were tested using a grammaticality judgment test. The learners also took an SRT and a working memory test. The only significant result was a strong negative correlation ($r = -0.54$) between the SRT and the most difficult structure under the incidental learning condition. The researchers indicated that the incidental treatment encouraged explicit learning and the structures were salient. Thus, the study suggests that implicit aptitude may have a negative effect if learners engaged in conscious processing of a complex structure.

**Data Elicitation**

**Research Design**

There are three basic research designs in aptitude research, which can be loosely called construct validation, predictive research, and aptitude–treatment interaction. Studies on construct validation seek to identify the components or composition of aptitude and to map the relationships between the components and between aptitude or aptitude components and other cognitive and affective variables. The primary objective of this strand of research is to seek evidence for the convergent and divergent validity of aptitude, that is, aptitude and aptitude components should correlate or behave similarly with variables in the same or similar domains and be uncorrelated or behave differently from variables in other domains. To achieve the objective, one would need to test the variables of interest and perform statistical analyses of a correlational nature, such as Pearson’s correlation, factor analysis, and structural equation modeling, to identify the relationships between the tested variables. For example, in order to explore the relationship between working memory and traditional aptitude, Yalçın et al. (2016) tested Turkish EFL learners’ working memory using two reading span tests and one operation span test and their aptitude using the LLAMA test battery. Correlation and factor analysis showed that measures of aptitude and working memory loaded on separate factors, suggesting that they are distinct constructs. The results provide evidence against the call to consider working memory an aptitude component or to replace traditional aptitude with working memory. One promising area of research is to validate the construct of implicit aptitude by showing that measures of implicit aptitude converge and that they diverge from measures of explicit aptitude or other cognitive abilities for explicit learning, such as working memory and attention control. Despite its importance, construct validation is normally integrated with other designs such as predictive and ATI research where the researcher normally maps the relationships between predictor variables before exploring whether they are associated with learning outcomes.

The design of predictive research is straightforward. The primary objective is to determine whether aptitude is predictive of or correlated with a learning outcome. In this type of research, the researcher obtains two sets of scores, one for aptitude and one for learning outcome, and then runs correlational or regression analyses to see whether aptitude has a significant correlation or is significantly predictive of learning outcome. The predictor variable can be overall aptitude represented by composite scores on whole test batteries (LLAMA, MLAT, etc.) or aptitude components operationalized as learners’ scores on the subtests of a test battery, such as language analytic ability measured by LLAMA_F and MLAT_4. The outcome variable can be overall L2 proficiency or aspects of L2 proficiency. Overall proficiency can be gauged through proficiency tests, such as TOEFL, cloze tests, etc., or through course grades (midterm or final exam scores). Li (2016) found no difference between proficiency tests and course grades in terms of aptitude’s predictive power.
Explicit and Implicit Language Aptitudes

Specific aspects of L2 proficiency include L2 skills—listening, speaking, reading, and writing—and L2 knowledge—pronunciation, grammar, and vocabulary.

Among predictive studies, age-related studies stand out from other studies. These studies all involve naturalistic learners who have lived in the country of the L2 for at least ten years, which is of no surprise because the purpose is to see the influence of starting age on ultimate attainment. These studies have been conducted following two templates. One is to recruit L2 learners who pass for native speakers based on rated speeches or interviews and divide them into different groups according to the age at which they start to learn the L2 or arrive in the country (Abrahamsson & Hyltenstam, 2008; Granema & Long, 2013). One way to analyze the data based on this template is to determine whether aptitude is correlated with L2 attainment differently as a function of starting age. Alternatively, the researcher may explore whether early starters’ aptitude scores are spread out while late starters’ scores are uniformly high to see whether later starters have to rely on analytic abilities to reach a high level of attainment. In the other template, the ultimate attainment is not controlled but, as with the first type of design, the learners are divided into different age groups to see whether aptitude correlates with their L2 attainment differently depending on the age group. The outcome measures in these studies are grammaticality judgments, pronunciation ratings, and lexical knowledge.

Aptitude–treatment interaction seeks to ascertain whether aptitude and aptitude components are implicated differently in different types of instruction. Therefore, the two basic elements of this type of research are aptitude and treatment. Typically, learners receive different types of treatment, such as different types of corrective feedback targeting a certain linguistic structure. They are given a pre-test before the treatment and a post-test after the treatment to measure treatment effects (some studies also include one or more delayed post-tests). They are also tested on their language aptitude. Then a correlation or regression analysis is performed for each group to determine whether aptitude or aptitude components are predictive of treatment effects represented by gain scores (post-test scores subtracted by pre-test scores) or post-test scores. If post-test scores serve as the dependent variable, pre-test scores should be entered as a covariate to see whether aptitude explains a unique portion of the variance of the post-test scores after pre-test scores are accounted for. Alternatively, learners can be divided into high and low aptitude based on their aptitude scores, in which case there are two categorical variables: high/low aptitude and treatment type. An ANOVA analysis can be performed to identify main and interaction effects.

Learner Characteristics

Among the 66 samples included in Li’s (2016) meta-analysis of predictive aptitude studies, 22 were high school foreign language learners and 26 were university students. Thus, there is a lack of research with other learner groups such as children, middle school learners, and naturalistic learners. Li pointed out that most of the learners in predictive research received traditional form-focused instruction, and there is a lack of research on learners that receive meaning-oriented instruction, such as immersion and communicative instruction. Form-oriented instruction favors traditional/explicit aptitude, hence the strong correlations between aptitude and learning outcomes yielded by these studies. Although it is possible to manipulate instruction type in experimental research (Li, 2015), the heavy form-focused instruction learners receive in the local context may preempt them to linguistic forms and encourage them to utilize their aptitude, increasing the chances of finding significant results for explicit aptitude. This poses a challenge for implicit aptitude, which is posited to be implicated in implicit learning. Existing research on implicit aptitude shows that it is more likely to be drawn on by advanced learners and learners who live in the host country for longer periods. Therefore, research investigating implicit aptitude needs to ensure that the learners are advanced enough or have a certain amount of implicit knowledge. Learners in age-related studies typically have a length of residence in the host country.
of over ten years. For the purpose of investigating aptitude effects for different age groups, they are typically divided into early or late starters based on their age of onset, defined as the age at which one arrives in the country or starts to learn the L2 formally. Age groups formed based on age of onset in the research include 1) >/< 12 and 2) 3–6, 7–15, and >16 (Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000; Granena & Long, 2013). In the first scheme, 12 is considered a cut-off point for the critical period, and in the second, there are multiple sensitive and critical periods for different aspects of L2 attainment.

Aptitude Measures

It is self-evident that the measurement of aptitude is an important aspect of aptitude research because, to a certain extent, aptitude is defined as what is measured by aptitude tests. Table 2.1 lists major tests of explicit and implicit aptitudes. The most influential test of explicit aptitude is the MLAT (Carroll & Sapon, 2002). The test targets native speakers of English and was validated with over 5,000 foreign language learners in the US. Li’s (2015, 2016) meta-analyses showed that the MLAT demonstrated higher predictive validity than other aptitude tests including those adapted from the MLAT. The test consists of five subtests measuring three abilities: phonetic coding, analytic ability, and rote memory. Among the five subtests, the components measured by MLAT_1 and MLAT_3 are not straightforward, although they have strong predictive validities (Li, 2016). The whole test takes about one hour to complete, and the short version (MLAT_3, _4, and _5) whose predictive validity is equivalent to the full battery, takes about 30 minutes. The test is copyrighted by the Language Learning and Testing Foundation (LLTF) and is currently unavailable to individual researchers. However, a computerized version is under development and will potentially be open to researchers in the near future. Modeled on the MLAT, the LLAMA (Meara, 2005) is probably the most frequently used test in recent aptitude research, partly because the MLAT is no longer available. It measures the same components as the MLAT, is freely available electronically, generates scores automatically, is language neutral, and only takes about 30 minutes to complete. Although the LLAMA has demonstrated predictive validity, it has not been validated. Recently, a study by Bokander and Bylund (2020) reported that the LLAMA subtests, except for LLAMA_B, lack internal validity. Therefore, researchers should exercise caution in interpreting the results based on LLAMA. The PLAB (Pimsleur et al., 2004), which is intended for learners in grades 7–12, was validated with 6,000 foreign language learners and has demonstrated high predictive validity. However, among the six subtests or sections, only four to six measure aptitude, and the other three—GPA, motivation, and vocabulary—do not, although they may be correlated with language learning. PLAB 4 measures inductive ability (also measured by LLAMA_F), which is one of the two components of language analytic ability, the other being grammatical sensitivity, which is measured by MLAT_4. The other two PLAB tests gauge phonetic coding. The PLAB is also managed by the LLTF but unlike the MLAT, it is available for purchase by researchers. The Hi-LAB was developed to measure abilities for high-level proficiency. It includes 12 tests measuring nine constructs, featuring a heavy emphasis on working memory (three executive functions and phonological short-term memory) and inclusion of implicit aptitude (SRT). However, the validation study only identified three significant predictors: implicit aptitude, phonological short-term memory, and rote memory, which is a component of traditional aptitude.

To date, SRT and LLAMA_D have been employed as measures of implicit aptitude. In one variant of SRT (e.g., Granena, 2013), subjects respond to a symbol, such as a dot, that appears in different locations on the computer screen following a fixed sequence (e.g., 1214321432), called the target sequence. A control sequence (e.g., 124312413234) is also built in the stimuli. The target sequence appears more frequently (e.g., 85%) than the control sequence (e.g., 15%). Learners’ performances are calculated by subtracting their mean reaction time for the target stimuli from the mean reaction time for the control stimuli. While the SRT measures the ability to learn visually presented sequences, LLAMA_D is a measure of the ability to learn phonological sequences.
Table 2.1 Major Aptitude Tests

<table>
<thead>
<tr>
<th>Tests</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit Aptitude</td>
<td>- Part 1: Number learning. It measures memory and auditory alertness. Learners are asked to learn numbers in a new language.</td>
</tr>
<tr>
<td>MLAT</td>
<td>- Part 2: Phonetic script. It measures phonetic coding. Learners learn sound-symbol associations.</td>
</tr>
<tr>
<td></td>
<td>- Part 3: Spelling clues. It measures English vocabulary and phonetic coding. Learners answer questions about English vocabulary.</td>
</tr>
<tr>
<td></td>
<td>- Part 4: Words in sentences. It measures grammatical sensitivity. Learners are asked to identify linguistic functions of sentence elements.</td>
</tr>
<tr>
<td></td>
<td>- Part 5: Paired associates. It measures rote memory. Learners are asked to memorize words and their meanings.</td>
</tr>
<tr>
<td>LLAMA</td>
<td>- LLAMA_B measures rote memory. Learners are asked to memorize shapes and their pronunciations.</td>
</tr>
<tr>
<td></td>
<td>- LLAMA_D measures phonetic recognition. Learners listen to some syllables and then discriminate between old and new syllables.</td>
</tr>
<tr>
<td></td>
<td>- LLAMA_E measures phonetic coding. Learners memorize symbols and their pronunciations.</td>
</tr>
<tr>
<td></td>
<td>- LLAMA_F measures inductive learning ability/language analytic ability. Learners see pictures as well as related sentences and learn grammar rules.</td>
</tr>
<tr>
<td>PLAB</td>
<td>- Part 1: GPA. Learners report their GPAs.</td>
</tr>
<tr>
<td></td>
<td>- Part 2: Motivation. Learners report the intensity of their interest in learning a foreign language.</td>
</tr>
<tr>
<td></td>
<td>- Part 3: Vocabulary. Learners are tested on their English vocabulary.</td>
</tr>
<tr>
<td></td>
<td>- Part 4: Language analysis. It measures inductive learning ability (language analytic ability). Learners learn rules of an artificial language and then answer questions applying the rules.</td>
</tr>
<tr>
<td></td>
<td>- Part 5: Sound discrimination. It measures the ability to distinguish sounds. Learners are asked to learn three words that sound similar. They then hear sentences and tell which word is heard.</td>
</tr>
<tr>
<td></td>
<td>- Part 6: Sound–symbol association. It measures phonetic coding. Learners hear nonsense words and choose the correct spelling of the word based on their knowledge about English pronunciation.</td>
</tr>
<tr>
<td>Hi-LAB</td>
<td>- Executive control</td>
</tr>
<tr>
<td></td>
<td>a. Updating. Measured using a running memory span test where learners are presented with digit strings and recall the last x number of digits.</td>
</tr>
<tr>
<td></td>
<td>b. Inhibition. Measured using an antisaccade test and a stroop test. In the antisaccade task, learners are asked to look in the opposite or same direction of a target. The score is the difference between the reaction times for the same and opposite conditions. In the stroop test, learners respond to words for three colors: red, green, and blue. For some words, the color and the word are congruent; for others, they are incongruent. Learners are asked to ignore the meaning of the word and only respond to the color. Scores are based on reaction time differences between congruent and incongruent words.</td>
</tr>
<tr>
<td></td>
<td>c. Shifting. Measured using a switching test where learners are asked to perform two tasks. The score is the reaction time difference between the switching and non-switching trials.</td>
</tr>
<tr>
<td></td>
<td>- Phonological short-term memory. Measured through a letter span test where learners are presented with lists of letters and asked to recall, and through a non-word test in which learners see lists of nonwords, after each list, they are shown another list and asked to recognize which words are on the list just presented.</td>
</tr>
</tbody>
</table>

(Continued)
The test includes a study phase and a testing phase. In the study phase, the subject listens to some monosyllabic sound sequences, and in the testing phase they listen to a mix of old and new syllables and are asked to recognize whether they appeared in the study phase. Subjects' performances are scored based on whether their judgments are accurate. As previously discussed, LLAMA_D has been found to be an inconsistent measure of implicit aptitude, probably because of the way it is administered. The test requires the learner to listen to recorded sounds and is informed of a test following the listening task, which may encourage explicit learning. Therefore, it is necessary not to inform learners of the testing phase and encourage them not to focus their attention on the phonetic features of the sounds, such as by instructing them to listen to the auditory stimuli to check the volume (Saito et al., 2019). Also, LLAMA_D is more likely to predict speech performance than other aspects of L2 proficiency (Granena, 2019; Saito et al., 2019), which is of no surprise given the nature of the test stimuli.

Practical Applications

Because of the strong, consistent predictive power of language aptitude demonstrated by the various streams of research, the findings of aptitude research have valuable practical implications and can be effectively applied in classroom teaching and policymaking. The implications can be discussed from two perspectives—testing and research—although it is difficult to entirely tease them out because test validity must be established by research, and the validity of research findings depends on test validity. From the standpoint of testing, learners’ aptitude scores can be utilized for a multitude of purposes. First, aptitude scores can be used to select learners that fulfill the mission of a language program or class, especially if the selected learners are expected to achieve a high level of success within a limited amount of time. Second, aptitude scores, together with learners’ proficiency levels, can be used to place students into different sessions to make sure they progress at a similar rate. Students may also be placed into different levels or streams depending on their
aptitude levels such that students of different aptitude levels all make maximal achievements. Third, aptitude tests can serve diagnostic purposes to identify learners’ strengths and weaknesses in certain components. Teachers and policy makers can then provide informed advice, tailored assistance, and appropriate accommodation. For example, if a learner has a low score in memory but a high score in analytic ability, then it will be necessary to advise the learner to engage in activities that fit his/her profile. If a learner gets a low score in phonetic coding, the teacher may provide more assistance to the learner in pronunciation learning. If a learner scores below 20% according to a certain norm (see MLAT manual), then the institution may consider waiving or lowering foreign language requirements for the learner (e.g., taking one year of class rather than two years).

Aside from test results, research findings may also guide practice. The age-related studies show that children and adults may learn differently. So far, the research seems to show that children rely on domain-specific LAD and adults draw on explicit aptitude. Therefore, it is necessary to teach children and adults differently, such as using implicit instruction for children and incorporating form-focused instruction for adults. However, research (Abrahamsson & Hyltensae, 2008; Granena, 2013; Roehr-Brackin & Tellier, 2019) seems to suggest that both children and adults draw on both explicit and implicit aptitudes, which poses a challenge to previous findings and theories on L2 learning. With the caveat that more research is needed on the cognitive processes of child and adult learning, it seems advisable for teachers to use a mix of implicit and explicit instruction to fit students with either propensity and maximize learning gains for all learners because, for all learners, learning occurs in both ways: explicitly and implicitly. Exclusive use of either explicit or implicit instruction would result in a waste of learners’ cognitive resources in the other pool. In a similar vein, it seems advisable for teachers to mix deductive and inductive instruction given that they favor low- and high-aptitude learners, respectively. Note that the idea of using mixed instructional approaches suits settings where it is challenging to cater to learners with particular strengths or weaknesses. In situations where it is possible to streamline learners based on their aptitude profiles, it would be feasible to implement certain instruction types systematically that match learners’ profiles. Finally, although it is challenging to adapt instruction in the classroom, it is certainly possible to provide counseling to individual learners based on research findings. For instance, for learners strong in implicit aptitude but weak in explicit aptitude, it would be more effective to engage in learning activities or adopt strategies that implicate implicit aptitude. However, the idea of matching aptitude profiles and learning styles is an intuitively appealing idea, and empirical evidence is lacking regarding whether matching indeed brings benefits to learning outcomes.

Future Directions

Current research has been dominated by explicit aptitude, and research on implicit aptitude will contribute significantly to the advancement of aptitude research and SLA research in general. Implicit aptitude is an essential piece of the tapestry of SLA, and the addition of implicit aptitude will enhance our understanding of the mechanism of L2 learning by informing us of how implicit and explicit aptitudes jointly and independently contribute to the process and outcome of L2 learning. Given that it is a new concept, there is an urgent need for more research on its conceptualization, measurement, and validation. It is also unclear what components implicit aptitude entails, and there has been little theorization on its mechanism and composition. To what extent implicit aptitude is language specific is unknown. Currently, SRT and LLAMA_D have been used to measure implicit aptitude. However, SRT is semantically vacuous and does not involve form-meaning mapping, which is probably why it is only predictive of agreement structures but not structures involving form-meaning mapping in Granena’s studies (e.g., Granena, 2013). The evidence for LLAMA_D as a measure of implicit aptitude is meager and inconsistent.

There is also a need for more research on explicit aptitude. Research on explicit aptitude started in the 1950s, which seems to suggest a large amount of research. Li’s (2016) meta-analysis collected
66 studies, but when the studies were categorized based on aptitude components and aspects of learning such as L2 knowledge and skills, the number of studies for each category was small, mostly below ten. There is also a lack of research on factors mediating the role of aptitude, such as learners’ proficiency level, age, and instructional setting. With respect to aptitude–treatment interaction, most of the research focused on language analytic ability and a limited number of instruction types, such as corrective feedback and inductive and deductive instruction. Therefore, although there has been a substantial amount of research, more focused research is needed to arrive at a more precise understanding of the role of aptitude.

References


Explicit and Implicit Language Aptitudes


53