RESEARCH ARTICLE

The prediction from MLAT to L2 achievement is largely due to MLAT assessment of underlying L1 abilities

Richard L. Sparks[1]* and Philip S. Dale[2]

[1]Mt. St. Joseph University, Cincinnati, OH, USA; [2]University of New Mexico, Albuquerque, NM, USA
*Corresponding author. E-mail: Richard.Sparks@msj.edu

Abstract

Widespread use of the Modern Language Aptitude Test (MLAT) in L2 studies of individual differences implicitly assumes that L2 aptitude is a distinct cognitive facet. There is considerable evidence for prediction from L1 abilities to L2 learning. In this longitudinal study, L1-MLAT-L2 relations were examined in 307 US secondary students based on six L1 and six L2 measures of language and literacy, and the MLAT. Mediation and regression analyses revealed that each L1 measure individually predicted all L2 scores and MLAT; the L1 measures collectively substantially predicted MLAT scores; MLAT is a significant but moderate mediator of prediction from L1 to L2 scores; and prediction from MLAT to L2 scores is significantly and substantially due to variance in L1 abilities captured by MLAT. Overall, prediction from MLAT is due primarily to its functioning as a measure of L1 abilities, although substantial L1 variance which predicts L2 scores is not captured by the MLAT.

Introduction

Aptitude for learning a foreign, or second, language (L2) as a predictor of L2 learning outcomes has been a topic of interest for almost a century. The first attempt to develop L2 aptitude tests occurred in the 1920s and 1930s when language specialists developed "prognosis" tests to determine who might benefit from L2 instruction and how one would perform in L2 learning (e.g., see Hunt et al., 1929; Luria & Orleans, 1928). Symonds (1930) developed a prognosis test based on three components: L1 ability, general intelligence, and "quick learning" tests in the target language. During World War II, the US Army funded a study by Dorcus et al. (1953), who developed a test of "major aptitude skills."

More rigorous and theoretically focused study of L2 aptitude commenced with the publication of the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959 [2001]). Carroll proposed that language aptitude is a specialized talent, or group of talents, independent of intelligence, and found that four independent variables were most relevant to L2 learning: phonetic coding, grammatical sensitivity, inductive

language learning ability, and rote memory (Carroll, 1962, 1990). Even with the development of other language aptitude batteries developed by Pimsleur (1966) as well as newer batteries such as the Hi-LAB (Linck et al., 2013) and LLAMA (Meara, 2005), the MLAT has largely dominated aptitude research since the 1960s.

Recently, Li (2019) provided an update of research on language aptitude and reviewed theories underlying aptitude research. One of those theories, the Linguistic Coding Differences Hypothesis (LCDH; Sparks, 1995; Sparks & Ganschow, 1993a, 1995), proposes that L1 and L2 learning draw on the same pool of cognitive abilities and that language aptitude is componential. In addition, the theory proposes that there are large and stable individual differences (IDs) in L1 ability developed prior to L2 exposure, and these differences are reflected in students' IDs in L2 aptitude and L2 achievement. Sparks's and Ganschow's longitudinal investigations over 10 years found that students' L1 literacy skills and L1 vocabulary knowledge in elementary school explained from 65% to 72% of the variance in L2 aptitude on the MLAT measured several years later in ninth grade (Sparks et al., 2006). Even so, most studies have found that the MLAT is the strongest single predictor of L2 achievement, even when L1 variables were also being examined.

In the present research, we propose that mapping the quantitative relations among L1 and L2 abilities and L2 aptitude would benefit by distinguishing two aspects of prediction from L2 aptitude measures such as the MLAT to L2 achievement. The first aspect, *uniqueness*, is the extent to which the L2 aptitude measure uniquely *adds* to the prediction of L2 outcomes beyond that explained by L1 measures alone. The second aspect, *efficiency*, is the extent to which the total prediction from a specific L1 measure to L2 is *mediated* by L2 aptitude. The importance of these definitions is that they are in principle independent: *uniqueness* and *efficiency* could be both high or low, or either one could be high or the other low. Further, *uniqueness* and *efficiency* may be influenced by diverse features of the L1 measures and/or the L2 aptitude measure.

In the review, we briefly examine the L2 aptitude concept and research on L2 outcomes with the MLAT. Next, we review the literature on IDs in L1 development and attainment, and research on the relationships between early L1 ability and L2 aptitude. Then, we review more recent research on the relationships among IDs in L1 ability, L2 aptitude on the MLAT, and L2 achievement.

L2 aptitude and the MLAT

The MLAT was developed by John Carroll and Stanley Sapon and published in 1959. Carroll based his investigations into language aptitude on the proposition that the facility to learn a L2 is a specialized talent independent of intelligence, based on empirical findings that IQ tests had been relatively unsuccessful in screening who would and would not be successful in language training. Carroll's impression that IQ tests were not predictive of L2 proficiency was confirmed by later work of other researchers (e.g., Gardner & Lambert, 1965; Sasaki, 1996). Through factor analytic studies, he found that four factors were most relevant for L2 learning: (a) phonetic coding, or the ability to code and remember phonetic material and sound-symbol relationships over time; (b) grammatical sensitivity, or the ability to recognize the grammatical function of words; (c) inductive language learning ability, or the ability to infer or induce linguistic forms, rules, and patterns from examples; and (d) rote memory, or the ability to learn associations between native language words and foreign language equivalents (Carroll 1962 [1990]). The MLAT assessed these four

components using five different subtests. Stansfield and Reed (2019) noted that the MLAT subtests were not designed to tap into Carroll's four components in a "one-to-one fashion" but instead were designed to "work together" to predict L2 proficiency. The MLAT has been found to be a reliable predictor of language learning success, demonstrating correlations of .40–.65 with end-of-year course performance in intensive L2 instruction (Skehan, 1998). For example, studies at the US Foreign Service Institute found comparable correlations between MLAT scores and language learning outcomes (Ehrman, 1998; Ehrman & Oxford, 1995).

Li (2015, 2016, 2017) has conducted three meta-analyses/research syntheses of empirical studies on various aspects of language aptitude. In a recent publication, he provided an overview of the research on language aptitude, including but not limited to the MLAT, collected over the previous 60 years (Li, 2019). Overall, aptitude measured with a composite score based on test batteries like the MLAT correlates about .50 with overall L2 proficiency. When L2 aptitude composite scores were examined in relationship to specific L2 skills, for example, reading, writing, listening, speaking, and findings revealed correlations from .30–.39. But there were only weak correlations between L2 aptitude and L2 vocabulary (.15) and no significant correlations with L2 writing. However, certain subtests on the MLAT were more strongly predictive of specific L2 skills. For example, the Phonetic Coding subtest was a significant predictor of L2 vocabulary learning (.38) and the Number Learning and Spelling Clues subtests were significant predictors of L2 writing (.42). Li suggested that researchers should examine aptitude components rather than overall aptitude.

In sum, language aptitude measures, including the MLAT, have been found to be predictive of L2 proficiency, and also more predictive of L2 outcomes than other ID variables.

Individual differences in L1 development and attainment

IDs in the development of oral language skills have been well-known for some time (Brown, 1973). For example, Bloom and Lahey (1978) presented evidence showing that developmental variation in children's language is a "fact that can be taken pretty much for granted" (p. 165), marking a shift from emphasis on commonalities (universals) to one that included variation. In an influential book, Bates (1988) reviewed investigations showing that while there are regularities in childhood language development, there are also important differences among (interindividual) and within (intraindividual) children in most aspects of language development. In a comprehensive report of the first very large-sample study of early development, Bates, Dale, and Thal (1995) documented variation in the early development of normal children that were substantial and stable in gestural communication, word comprehension and production, and first stages of grammar. They concluded that there are "enormous individual differences in onset time and rate of growth in each of these components" (p. 1). The variations are stable and cannot be explained by a single causal factor. Further, the relations among skills, such as language comprehension and production can be variable as well (Fenson et al., 2007).

Although most children learn to communicate in their L1, that is, talk and listen effectively in everyday contexts, more recent research with increasingly large and representative samples has continued to find variation in the onset and rate of acquisition across all components of the language system (Gilkerson et al., 2017; Huttenlocher et al., 2010). IDs in L1 skills are both substantial and stable across

development (Bornstein & Putnick, 2012), and these differences are strongly related to later acquisition of L1 literacy skills (Kendeou et al., 2009). In a recent review, Kidd and Donnelly (2020) concluded that IDs in first language proficiency are (a) a pervasive feature of language development, (b) the norm rather than the exception, (c) large and stable across development, and (d) observed across all domains of language development. Moreover, these IDs in early language developments do not disappear after childhood and are observed among typically developing adults in their ultimate ability.

In sum, converging evidence over several years has found that there are IDs in early L1 ability in all components of language development, IDs in language ability are both large and stable over time, and early IDs in language skills predict later language outcomes.

Individual differences in L1 development and L2 aptitude

The Bristol Language Project (Wells, 1985) was a pioneering study of IDs in L1. Children born in Bristol, United Kingdom were followed for several years. The results revealed wide variation in the speed at which children acquire their first language. For example, some children reached a point in language development well in advance of others by ages 3–4 years, while other children whose language development was the slowest fell more than several years behind them.

In his seminal work on language aptitude, Skehan (1986) followed students from the Bristol Project to the time they entered L2 classes at 13–16 years of age, which allowed him to study connections between the students' L1 development and their subsequent L2 aptitude (York Language Aptitude Test, Elementary MLAT [two subtests], Pimsleur Language Aptitude Battery [two subtests]) and L2 achievement, that is, Skehan's "triangle of relationships" (1989, p. 32), depicted in Figure 1. The children's L1 development and attainment prior to age 5 were strongly correlated with their L2 aptitude and L2 achievement several years later in secondary school (Skehan & Ducroquet, 1988). Their results also revealed that while L2 aptitude scores in secondary school were predicted "reasonably well" by L1 indices, L2 aptitude scores were more successful predictors of L2 achievement. Nonetheless, prediction of L2 achievement by the L2 aptitude tests was improved by specific L1 achievement measures, notably vocabulary growth and language comprehension. Skehan and Ducroquet concluded that the L2 aptitude tests "captured the useful predictive variance of many of the first language indices … and so preempt them" for predicting L2 achievement (p. 102).

Skehan et al.'s findings suggested that L2 aptitude tests are predictive of L2 achievement because their items "partly measure an underlying language learning capacity which is similar in first and foreign language learning." He further hypothesized that
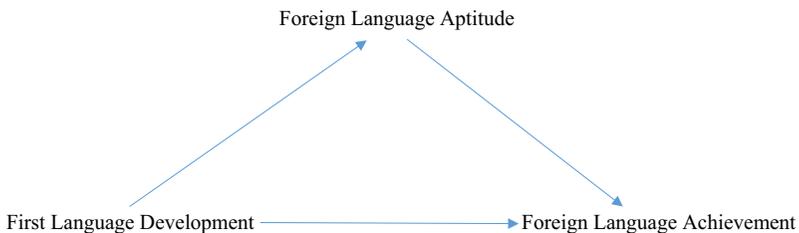

Figure 1. Skehan's triangle of relationships for the study of connections of L1 development, L2 aptitude, and L2 achievement.

aptitude tests' "main emphasis is probably to function as a measure of the ability to learn from decontextualized material" (p. 34). On aptitude tests, students must be able to go beyond everyday "real life" L1 activities to use their language analytic abilities to "think" about how language works.

In sum, findings from the Bristol Language Project showed that there are significant relationships among early L1 ability and later L2 aptitude and L2 achievement. The findings suggest that L2 aptitude tests measure skills similar to those mastered in a student's L1, for example, sound-symbol relationships, knowledge of grammar, L1 vocabulary, but that L2 aptitude tests predict L2 achievement better than L1 measures because of their capacity to measure material that has been isolated from a context.

Relationships among L1 ability, L2 aptitude, and L2 achievement

Historically, SLA/L2 researchers have been engaged primarily in searching for universal characteristics and processes of language development (see Dabrowska, 2016). However, there is growing evidence that there are large IDs in adult L1 speakers' linguistic skills due to internal and external factors that have important implications for L2 research. For example, a recent issue of *Language Learning* was devoted to studying IDs in L1 and L2 attainment from different perspectives (see summary by Dabrowska, 2019). These developments are unsurprising because it is well-known that students with stronger L1 oral and written language skills exhibit stronger L1 reading achievement and display more positive educational outcomes than students with weaker L1 skills (e.g., see Bleses et al., 2016; Hayiou-Thomas et al., 2010).

However, other than Skehan's investigations, by 1990 little research had examined the relationships among IDs in L1 skills, L2 aptitude, and L2 achievement. To investigate these relationships, Sparks and Ganschow (1991, 1993a) authored the LCDH, which proposes that L1 skills are a foundation for L2 learning, the primary causal factors in more and less successful L2 learning are linguistic, and IDs in L1 explain ultimate ability in the L2 (see also Sparks, 1995). The LCDH also posits that IDs in students' L1 skills are related to and consistent with their aptitude for L2 learning. The tenets of the LCDH are similar to Cummins's (1979) Linguistic Interdependence Hypothesis (L1 and L2 have a common underlying foundation) and Linguistic Threshold Hypothesis (L2 proficiency is moderated by one's level of ability in L1).

Starting in 1990, Sparks's and Ganschow's studies with secondary and postsecondary US L2 learners have generated broad empirical support for the LCDH by showing that there are strong relationships between IDs in L1 ability and IDs in L2 aptitude and L2 achievement. In their longitudinal investigations covering from 3 to 10 years, they have found that high-, average-, and low-achieving L2 learners exhibit significant IDs in their L1 skills in elementary school as early as second grade and in L2 aptitude on the MLAT; L1 skills, especially L1 literacy, are strong predictors of L2 aptitude and L2 achievement; and IDs in learners' L2 aptitude are robust predictors of L2 achievement. (Comprehensive reviews of these studies and those cited in the following text can be found in Sparks [2012], Sparks & Patton [2013], and Sparks et al. [2019].) Their factor analyses have shown that L2 aptitude as measured by the MLAT is comprised of different language components, and that language skills measured by the specific MLAT subtests load with similar L1 skills (e.g., MLAT Phonetic Coding subtest and L1 phonetic word and pseudoword decoding tests). Notably, L1 skills and the MLAT together explain from 67% to 76% of the variance in overall L2 proficiency. Studies investigating relationships between L1 print exposure and L2 achievement found that

L2 learners who displayed significant IDs in L1 reading volume also exhibited significant differences in L2 aptitude and L2 achievement, and that IDs in L1 print exposure contribute unique variance to L2 skills even after controlling for L1 literacy, L1 verbal skills, and L1 cognitive ability in primary school, and for L2 aptitude in ninth grade. These longitudinal studies have also found support for L1-L2 cross-linguistic transfer. Other researchers have also found strong relationships between IDs in L1 skills and L2 learning (e.g., see Dufva & Voeten, 1999; Kahn-Horwitz et al., 2006; Lervåg & Aukrust, 2010; Melby-Lervåg, & Lervåg, 2011; Meschyan & Hernandez, 2002).

Two of Sparks et al.'s longitudinal studies have investigated the questions raised by Skehan about the relationships among early L1 ability, L2 aptitude, and L2 achievement. In one study, they measured students' L1 literacy and oral language skills from first to fifth grades, then followed them into high school where they administered the MLAT in ninth grade and L2 oral and written achievement measures in tenth grade after two years of L2 courses (Sparks et al., 2006). The results showed that L1 literacy and L1 vocabulary measures in elementary school predicted from 58% to 73% of the variance on the MLAT and from 30% to 43% of the variance in L2 achievement. In another study with these students, Sparks et al. (2009) found that the MLAT was the single best predictor of L2 reading comprehension, spelling, writing, and oral proficiency even in the presence of the L1 skills. The findings in the latter study revealed strong correlations between the L1 skill measures and the L2 achievement tests (.49–.68), as well as strong correlations between the MLAT and all L2 outcomes, that is, L2 writing (.50), L2 word decoding (.61), L2 spelling (.72), L2 oral proficiency (.54), and overall L2 proficiency (.75).

The results of these two studies prompted Sparks et al. to ask the same question as Skehan: Given the strong relationships between L1 skills and L2 achievement, why is the MLAT the most important predictor of L2 achievement? They proposed that L2 aptitude tests may preempt (cut out) the variance explained by L1 skills. Their simple explanation for the superiority of the MLAT is that L2 aptitude tests are comprised of basic language tasks that measure the skills necessary for language learning generally in both L1 and L2. But while the MLAT measures skills similar to L1 tests, it, like other aptitude tests, also includes tasks that can measure students' ability to learn from "decontextualized material" (Skehan, 1989, p. 34) that are related to language ability but not encountered in everyday life. For example, a student can speak or write a sentence without awareness of the grammatical function of each word, but the MLAT Words in Sentences requires knowledge of words' grammatical function (part of speech). Sparks et al. speculated that aptitude tests may draw their predictive value from tapping into students' metalinguistic skills, and concurred with Ranta (2002), who proposed that language analytic ability and metalinguistic ability are "two sides of the same coin" (p. 163).

In sum, studies with L2 learners have found that more and less successful L2 learners exhibit IDs in early L1 skills and in L2 aptitude; there are strong relationships among students' early L1 achievement, L2 aptitude, and later L2 achievement; and early L1 skills alone, especially L1 literacy, are strong predictors of L2 aptitude and later L2 achievement. Like Skehan, Sparks et al. have proposed that, unlike L1 measures, L2 aptitude tests assess the ability to learn from decontextualized material ability and to use language analytic, or metalinguistic, abilities.

Purpose of study and research questions

Underlying the present investigation is our view that both theory and research on variability in L2 achievement will benefit from a more comprehensive, quantitative

characterization of the relationships among L1 ability, L2 aptitude, and L2 achievement. Previous research has largely relied on individual bivariate correlations with a strong emphasis on determination of statistical significance, though regression has sometimes been used to establish a unique contribution of L2 aptitude. Furthermore, the use of simple correlations fails to acknowledge the role of limited reliability of measures in constraining those correlations. In addition, the hypothesis that much of the correlation from L2 aptitude to L2 achievement is due to the aptitude measure preempting (cutting out) L1 variance suggests two independent indices for the predictiveness of the aptitude measure.

The first index, which we label *uniqueness*, is the extent to which the L2 aptitude measure adds to prediction of L2 achievement beyond that predicted by L1 measures alone. Although regression analyses have been used for this purpose previously, we specifically propose that the relevant measure is not the absolute level of prediction by aptitude, but the proportion it represents of the total prediction. This adjustment acknowledges the role of limited reliability. The second index concerning predictiveness of the aptitude measure is the extent to which the aptitude measure captures the potential predictive potential of each L1 measure. We label this *efficiency*, which is measured as the degree of mediation by the aptitude measure of the total prediction from the L1 measures to the L2 measures. As stated earlier, the importance of these definitions is that they are in principle independent: *uniqueness* and *efficiency* could be both high or low, or either one could be high or the other low. Assessing the role of L2 aptitude in this way across diverse L2 measures will provide a more complete empirical basis for formulating hypotheses about the role of aptitude. The conceptual model underlying the calculation of these two indices is summarized in Figure 2.

In the present study, our measure of L2 aptitude was the MLAT. Having a single measure for an abstract construct entails that the conclusions can be relatively definitive with respect to that measure, while an inferential leap is required for the same conclusions about the construct more generally. However, given the prevalence of use of the MLAT in the field, similar networks of correlations of various aptitude measures with L1 and L2 measures, and the fact that the MLAT has norms for secondary
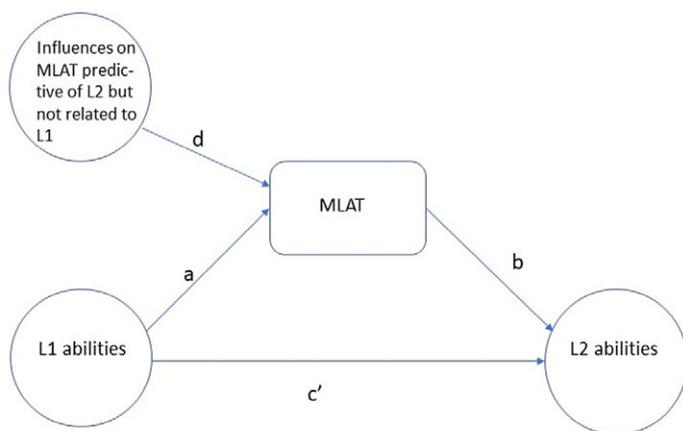
Uniqueness: d / (a+d) = unique contribution of MLAT to predicting L2 as a proportion of total prediction from MLAT to L2
Efficiency: a*b / (a*b + c′), the degree to which MLAT mediates the prediction from L1 to L2, as a proportion of the total prediction from L1 to L2

Figure 2. Conceptual mediation model of the role of L1 abilities and the MLAT for predicting L2 abilities.

level participants, led us to focus on it. The results can be taken as likely features of the construct more broadly. We return to this issue in the "Discussion" section.

In the present study, we included multiple L1 and L2 measures along with MLAT scores from 307 high school students engaged in Spanish instruction. We addressed four interlocking research questions about relations among early L1 achievement, MLAT performance, and subsequent L2 achievement: (a) How well do L1 scores predict MLAT scores and L2 achievement, (b) how well do MLAT scores predict L2 achievement, (c) to what extent does MLAT add unique variance to the prediction of each L2 measure beyond the prediction from L1 measures alone, and (d) how much of the variance in each L1 ability that is predictive of each L2 skill is captured by the MLAT as a mediator of the correlation between them? Research questions 3 and 4 capture the distinction described earlier between the "uniqueness" and "efficiency" of MLAT prediction.

Method

Participants

The study began with 307 participants randomly chosen from students enrolled in first-year Spanish courses at one of four high schools in a large suburban school district in the Midwest near a metropolitan US city. There were 154 males and 153 females whose mean age was 15 years, 7 months (ages ranged from 13 years, 7 months to 17 years, 6 months) enrolled in ninth, tenth, and eleventh grades at the beginning of the study. Participants included 301 Caucasian, 4 African American, and 2 East Asian students. A total of 293 (148 females and 145 males) of the 307 students completed the first-year Spanish course. All participants were monolingual English speakers who had no prior experience with Spanish, were not routinely exposed to Spanish outside school, and spoke no language other than English. Parental permission was obtained for each participant.

The sample size for this study represents the maximum size that could be assessed with project resources. It substantially exceeds the widely used informal guideline of ten subjects per predictor variable for regressions and provides 80% power for detecting significant correlations of $r = .16$ or greater.

Testing instruments

There were several types of testing measures used in this study: L1 achievement, L1 working memory, L2 aptitude, and L2 (Spanish) achievement. Each of the measures is briefly described. A complete description of each L1 measure is provided in Supplementary Appendix A. A complete description of the L2 aptitude and L2 achievement measures is provided in Supplementary Appendix B. Reliability coefficients for the instruments are reported in the appendices.

Measures of L1 skills

L1 Achievement

L1 word decoding

The two measures of word decoding were the Woodcock Reading Mastery Test-Revised Basic Skills Cluster (Woodcock, 1998) and the Test of Word Reading Efficiency (Torgesen et al., 1999). The L1 Word Decoding score was obtained by averaging a student's standard scores ($M = 100$, $SD = 15$) on the Woodcock Basic Skills Cluster and the TOWRE Composite.

L1 reading comprehension

The measure of L1 reading comprehension was the Stanford Achievement Test 10 (Pearson, 2007).

L1 vocabulary

The measure of L1 vocabulary was the Woodcock-Johnson-III/NU Picture Vocabulary subtest (Woodcock et al., 2001).

L1 language analysis

The measure of language analysis was the Test of Language Competence-Expanded Edition Figurative Language subtest (Wiig & Secord, 1989).

L1 writing

The measure of L1 writing was the On-Demand Writing assessment, a state-required outcomes assessment that is a timed, group-administered standardized measure of writing.

L1 working memory

The measure of phonological short-term memory was the Comprehensive Test of Phonological Processing, Phonological Memory Composite (CTOPP) (Wagner et al., 1999). The measure of working memory was the Woodcock-Johnson-III/NU Working Memory Cluster (Woodcock et al., 2001). The L1 Memory score was obtained by averaging a student's standard scores ($M = 100$, $SD = 15$) on the two tests. Support for combining these two measures was provided by the correlation of $r = .62$ between them in the present sample.

L2 Aptitude

The measure of L2 aptitude was the MLAT (Carroll & Sapon, 1959 [2000]). This standardized test measured L2 aptitude with a simulated format to provide an indication of the probable degree of success in learning a L2 (see Appendix B). The test does not provide normed subtest scores, only an overall aptitude score, obtained by summing subtest raw scores and referencing the test manual tables.

L2 (Spanish) Achievement

A standardized measure of Spanish achievement, the *Batería III Woodcock-Muñoz Pruebas de aprovechamiento* (Woodcock et al., 2004) designed for students whose native language is Spanish, was used to measure participants' Spanish achievement. This standardized measure of Spanish has been used in several studies to measure the Spanish achievement of US students (see Sparks et al., 2017; Sparks et al., 2019). The measure was chosen for use in the present study for several reasons. First, the Woodcock-Muñoz provides an explicit numerical value, that is, standard scores with M = 100, SD = 15, for all Spanish skills that identifies the level of Spanish achievement when US students are compared to native Spanish speakers. In a previous study, Sparks et al. (2017) reported that the first-year students' Spanish writing skills on the Woodcock-Muñoz reflected writing at the ACTFL novice-low to novice-mid level;

the second-year students' writing skills reflected novice-mid to novice-high level; and the third-year students' writing skills reflected novice-high to intermediate-low level. Second, the test was able to be administered in the time allotted by the school for testing the participants in this study. Third, the test provides separate scores for important components of reading (including word decoding, comprehension), writing (including spelling, writing sentences), and listening comprehension (including vocabulary, oral language comprehension). The subtests are listed here and described in Appendix B.

L2 reading

On the *Identificación de letras y palabras* subtest, a measure of Spanish word decoding, a student reads aloud a list of increasingly difficult Spanish words. On the *Comprensión de textos* subtest, a student reads a short passage and identifies a key missing word. The L2 Reading score was obtained by averaging a student's standard scores ($M = 100$, $SD = 15$) on the two subtests.

L2 writing

On the *Ortografía* subtest, a student spells (writes) increasingly difficult words presented orally. On the *Muestras de redacción* subtest, a student writes sentences in Spanish that were evaluated with respect to their quality. The L2 Writing score was obtained by averaging a student's standard scores ($M = 100$, $SD = 15$) on the two subtests.

L2 listening comprehension

On the *Vocabulario sobre dibujos* subtest, a student is asked to name common to less common objects shown in a picture. On the *Comprensión Oral* subtest, a student listens and comprehends a short, audio-recorded passage and supplies a missing word. The L2 Listening Comprehension score was obtained by averaging a student's standard scores ($M = 100$, $SD = 15$) on the two subtests.

L2 oral proficiency

At the end of the second-year only, students' oral proficiency in Spanish was assessed through a 10–15 minute individual interview. The interviews were conducted by two L2 (Spanish) educators, who had been trained to conduct oral interviews, and graduate students fluent in Spanish trained by them. The interviewers had no previous knowledge about the participants, who were assigned randomly to an interviewer.

Procedure

The testing instruments were administered to participants at different times over the course of the study. The MLAT was administered in groups of 25–30 students by the first author in the first 3–4 weeks of the first-year Spanish course. The L1 measures were administered individually by the first author, a Spanish professor, and graduate students trained by the first author at the beginning of the Spanish course. The participants' scores on the L1 reading comprehension and L1 writing measures were obtained from school records.

The measures of Spanish achievement were administered individually to the participants at the end of the first and second year courses by the first author, the L2 Spanish professor, and graduate students trained by them. Participants' raw scores for the six measures were transformed to standard scores ($M = 100$, $SD = 15$) using the Woodcock-Johnson-III Normative Update Compuscore and Profiles Program Version 3.1 (Schrank & Woodcock 2008). Because the Woodcock-Munoz is a standardized, norm-referenced test calibrated to measure the skills of native Spanish-speaking test-takers, norms were available for a wide range of grade levels, consequently, participants' scores on the six subtests could be compared to native Spanish-speaking students in first through twelfth grades. For this study, participants' scores according to ninth grade native Spanish speaker norms were used. The oral proficiency interviews were conducted at the end of the second-year Spanish course by the Spanish professor and graduate students trained by her.

Data imputation and analysis

There was a moderate proportion (7.5%) of missing data in the dataset, primarily due to students who did not take second-year Spanish. Little's Missing Completely at Random (MCAR) test did not reject the hypothesis of data missing completely at random ($\chi2 = 84.723$, $df = 69$, $p = .096$). Data imputation was conducted by expectation maximization, using the SPSS Missing Values program.

Data analysis began with the computation of descriptive statistics for all variables, and correlations among them. The correlations addressed the specific issues referenced in Research Questions 1 and 2. Research Question 3 was addressed by comparing the zero-order prediction from MLAT to each L2 measure with the increment to prediction by MLAT after all the L1 measures had been entered first in a multiple regression analysis. The proportional decrement in correlation was interpreted as a measure of the extent to which prediction by the MLAT was due to inclusion of L1 variance. Finally, we conducted simple mediation analyses of each L1 measure as a predictor of each L2 measure as an outcome. The analysis estimated the direct effect from the L1 measure and the indirect effect, which was mediated by MLAT. These analyses, which were conducted using the SPSS PROCESS macro version 4.0 for mediation, addressed Research Question 4.

Results

Descriptive statistics for all measures following imputation are provided in Table 1. Four of the L1 measures—vocabulary, working memory, word decoding, and reading comprehension—are standardized measures with $M = 100$, $SD = 15$, or averages of two or more such measures. Overall, the performance of these students is in the average range. Table S1 in the Supplementary File presents correlations among all analyzed variables in this study; subsets of relevant correlations for the research questions are presented in the following text.

RQ1: How well do L1 achievement scores predict MLAT and L2 achievement scores?

Table 2 includes the predictive correlations from L1 to L2 measures. All correlations are positive and significant. In general, the strongest L1 predictor is word decoding,

Table 1. Means, standard deviations, and ranges for L1 (English), L2 (Spanish), and MLAT measures

Measure	M	SD	Range
L1 Vocabulary	93.30	10.57	60–129
L1 Working Memory	97.18	10.59	66.5–127.0
L1 Language Analysis	9.25	1.68	1–12
L1 Word Decode	100.42	10.16	69.5–131.5
L1 Reading Comp	107.99	12.32	53–135
L1 Writing	10.83	1.97	5–16
L2 Reading Yr1	28.67	12.78	3.0–68.5
L2 Reading Yr 2	38.26	12.96	1.0–69.0
L2 Listening Comp Yr 1	19.50	7.11	4.0–52.0
L2 Listening Comp Yr2	19.91	7.98	4.0–40.0
L2 Writing Yr 1	35.88	9.84	19.5–79.0
L2 Writing Yr 2	46.23	12.16	19.5–89.5
L2 Oral Proficiency	8.29	2.09	1.0–16.0
MLAT SS	97.39	11.69	65–128

Note: $n = 307$ for all measures.

Table 2. L1 – L2 predictive correlations, and correlations with MLAT

L2 Measure/ MLAT L1 Measure/MLAT	Reading Yr 1	Reading Yr 2	Listening Comp Yr 1	Listening Comp Yr 2	Writing Yr 1	Writing Yr 2	Oral Proficiency	MLAT
Vocabulary	.277**	.303**	.364**	.402**	.338**	.392**	.178**	.302**
Working Memory	.291**	.324**	.174**	.223**	.261**	.322**	.151**	.365**
Lang Analysis	.235**	.248**	.201**	.278**	.226**	.270**	.172**	.325**
Word Decoding	.545**	.632**	.310**	.441**	.414**	.521**	.343**	.443**
Reading Comp	.245**	.326**	.185**	.348**	.197**	.282**	.325**	.280**
Writing	.200**	.367**	.210**	.351**	.260**	.285**	.324**	.223**
MLAT	.389**	.457**	.422**	.452**	.425**	.479**	.326**	

Note: $n = 307$ for all measures.
**$p < .01$.

followed by vocabulary. There is also a consistent pattern of L2 Year 2 measures being more strongly predicted than the parallel L2 Year 1 measure for all L1 predictors.

As also shown in Table 2, all L1 measures significantly predict MLAT performance, with correlations of .223–.443. Word decoding again emerged as the strongest predictor. Table 3 summarizes a multiple regression analysis utilizing all six L1 measures to predict MLAT. The multiple $R = .524$ reflects significant influence from all L1 measures, except reading comprehension and writing.

Table 3. Multiple regression prediction from L1 measures to MLAT

Variable	B	SE	β	t	p
(Constant)	21.486	7.653		2.807	.005
L1 Vocabulary	.121	.060	.109	2.000	.046
L1 Working Memory	.168	.063	.152	2.661	.008
L1 Lang Analysis	.934	.391	.134	2.388	.018
L1 Word Decoding	.291	.071	.253	4.108	< .001
L1 Read Comp	.073	.053	.077	1.389	.166
L1 Writing	.233	.322	.039	.722	.471

Note: Total prediction is $R = .524$; adjusted $R^2 = .26$.

RQ2: How well do MLAT scores predict L2 achievement?

Table 2 also includes the correlations from MLAT to each of the L2 skills, which addresses Research Question 2. All correlations are positive and significant. The prediction to L2 Year 2 scores are consistently, though only modestly, higher than the predictions to L2 Year 1 scores. There is little variation of these predictions across the L2 measures.

RQ3: To what extent does MLAT uniquely add to the prediction from L1 to L2?

To address Research Question 3, we used multiple regression analyses to estimate the proportion of MLAT to L2 predictions that are due to inclusion of L1 variance in the MLAT. A separate regression was conducted for each of the seven L2 measures. Table 4 summarizes the most important results from the seven regression analyses. The final column of the table estimates the proportion of the prediction from MLAT to L2 achievement that is unique, in other words, not predicted by the L1 measures. The regression tables for the remaining regression predictions are included in the Supplementary File Table S2. As shown in Table 4, the majority of the prediction (from 59.6–87.4% of the variance) is due to L1, with only a small, though significant proportion (except for L2 Writing Yr 2), due to a unique contribution from MLAT.

RQ4: How much of the variance in L1 abilities that is predictive of L2 skills is captured by the mediation by MLAT of L1-L2 correlations?

The preceding analysis evaluates the extent to which L1 scores account for the prediction from MLAT. A complementary analysis addresses Research Question 4, which can be rephrased as how efficiently, or thoroughly, the MLAT extracts (or estimates) information about L1 abilities in serving as a mediator. The question can be addressed for each combination of L1 and L2 measures, for a total of 36 analyses. The analyses were conducted using the PROCESS Macro v4.0 (Hayes, 2021) within SPSS. The relevant output from this analysis, which is relevant for Research Question

Table 4. Proportion of MLAT–L2 predictions due to inclusion of L1 variance in MLAT

L2 measure	r	Prediction from MLAT to L2 % variance	% variance for "direct" MLAT prediction *after* all L1 measures entered in multiple regression	Proportion (%) of MLAT prediction effect due to inclusion of L1 measure; (col 3–col 4)/col 3	Proportion (%) of effect size for MLAT which is new, not predicted by L1 measure, col 1–5
Reading Yr 1	.389	15.1	1.9	87.4	12.6
Reading Yr 2	.457	20.9	2.8	86.7	13.3
Listening Comp Yr 1	.422	17.8	7.2	59.6	40.4
Listening Comp Yr 2	.452	20.4	4.7	77.0	23.0
Writing Yr 1	.425	18.1	4.9	72.9	27.1
Writing Yr 2	.479	22.9	4.8	79.0	21.0
Spanish Oral Proficiency	.326	10.6	2.6	75.5	24.5

Table 5. Mediation analysis for MLAT mediation of predictions from L1 to L2 measures

L1 measure	L2 measure-Year	Total effect	Indirect (mediated by MLAT) effect (% of total effect)	Direct effect from L1 measure
Vocabulary	Reading-1	.335	.123 (36.7%)	.212
	Listening Comp-1	.245	.070 (28.6%)	.175
	Writing-1	.315	.100 (31.7%)	.215
	Reading-2	.372	.149 (40.1%)	.223
	Listening Comp-2	.303	.083 (27.4%)	.220
	Writing-2	.451	.138 (30.6%)	.313
	Spanish Oral Prof.	.035	.018 (51.4%)	.017
Word Decoding	Reading-1	.686	.102 (14.9%)	.584
	Listening Comp-1	.217	.110 (50.7%)	.107
	Writing-1	.401	.129 (32.2%)	.272
	Reading-2	.807	.124 (15.4%)	.682
	Listening Comp-2	.347	.111 (32.9%)	.236
	Writing-2	.624	.164 (26.3%)	.460
	Spanish Oral Prof.	.070	.020 (28.6%)	.051
Reading Comprehension	Reading-1	.255	.154 (60.4%)	.101
	Listening Comp-1	.107	.065 (60.7%)	.042
	Writing-1	.158	.090 (57.0%)	.068
	Reading-2	.343	.117 (34.1%)	.226
	Listening Comp-2	.225	.070 (31.1%)	.156
	Writing-2	.279	.120 (43.0%)	.159
	Spanish Oral Prof.	.055	.012 (21.8%)	.043
Writing	Reading-1	1.299	.548 (42.2%)	.751
	Listening Comp-1	.760	.332 (43.7%)	.429
	Writing-1	1.302	.449 (34.5%)	.853
	Reading-2	2.421	.602 (24.9%)	1.819
	Listening Comp-2	1.427	.369 (25.9%)	1.057
	Writing-2	1.765	.627 (35.5%)	1.138
	Spanish Oral Prof.	.344	.066 (19.2%)	.279
Language Analysis	Reading-1	1.790	.866 (48.4%)	.924
	Listening Comp-1	.852	.550 (64.6%)	.302
	Writing-1	1.326	.751 (56.6%)	.575
	Reading-2	1.915	1.057 (55.2%)	.858
	Listening Comp-2	1.324	.625 (47.2%)	.699
	Writing-2	1.963	1.030 (52.5%)	.934
	Spanish Oral Prof.	.215	.122 (56.7%)	.092
Memory	Reading-1	.351	.144 (41.0%)	.208
	Listening Comp-1	.117	.101 (86.3%)	.016
	Writing-1	.243	.129 (53.0%)	.114
	Reading-2	.397	.174 (43.8%)	.222
	Listening Comp-2	.169	.118 (69.8%)	.051
	Writing-2	.370	.175 (47.3%)	.195
	Spanish Oral Prof.	.030	.023 (76.7%)	.007

4 is the proportion of the total prediction effect from the L1 measure that is mediated by the MLAT (the indirect effect). As shown in Table 5 and illustrated in Figure 3, the MLAT is only moderately efficient, with an average indirect effect of about half. Higher MLAT efficiency is found for L1 reading comprehension, language analysis, and working memory, although these measures are the weakest predictors of L2 scores generally.
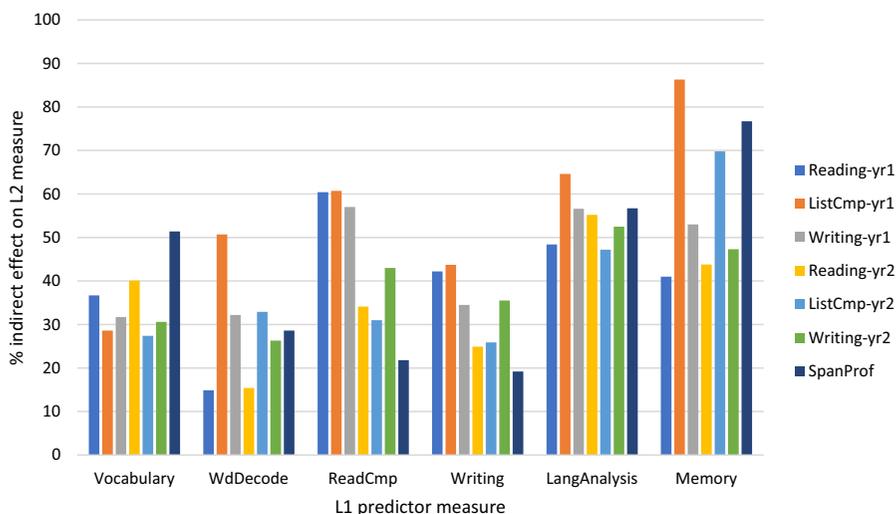
Figure 3. Degree of mediation of L1 prediction of L2 measures by MLAT.

Discussion

We proposed that theory and research investigating variability in L2 achievement will benefit from a more comprehensive, quantitative characterization of the relationships among L1 ability, L2 aptitude, and L2 achievement, and suggested that research examining these relationships has been constrained by relying on individual bivariate correlations with an emphasis on determining statistical significance and using simple correlations with their inherent limitations. Researchers have hypothesized that much of the correlation between L2 aptitude and L2 achievement may be due to the aptitude measure(s) preempting (cutting out) prediction from L1 variance. However, we speculated that there may be two independent indices, uniqueness and efficiency, involved for the prediction of L2 aptitude measures for L2 achievement. To explore our hypothesis, we asked four research questions, each of which are discussed in this section.

Our first research question asked how well L1 scores predict MLAT scores and L2 achievement. For L2 achievement, the results in Table 2 show that the predictive correlations from the L1 achievement to L2 achievement measures are all positive and significant. The strongest L1 predictor for L2 achievement was L1 word decoding; in particular, L1 word decoding was a strong predictor of both first-year (.545) and second-year (.632) L2 reading. Following L1 decoding, the L1 vocabulary measure was also a strong predictor of L2 achievement, particularly for first-year (.364) and second-year (.402) L2 listening comprehension. These findings are similar to those from a 10-year longitudinal study conducted by Sparks et al. (2006, 2009) who found that measures of L1 achievement, especially those related to L1 literacy (word decoding, spelling, reading comprehension, reading readiness) from first through fifth grades, predicted from 30% to 40% of the variance in oral and written L2 achievement in high school. In that study, L1 word decoding was a strong predictor of L2 word decoding (.66) and L2 reading comprehension (.44); L1 reading comprehension was a strong predictor of L2 reading comprehension (.50); and L1 vocabulary was a strong predictor of L1 listening comprehension (.46). Several years earlier, Skehan (1986) reported that

early L1 development in preschool, specifically vocabulary and language comprehension, was strongly correlated with L2 achievement several years later in high school. A new finding in the current study showed that L1 word decoding is also the strongest predictor of L2 oral proficiency (.343), followed by L1 reading comprehension (.325) and L1 writing (.324). Likewise, in Sparks et al.'s longitudinal study, measures of L1 literacy in elementary school—word decoding (.40), spelling (.45), and reading comprehension (.51) —were strongly predictive of L2 oral proficiency in high school. Taken together, these findings suggest that L1 literacy skills developed prior to L2 exposure are important for both written and oral L2 achievement.

Table 2 also shows that all L1 measures significantly predicted MLAT performance with correlations ranging from .223–.443. In this study, L1 word decoding, a measure of L1 literacy, was the strongest predictor of MLAT scores. The finding that a measure of L1 literacy was also the strongest predictor of L2 aptitude on the MLAT is similar to those in Sparks et al.'s (2006) longitudinal study, which found that measures of L1 literacy and L1 vocabulary from first through fifth grades predicted from 58% to 73% of the variance on the MLAT administered in high school. In other studies with high school L2 learners, L1 achievement measures have also been found to be strongly correlated with participants' MLAT scores (e.g., Sparks et al., 1998; Sparks et al., 1997; Sparks et al., 2008). In their study with children from the Bristol Language Project, Skehan and Ducroquet (1988) reported that measures of oral language in preschool prior to age 5 were significantly correlated with L2 aptitude (on the York Language Aptitude Test, Elementary MLAT, Pimsleur LAB) several years later in high school. The aforementioned findings suggest that the skills involved in early oral language development as well as the skills necessary for L1 literacy competence are related to those measured by an L2 aptitude test, in this case, the MLAT.

Table 3 summarizes the results of a multiple regression analysis that used the six L1 measures together to predict participants' MLAT scores. The results indicated that four of the measures—L1 vocabulary, L1 working memory, L1 language analysis, and L1 word decoding—contributed significantly to the prediction of MLAT scores ($R = .524$). However, two measures were not significant—L1 reading comprehension and L1 writing. These findings suggest that skills necessary for the development of L1 literacy (word decoding) and L1 oral language (vocabulary, language analysis), as well as a skill found to be important for L1 reading comprehension (L1 working memory), have a significant relationship for prediction of L2 aptitude on the MLAT. These findings are similar, in part, to Sparks et al.'s (2006) study cited earlier in which measures of L1 literacy and a measure of L1 vocabulary in first through fifth grades accounted for large amounts of variance on the MLAT several years later. For the present study, findings suggest that even though L1 reading comprehension and L1 writing were significantly correlated with the MLAT, much of their variance in the regression was captured by the other L1 measures.

Our second research question asked how well MLAT scores predict L2 achievement. Table 2 shows that not only are all correlations between MLAT and L2 achievement positive and significant but also that the predictions from MLAT, like those from the L1 measures to second-year L2 achievement, were modestly, but consistently, higher than the predictions to first-year L2 achievement. The first finding is in accord with research over many years which has found that the MLAT is a reliable predictor of overall L2 achievement (e.g., see Ehrman & Oxford, 1995; Skehan, 1998; Stansfield & Reed, 2019). In their longitudinal study over 10 years, Sparks et al. (2009) found that the MLAT was the single best predictor of overall L2 proficiency, accounting for 56% of the variance, and was also the best predictor for most L2 skills, that is, L2 spelling (52%), L2 reading

comprehension (39%), L2 writing (34%), and L2 oral proficiency (29%). (However, L1 word decoding was the single best predictor of L2 word decoding [see Sparks et al., 2008].) In his review, Li (2019) reported correlations of .30–.39 between several aptitude batteries and specific L2 skills (reading, writing, listening, speaking). The results of the present study are generally consistent with but somewhat stronger than those reported by Li (2019), for example, up to .457 for L2 reading, .452 for L2 listening comprehension, and .479 for L2 writing. However, in their 10-year study, Sparks et al. (2006) found stronger correlations between MLAT scores and students' performance in L2 word decoding (.61), L2 reading comprehension (.62), L2 writing (.58), and L2 listening/speaking (.54). The second finding, a consistent pattern of all second-year L2 measures being more strongly predicted by the L1 measures than were the parallel first-year L2 measures, suggests that Year 2 performance in the L2 is a better measure of a student's L2 achievement. This may be so because students in a second year of L2 are necessarily working with a stronger and broader understanding and grasp of the language structure of the target language. It may also be the case that students can more easily find alternative ways to cope with the difficulties they encounter in the early stages of language learning. The findings from the current study and others confirm that the MLAT is a consistently strong and reliable predictor of L2 achievement.

Our third research question asked the extent to which the MLAT adds unique variance to the prediction of each L2 measure beyond the prediction from L1 measures alone. The results indicated that this proportion (uniqueness) is relatively modest for all L2 measures (see Table 4, last column). The unique contribution of the MLAT for the prediction of L2 measures expressed as a proportion of the total prediction was ≤ 40.4% for all seven measures, and ≤ 23.0 % for the majority of them. These findings are quite similar to a previous investigation in which hierarchical regression analyses showed that IDs in L1 achievement alone (reading, writing, vocabulary, print exposure) accounted for substantial unique variance (20%–50%) in L2 reading, writing, listening comprehension, and oral proficiency, while the MLAT accounted for a small absolute amount of unique variance (2%–6%, compared to 1.9%–7.2% in the present study, as shown in Table 4, column 3) at the end of first- and second-year L2 achievement. One reason why the MLAT explains some unique proportion of variance for L2 achievement is that the L1 measures in this study did not directly assess some important abilities, including grammar measured directly by the Words in Sentences subtest, which are necessary for L2 achievement. Even so, the findings raise the question of why the unique variance accounted for the MLAT is so modest while the test has been found over many years to be the best single predictor of L2 achievement. The strongest answer to this question is that there is considerable overlap between the skills measured by the MLAT and the L1 measures, reducing the opportunity for MLAT to add unique variance. This may be especially true for phonetic coding ability, which is measured by MLAT Phonetic Script subtest and the L1 word decoding measures, and vocabulary knowledge, which is measured by MLAT Spelling Clues subtest and the L1 vocabulary measure. In addition, rote learning ability, measured by the MLAT Paired Associates and Number Learning subtests, is also required for several of the skills assessed by the L1 measures.

Another possible but not mutually exclusive explanation for the aforementioned findings is that the MLAT and the L1 measures are to some extent assessing different abilities. Roehr-Brackin (2018) has noted that despite similarities in the types of tasks measured by L1 and L2, "measures of metalinguistic awareness are typically based on L2, while measures of language-analytic ability as a component of language learning aptitude are typically based on L1" (p. 127). She defines language-analytic ability as the ability to "treat language as an object of analysis and arrive at linguistic generalizations,"

which is "at the core of the constructs of language learning aptitude and metalinguistic awareness, which are implicated in our ability to learn explicitly" (Roehr-Brackin & Tellier, 2019, p. 1111). Her claim leads to speculation that because the participants had learned and used their L1 oral (listening, speaking) and written (literacy) language skills for many years prior to engaging in L2 courses, they likely possessed sufficient linguistic ability to use and analyze English. However, the MLAT challenged their metalinguistic ability, that is, "the capacity to use knowledge about language as opposed to the capacity to use language" (Bialystok, 2001, p. 124). For example, the MLAT requires students to use knowledge *about* language to, that is, to learn a new sound-symbol system (Phonetic Coding subtest), to use their knowledge *about* English to decode incorrectly spelled words (Spelling Clues), and to use their knowledge *about* English grammar to perform a task of grammatical structure (Words in Sentences). However, explicit knowledge about English was not needed by the participants to read, spell, and write words and sentences. In some studies, metalinguistic awareness and language learning aptitude have been found to be partially overlapping constructs (e.g., see Jessner, 2006), but others have found that metalinguistic knowledge and language aptitude are distinguishable constructs (Roehr-Brackin, 2018). In either case, some factor analytic investigations have found that L1 tests and MLAT subtests that measure similar skills, for example, phonetic coding, load on the same factor (Sparks et al., 2011), while other factor analyses have shown that all five MLAT subtests load on the same factor (Sparks et al., 2019), depending on the L1 measures used for the study. These findings lend credence to the claim that L1 tests and MLAT subtests may be measuring similar skills, for example, phonology, grammar, rote memory, but assessing different specific abilities in these domains. Consequently, the MLAT may provide additional prediction for L2 achievement.

Our fourth research question asked how much of the variance in L1 abilities that is predictive of L2 skills is preempted (captured) by the mediation by MLAT of L1-L2 correlations. This question can be rephrased to ask how efficiently the MLAT extracts (or estimates) information about L1 abilities while serving as a mediator. The relevant output from this analysis is the proportion of the total prediction effect from a specific L1 measure to a specific L2 measure that is mediated by the MLAT (indirect effect). Table 5 and Figure 3 show that in total terms, the MLAT is only moderately efficient with an average indirect effect of about 50%. The MLAT became less efficient for extracting information for L1 reading comprehension over time from Year 1 to Year 2 for all three L2 variables, that is, from 60% to 34% for L2 reading, 61% to 36% for L2 listening comprehension, from 57% to 43% for L2 writing. Likewise, the MLAT became less efficient for extracting information for L1 reading comprehension, L1 working memory, and L1 language analysis from Year 1 to Year 2 for L2 listening comprehension. One explanation for these results could be based on findings from Sparks et al., who found that the MLAT scores of both low-achieving and average- to high-achieving groups of secondary level L2 learners increased one standard deviation after one year of Spanish instruction, and that the gains were maintained over a second year of Spanish (Sparks & Ganschow, 1993b; Sparks et al., 1992, 1998). The findings from those studies and the present investigation suggest that the longer students engage in the study of a L2, the better the development of their metalinguistic ability to use knowledge about language and, over time, the better they can use and understand "decontextualized material," that is, knowledge about language as measured by the MLAT, which may increase the efficiency of IDs in MLAT for predicting achievement in some L2 skills.

In contrast to the aforementioned L1 skills where the MLAT was relatively more efficient, the MLAT was much less efficient at extracting variance about L1 word

decoding, vocabulary, and writing. In particular, the MLAT was largely inefficient at extracting variance about L1 word decoding for L2 reading (15%) and L2 writing (26%–32%) in Years 1 and 2; L1 vocabulary for L2 listening comprehension (27%–28%) and L2 writing (30%–31%) in Years 1 and 2; and L1 writing for L2 writing (35%–36%) in Years 1 and 2 and for L2 listening comprehension in Year 2 (26%). As noted earlier, the MLAT is highly dependent on students' L1 literacy ability. In addition, there is likely to be some degree of overlap between the L1 vocabulary measure and the MLAT Spelling Clues subtest, which assesses L1 vocabulary (after words are decoded using decontextualized material). A study described earlier in the review found that L1 skills, especially L1 literacy and L1 vocabulary, predicted from 58% to 73% of the variance on the MLAT and from 30% to 43% of the variance in L2 achievement (Sparks et al., 2006). Taken together, the findings from the present investigation and others show that although there is considerable overlap between L1 skills and L2 aptitude and L1 skills and L2 achievement, much of the predictive power of L1 skills for L2 achievement is "missed" by the MLAT.

Toward a theory of L1 abilities and L2 aptitude

Our findings raise two distinct but related questions: Why is there substantial overlap between L1 ability and L2 aptitude as measured by the MLAT, and why is much of the predictive power of L1 ability for L2 achievement "missed" by the MLAT? One easy answer is that the L1 measures and the MLAT both measure *language* ability but do so differently, that is, assessing contextualized versus decontextualized material. Another straightforward answer is that the MLAT is heavily dependent on literacy ability. But, these answers do not explain the findings by Skehan that preschool L1 oral language abilities, particularly vocabulary and language comprehension, predict L2 aptitude years later in high school, or the research of Sparks et al., that has found strong relationships between L1 ability, particularly literacy in elementary school, and L2 aptitude in high school, most strikingly the finding that early L1 literacy and L1 vocabulary skills accounted for up to 73% of the variance in L2 aptitude several years later. This developmental evidence suggests that there must be a more ambitious explanation for the overlap between L1 ability and L2 aptitude.

We propose a different, two-pronged explanation for the substantial overlap between L1 ability and L2 aptitude along with the amount of predictive power of L1 ability for L2 achievement "missed" by the MLAT. Our explanation draws on the predictive results just mentioned, along with other new evidence for relationships among L1 literacy, metalinguistic awareness, and L2 aptitude. For some time, researchers have suggested that metalinguistic awareness and language aptitude can be considered partially overlapping constructs (Herdina & Jessner, 2002; Jessner, 2006); consequently, tests of language aptitude and metalinguistic awareness may assess overlapping skills (Ellis, 2004). Ranta (2002) has gone as far as suggesting that language aptitude and metalinguistic awareness are "two sides of the same coin." Integrating this insight with previous research, we suggest that the MLAT is both (a) preempting ("cutting out") variance explained by L1 ability *and* (b) extracting the "meta" parts (decontextualized material) of L2 aptitude.

We further propose that a central role for L1-L2 connections is the development of L1 literacy. Roehr-Brackin (2018) has reviewed research underscoring the link between the onset of literacy and the development of metalinguistic awareness (see also, Bialystok, 2001; Yelland et al., 1993). In L1 research, it is well-known that learning to

read (literacy) is "parasitic" on speech (Kavanagh & Mattingly, 1972) and language development (Snowling & Hulme, 2012). Prior to literacy development, some meta-linguistic awareness can be drawn from oral L1, for example, rhyming and alliteration (Snow et al., 1998). However, it is the development of literacy that leads to enhanced metalinguistic awareness (e.g., Kurvers et al., 2006; Tunmer et al., 1988), which further enhances literacy skills. Within L2, Koda (2005) has found that metalinguistic aware-ness and literacy are "developmentally interdependent" (p. 312) We posit that the "path" from L1 ability to L2 achievement begins with oral L1 ability followed by the development of L1 literacy, which leads to the development of metalinguistic aware-ness. L1 literacy and metalinguistic awareness together are the foundation of L2 aptitude, which predicts L2 achievement.

In sum, the results of our study suggest a much more important role for L1 literacy in explaining L2 aptitude and L2 achievement than has previously been acknowledged. Analogous to the connection between metalinguistic awareness and language aptitude, L1 literacy and metalinguistic awareness may be overlapping constructs that provide the foundation for L2 aptitude. Our proposal notably accounts for developmental findings in the L2 literature that have found predictions from early L1 literacy and language abilities to later L2 outcomes, along with strong prediction from L1 literacy to L2 aptitude, and that L1 print exposure explains unique variance in L2 achievement.

Limitations and implications

The strengths of this study are a large and representative sample, comprehensive L1 measures, and a prospective design. At the same time, there are some limitations that may limit generalizability of the conclusions. The first and most general of these was introduced earlier: the use of the MLAT as the single aptitude measure, so that results concerning the specific measure and the conclusions about aptitude may not be clearly distinguished. However, on grounds of face validity and actual similarity of items across aptitude tests, it is likely that tests of aptitude will have a strong "family resemblance" along with individual differences. Ultimately these issues can only be decided by research comparing aptitude measures in a unified, or at least comparable design.

A second limitation is that the present study is focused on pedagogically conven-tional classroom-based L2 learning, rather than immersion or immigration-based learning. The instruction is occurring after the development of literacy, as opposed to early simultaneous bilingualism. A related issue is that the study includes a single context of learning in the United States with only two years of L2 learning, which is typical for US L2 learners. Likewise, the L2 for all students is Spanish, for which there is a typologically close relationship with English. Furthermore, Spanish may be present in the oral and print environment to some extent, and reasonably valued, or at least not strongly disfavored in the larger society; these are features that may have a strong impact on motivation. The degree of orthographic similarity/difference may also be relevant, given the importance of phonological awareness and word decoding as predictors of outcome.

A third limitation is that some of the subtests on the Spanish measure used in this study called for single word responses to the test items. Although it is not yet known whether the format of test items interacts with teaching curriculum, replication of the present study using different kinds of measures that allow for broad-based responses of students' oral and written proficiency in Spanish should be conducted. Each of these limitations constitutes a recommendation for further, and more diverse research,

especially with other languages that have more typological distance. Despite these limitations, there are several implications that can be drawn from the study.

First, although the results of this study are critical of the conventional interpretation of the MLAT, they do not constitute a criticism of the MLAT. Clearly, it is more time-efficient to administer the MLAT (or another L2 aptitude measure) than a full battery of L1 measures. Nevertheless, the modest degree of mediation by the MLAT of L1 measures suggests that for prediction purposes, it would be cost-effective to include a measure of L1 word decoding, as the most predictive of L1 measures. The WRMT Basic Skills Cluster (less than 5 minutes) and the TOWRE (2 minutes) are brief but highly valid measures of L1 word decoding. As shown in Table S2, adding either of these measures to the MLAT approximately doubles the prediction to L2 reading and also makes a substantial contribution to the prediction of Listening Comprehension (especially Year 2) and L2 Writing.

Second, the conclusion that the prediction from MLAT to L2 achievement is largely due to measurement by the MLAT of L1 skills is an empirical finding, not an explanation of that prediction or a complete listing of the components of L2 aptitude. Language and literacy are central in a model of language aptitude because language skills are necessary for L2 learning. (Skehan, 2019; Sparks et al., 2019). However, there may be other distinctive cognitive and socioemotional skills important for L2 learning. For example, Wen et al. (2015) have explored the role played by working memory in L2 processing, interaction and performance, and instruction. Others have proposed models of aptitude that include domain-specific and domain-general variables and explicit and implicit processes for language learning (e.g., see Wen et al., 2017). Over time, these types of investigations will help to clarify the roles of variables important for L2 learning.

Third, L2 educators should be aware that L1 literacy plays an important role for L2 learning. Students with stronger reading ability in their L1 are more likely to have stronger metalinguistic ability than their peers with lower levels of literacy, and also to have read more extensively, which also improves other literacy-related skills, for example, vocabulary, grammar, and declarative knowledge. Although L2 teachers are not responsible for teaching L1 literacy skills, they should be cognizant of their students with more and less language and literacy ability and especially the roles that literacy and metalinguistic awareness play for L2 achievement. Roehr-Brackin and Tellier (2019) suggest that form-focused instruction in L2 may help to improve students' metalinguistic awareness (language analytic ability). To the extent that there is a bidirectional relationship between L2 aptitude and L1, form-focused instruction may also improve metalinguistic awareness in L1 (see Sparks et al., 1998).

Lastly, our study has shown that prediction from MLAT is primarily due to its functioning as a measure of L1 abilities, but a large proportion, often the majority, of L1 variance which predicts L2 scores is not captured by the MLAT. These findings require replication in longitudinal studies with other groups of L2 learners who have been administered a battery of L1 and L2 achievement measures in conjunction with the MLAT or other L2 aptitude measures. Investigators should also be aware of the need to identify additional specific abilities not captured by the MLAT to improve aptitude measures.

Competing Interests. The authors declare none.

References

Bates, E. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. University of Cambridge.

Bates, E., Dale, P., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 96–151). Blackwell.

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press.

Bleses, D., Makransky, G., Dale, P., Højen, A., & Ari, B. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, *37*, 1461–1476. https://doi.org/10.1017/S0142716416000060

Bloom, L., & Lahey, M. (1978). *Language development and language disorders*. Wiley.

Bornstein, M., & Putnick, D. (2012). Stability of language in childhood: A multiage, multidomain, multimeasure, and multisource study. *Developmental Psychology*, *48*, 477–491. https://doi.org/10.1037/a0025889

Brown, R. (1973). *A first language: The early stages*. Harvard University Press.

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

Carroll, J. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. Parry & C. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 1–29). Prentice-Hall.

Carroll, J., & Sapon, S. (1959 [2000]). *Modern Language Aptitude Test (MLAT): Manual*. Psychological Corp. Republished by Second Language Testing, Inc., http://www.2LTI.com

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*, 222–251.

Dąbrowska, E. (2016). Cognitive linguistics' seven deadly sins. *Cognitive Linguistics*, *27*, 479–491. https://doi.org/10.1515/cog-2016-0059

Dąbrowska, E. (2019). Experience, aptitude, and individual differences in linguistic attainment: A comparison of native and nonnative speakers. *Language Learning*, *69*, 72–100. https://doi.org/10.1111/lang.12323

Dorcus, R., Mount, G., & Jones, M. (1953). *Construction and validation of foreign language aptitude tests*. University of California, Los Angeles.

Dufva, M., & Voeten, M. (1999). Native language literacy and phonological memory as prerequisites for learning English as a foreign language. *Applied Psycholinguistics*, *20*, 329–348. https://doi.org/10.1017/S014271649900301X

Ehrman, M. (1998). The Modern Language Aptitude Test for predicting learning success and advising students. *Applied Language Learning*, *9*, 31–70.

Ehrman, M., & Oxford, R. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal*, *76*, 67–89. https://doi.org/10.1111/j.1540-4781.1995.tb05417.x

Ellis, R. (2004). Individual differences in second language learning. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 525–551). Blackwell.

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories. Users Guide and Technical Manual* (2nd ed.). Brookes.

Gardner, R., & Lambert, W. (1965). Language aptitude, intelligence, and second-language achievement. *Journal of Educational Psychology*, *56*, 191–199. https://doi.org/10.1037/h0022400

Gilkerson, J., Richards, J., Warren, S., Montgomery, J., Greenwood, C., Kimbrough, C., Oller, D., & Paul, T. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology*, *26*, 248–265. https://doi.org/10.1044/2016_AJSLP-15-0169

Hayiou-Thomas, M., Harlaar, N., Dale, P., & Plomin, R. (2010). Preschool speech, language skills, and reading at 7, 9, and 10 years: Etiology of the relationship. *Journal of Speech, Language, and Hearing Research*, *53*, 311–332. https://doi.org/10.1044/1092-4388(2009/07-0145)

Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism: Perspectives of change in psycholinguistics* (Vol. 121). Multilingual Matters.

Hunt, T., Wallace, F., Doran, S., Buynitzky, K., & Scharz, R. (1929). *Language Aptitude Test*. George Washington University.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*, 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Jessner, U. (2006). *Linguistic awareness in multilinguals: English as a third language*. Edinburgh University Press.

Kahn-Horwitz, J., Shimron, J., & Sparks, R. (2006). Weak and strong novice readers of English as a foreign language: Effects of first language and socioeconomic status. *Annals of Dyslexia*, *56*, 161–185. https://doi.org/10.1007/s11881-006-0007-1

Kavanagh, J., & Mattingly, I. (1972). *Language by ear and by eye: The relationships between speech and reading*. MIT Press.

Kendeou, P., van den Broek, P., White, M., & Lynch, J. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, *101*, 765–778. https://doi.org/10.1037/a0015956

Kidd, E., & Donnelly, S. (2020). Individual differences in first language acquisition. *Annual Review of Linguistics*, *6*, 319–340. https://doi.org/10.1146/annurev-linguistics-011619-030326

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge University Press.

Kurvers, J., van Hout, R., & Vallen, T. (2006). Discovering language: Metalinguistic awareness of adult illiterates. In I. van de Craats, J. Kurvers, & M. Young-Scholten (Eds.), *Low-educated adult second language and literacy acquisition* (pp. 69–88). LOT.

Lervåg, A., & Aukrust, V. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension between first and second language learners. *Journal of Child Psychology and Psychiatry*, *51*, 612–620. https://doi.org/10.1111/j.1469-7610.2009.02185.x

Li, S. (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, *36*, 385–408. https://doi.org/10.1093/applin/amu054

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

Li, S. (2017). The effects of cognitive aptitudes on the process and product of L2 interaction: A synthetic review. 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. (2019). Six decades of language aptitude research: A comprehensive and critical review. In Z. Wen et al., *Language aptitude: Advancing theory, testing, research, and practice* (pp. 78–96). Routledge.

Linck, J., Hughes, M., Campbell, S., Silbert, N., Tare, M., Jackson, S., …. & Doughty, C. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*, 530–566. https://doi.org/10.1111/lang.12011

Luria, M., & Orleans, J. (1928). *Luria-Orleans Modern Language Prognosis Test*. World Book.

Meara, P. (2005). *Llama Language Aptitude Tests*. University of Wales.

Melby-Lervåg, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading*, *34*, 114–135. https://doi.org/10.1111/j.1467-9817.2010.01477.x

Meschyan, G., & Hernandez, A. (2002). Is native-language decoding skill related to second-language learning? *Journal of Educational Psychology*, *94*, 14–24. https://doi.org/10.1037/0022-0663.94.1.14

Pearson. (2007). *Stanford Achievement Test 10*. Pearson.

Pimsleur, P. (1966). The *Pimsleur Language Aptitude Battery*. Harcourt Brace Jovanovich (republished by Second Language Testing, Inc., www.2lti.com).

Ranta, L. (2002). The role of the learner's language analytic ability in the communicative classroom. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 159–180). John Benjamins.

Roehr-Brackin, K. (2018). *Metalinguistic awareness and second language acquisition*. Routledge.

Roehr-Brackin, K., & Tellier, A. (2019). The role of language-analytic ability in children's instructed second language learning. *Studies in Second Language Acquisition*, *41*, 1111–1131. https://doi.org/10.1017/S0272263119000214

Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence: Quantitative and qualitative analyses*. Peter Lang.

Schrank, F., & Woodcock, R. (2008).Woodcock Interpretation and Instructional Interventions Program (WIIIP, Version 1.0) [Computer software]. Riverside.

Skehan, P. (1986). The role of foreign language aptitude in a model of school learning. *Language Testing*, 3, 188–221. https://doi.org/10.1177/026553228600300207

Skehan, P. (1989). *Individual differences in second language learning*. Arnold.

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

Skehan, P. (2019). Language aptitude implicates language and cognitive skills. In E. Wen et al., *Language aptitude: Advancing theory, testing, research, and practice* (pp. 56–77). Routledge.

Skehan, P., & Ducroquet, L. (1988). *A comparison of first and foreign language ability*. ESOL Department, Institute of Education, London University: Working Documents No. 8.

Snow, C., Burns, S., & Griffin, P. (Eds.) (1998). *Preventing reading difficulties in young children*. National Academy Press.

Snowling, M., & Hulme, C. (2012). Interventions for children's language and literacy difficulties. *International Journal of Language & Communication Disorders*, 47, 27–4. https://doi.org/10.1111/j.1460-6984.2011.00081.x

Sparks, R. (1995). Examining the Linguistic Coding Differences Hypothesis to explain individual differences in foreign language learning. *Annals of Dyslexia*, 45, 187–14. https://doi.org/10.1007/BF02648218

Sparks, R. (2012). Individual differences in L2 learning and long-term L1-L2 relationships. *Language Learning*, 62, 5–27. https://doi.org/10.1111/j.1467-9922.2012.00704.x

Sparks, R., Artzer, M., Patton, J., et al. (1998). Benefits of multisensory language instruction in Spanish for at-risk learners: A comparison study of high school Spanish students. *Annals of Dyslexia*, 48, 239–270. https://doi.org/10.1007/s11881-998-0011-8

Sparks, R., & Ganschow, L. (1991). Foreign language learning difficulties: Affective or native language aptitude differences? *Modem Language Journal*, 75, 3–16. https://doi.org/10.2307/329830

Sparks, R., & Ganschow, L. (1993a). Searching for the cognitive locus of foreign language learning problems: Linking first and second language learning. *Modern Language Journal*, 77, 289–302. https://doi.org/10.2307/329098

Sparks, R., & Ganschow, L. (1993b). The effects of a multisensory structured language approach on the native and foreign language aptitude skills of at-risk learners: A replication and follow-up study. *Annals of Dyslexia*, 43, 194–216. https://doi.org/10.1007/BF02928182

Sparks, R., & Ganschow, L. (1995). A strong inference approach to causal factors in foreign language learning: A response to Macintyre. *Modern Language Journal*, 79, 235–244. https://doi.org/10.1111/j.1540-4781.1995.tb05436.x

Sparks, R., Ganschow, L., Patton, J., Artzer, M., Siebenhar, D., & Plageman, M. (1997). Prediction of proficiency in a foreign language. *Journal of Educational Psychology*, 89, 549–561. https://doi.org/10.1037/0022-0663.89.3.549

Sparks, R., Ganschow, L., Pohlman, J., Artzer, M., & Skinner, S. (1992). The effects of a multisensory, structured language approach on the native and foreign language aptitude skills of high-risk foreign language learners. *Annals of Dyslexia*, 42, 25–53. https://doi.org/10.1007/BF02654937

Sparks, R., Humbach, N., & Javorsky, J. (2008). Individual and longitudinal differences among high and low-achieving, LD, and ADHD L2 learners. *Learning and Individual Differences*, 18, 29–43. https://doi.org/10.1016/j.lindif.2007.07.003

Sparks, R., Luebbers, J., & Castañeda, M. (2017). How well do US high school students achieve in Spanish when compared to native Spanish speakers? *Foreign Language Annals*, 50, 339–366.

Sparks, R., & Patton, J. (2013). Relationship of L1 skills and L2 aptitude to L2 anxiety on the Foreign Language Classroom Anxiety Scale (FLCAS): Language ability or anxiety? *Language Learning*, 63, 870–895. https://doi.org/10.1111/lang.12025

Sparks, R., Patton, J., Ganschow, L., & Humbach, N. (2009). Long-term relationships among early L1 skills, L2 aptitude, L2 affect, and later L2 proficiency. *Applied Psycholinguistics*, 30, 725–755. https://doi.org/10.1017/S0142716409990099

Sparks, R., Patton, J., Ganschow, L., & Humbach, N. (2011). Subcomponents of second-language aptitude and second-language proficiency. *Modern Language Journal*, 95, 1–21. https://doi.org/10.1111/j.1540-4781.2011.01176.x

Sparks, R., Patton, J., Ganschow, L., Humbach, N., & Javorsky, J. (2006). Native language predictors of foreign language proficiency and foreign language aptitude. *Annals of Dyslexia*, 56, 129–160. https://doi.org/10.1007/s11881-006-0006-2

Sparks, R., Patton, J., Ganschow, L., Humbach, N., & Javorsky, J. (2008). Early L1 reading and spelling skills predict later L2 reading and spelling skills. *Journal of Educational Psychology*, *100*, 162–174. https://doi.org/10.1037/0022-0663.100.1.162

Sparks, R., Patton, J., & Luebbers, J. (2019). Individual differences in L2 achievement mirror individual differences in L1 skills and L2 aptitude: Crosslinguistic transfer of L1 to L2 skills. *Foreign Language Annals*, *52*, 255–283. https://doi.org/10.1111/flan.12390

Stansfield, C., & Reed, D. (2019). The MLAT at 60 years. In E. Wen et al. (Eds.), *Language aptitude: Advancing theory, testing, research, and practice* (pp. 15–32). Routledge.

Symonds, P. (1930). *Foreign Language Prognosis Test*. Teachers College, Columbia University.

Torgesen, J., Wagner, R., & Rashotte, C. (1999). *Test of word reading efficiency*. PRO-ED.

Tunmer, W., Herriman, M., & Nesdale, A. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly*, *23*, 134–158. https://doi.org/10.2307/747799

Wagner, R., Torgesen, J., Rashotte, C. (1999). *CTOPP-2: Comprehensive Test of Phonological Processing*. PRO-ED.

Wells, G. (1985). *Language development in the pre-school years* (Vol. *2*). Cambridge University Press.

Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, *50*, 1–31. https://doi.org/10.1017/S0261444816000276

Wen, Z., Mota, M., & McNeil, A. (2015). *Working memory in second language acquisition and processing*. Multilingual Matters.

Wiig, E., & Secord, W. (1989). *Test of Language Competence-Expanded Edition*. Psychological Corporation.

Woodcock, R. (1998). *Woodcock Reading Mastery Test-R (WRMT-R)*. AGS.

Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson-III Tests of Achievement, Tests of Cognitive Abilities*. Riverside.

Woodcock, R., Muñoz-Sandoval, A., McGrew, K., & Mather, M. (2004). *Batería-III Woodcock-Muñoz*. Riverside.

Yelland, G., Pollard, J., & Mercuri, A. (1993). The metalinguistic benefits of limited contact with a second language. *Applied Psycholinguistics*, *14*, 423–444. https://doi.org/10.1017/S0142716400010687