RESEARCH ARTICLE

Mapping the predictive role of MLAT subtests for L2 achievement through regression commonality analysis

Philip S. Dale[1] (ID) and Richard L. Sparks[2] (ID)

[1]University of New Mexico, Albuquerque, NM, USA; [2]Mt. St. Joseph University, Cincinnati, OH, USA
Corresponding author: Philip S. Dale; Email: dalep@unm.edu

Abstract

Despite the widespread use and effectiveness of the Modern Language Aptitude Test (MLAT) composite score in predicting individual differences in L2 achievement and proficiency, there has been little examination of MLAT subtests, although they have potential for illuminating components of L2 aptitude and the mechanism of prediction. Here we use regression commonality analysis to decompose the predictive variance from the MLAT into unique components for each subtest alone and for each possible combination of subtests (duos, trios, etc.) that may have shared variance. The results, from a longitudinal study of 307 U.S. secondary students during 2 years of Spanish learning, provide strong evidence for the role of literacy-related skills in all subtests and in predicting all L2 outcomes. These and other results support a view of L1 literacy and language skills leading to metalinguistic development, which in turn leads to stronger L2 aptitude and achievement.

Introduction

The Modern Language Aptitude Test (MLAT) developed by Carroll and Sapon (1959/2000) has been the most influential and well-researched measure of language aptitude since its publication. Carroll proposed that aptitude is a predictor of the rate of language learning; therefore, one use of an aptitude test is to identify individuals who could master a foreign language (L2) in a prescribed, and often limited, time (Li, 2019). Based on factor analyses, Carroll (1962) identified four components of language aptitude, which are represented by five subtests on the MLAT. Over time, the MLAT has been found to be the strongest single predictor of L2 achievement and proficiency (see reviews by Li, 2016; Skehan, 2002; Stansfield & Reed, 2019).

Most studies with the MLAT have used the composite (total) score reported in the test manual for prediction of L2 achievement. This practice reflects the view that the individual subtests do not correspond in a one-to-one fashion to the underlying four components for L2 acquisition (Stansfield & Reed, 2019). This view of collective rather than discrete assessment by the subtests has been confirmed by correlational and

regression studies showing that overall aptitude has greater predictive power than the individual subtests (Li, 2015). At most, researchers have examined the relative role of the subtests for predicting L2 achievement by comparing the regression weights for each subtest when all five subtest scores are included as predictors in the regression analysis. However, regression analyses can be misleading when the predictor variables are substantially intercorrelated, as they are for the MLAT subtests (see Sparks et al., 2011, 2019). This is because shared variance is entirely assigned to the predictor with the larger total prediction, thus potentially overestimating its effect and consequently underestimating the role of other variables.

When predictor variables are substantially correlated, the results of conventional multiple regression are less informative about the role of each variable because shared variance will be attributed to the variable that is earlier in the equation, by either a priori ordering or the use of a hierarchical procedure in the analysis. A method that partially addresses this problem is *uniqueness* analysis. In this procedure, each predictor of interest is specified as entered last in an analysis so that the result estimates the effect that is due to that variable alone. This can be done in turn for each predictor variable. Regression commonality analysis (RCA; Nimon, 2010) extends this approach beyond individual variables to components of shared variance—that is, to a pair, trio, or other combination of variables. The RCA regression coefficients are measures of the *unique* variance associated with each variable and each subset of variables, and therefore they are free of the multicollinearity interpretation problem. The pattern of results concerning shared effects can provide helpful insight in substantive issues as well. As a hypothetical example, it might be that one MLAT subtest, Paired Associates, is a significant predictor but is less powerful uniquely than via the shared variance between Paired Associates and another subtest, Phonetic Script. Such a finding would have important implications for explaining the role of Paired Associates in predicting L2 achievement. In the present study, we employ RCA to examine the role of the five MLAT subtests in predicting L2 achievement.

A second set of analyses is motivated by the well-established finding that L1 measures are strong predictors of L2 achievement (see Sparks, 2022a, 2022b) and more recent findings showing that the MLAT adds only a modest amount to the prediction of L2 achievement after the variance accounted for by L1 skills (Sparks & Dale, 2023a, 2023b; Sparks et al. 2023). That is, most of the prediction from the MLAT for L2 achievement is because MLAT is capturing the variance in L1 skills. This raises the question of what precisely the MLAT is measuring that goes beyond L1 achievement. To explore this question directly, we conducted a second set of RCA analyses in which the dependent variable was not the full range of each L2 score but the variance that remained *after* the L1 measures were used to predict the L2 achievement score.

Review of the literature
Modern Language Aptitude Test

In the 1950s, John Carroll conducted studies on L2 aptitude based on the proposition that facility to learn an L2 is a specialized talent or group of talents, independent of intelligence. Through his factor analytic studies, Carroll found that four components treated as independent variables proved to be most relevant to L2 learning: (a) *phonetic coding*, or the ability to identify speech sounds and the symbols representing them and to retain the 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 linguistic forms, rules, and patterns from new linguistic content; and (d) *rote learning ability for foreign language materials*, or the ability to learn associations between sounds or words rapidly and to recall the associations (Carroll, 1962). The MLAT measures L2 aptitude using a simulated format (i.e., a "fake" L2) along with English grammar and vocabulary to provide an indication of one's probable degree of success in learning an L2.

Initially, Carroll developed 30 different types of test items, then through his analyses, identified five item types that provided some unique variance while also being predictive of the global construct of L2 aptitude. The MLAT uses these five item types, or subtests, to measure the four components that Carroll found to be important for L2 learning. Part I, *Number Learning*, involves learning numbers in an artificial language of nonsense words with English sounds, then listening to the words and writing the numbers in a timed format. Part II, *Phonetic Script*, requires the student to learn symbols, some previously learned and some unique, for English sounds and to remember the sound–symbol relationships by selecting a written word that represents a spoken stimulus. Part III, *Spelling Clues*, involves reading (decoding) a stimulus word written in an incomplete phonetic spelling for English, then choosing the correct meaning of the word from five English vocabulary words. Part IV, *Words in Sentences*, requires the student to read two English sentences, then choose the word in the second sentence that has the same grammatical function as the underlined word in the first sentence. Part V, *Paired Associates*, requires the student to learn a series of 24 words written in a made-up language and match these words with their English equivalents in a brief, designated period. Supplemental Appendix SA1 presents additional information and examples of items from each of the five subtests.

Since their introduction, L2 aptitude tests, particularly the MLAT, have consistently been found to be the single strongest predictors of L2 achievement. The MLAT has been shown to have reliably strong correlations ($r = .40$–$.60$) with performance in L2 courses and research has confirmed the utility of the MLAT for predicting ultimate L2 achievement (Skehan, 2002). For example, Ehrman (1998) found that the MLAT was the best predictor of L2 achievement for adults at the Foreign Service Institute among a diverse set of instruments. In another study with adults, Ehrman and Oxford (1995) found that among a number of variables, the MLAT showed the strongest correlation with L2 proficiency. In a 10-year study with students from 1st to 10th grades, Sparks et al. (2009) found that the MLAT was the strongest single predictor of overall L2 achievement ($r = .75$) and performance on L2 reading, writing, speaking, and listening comprehension measures ($r = .54$–$.72$). In a study that followed a large group of L2 learners ($n = 262$) over 2 years of Spanish and measured their Spanish skills each year with a standardized achievement measure, Sparks et al., 2023) found that there were strong correlations between the MLAT and L2 reading ($r = .38$–$.53$), L2 writing ($r = .29$–$.47$), L2 listening comprehension ($r = .42$–$.54$), and L2 oral proficiency ($r = .42$). A recent meta-analysis of aptitude studies over 50 years (Li, 2015) found a positive correlation ($r = .34$) between MLAT scores and L2 learning outcomes.

In summary, converging evidence over many years has shown that the overall MLAT score is a strong predictor of subsequent L2 achievement.

Relationships among L1 achievement, L2 aptitude, and L2 achievement

In addition to the consistent finding of strong prediction from the MLAT to L2 achievement, in recent decades much research has demonstrated a close relation between achievement in learners' first language (L1) and their L2 achievement. It is

now well recognized that although most children learn to communicate sufficiently well in their L1 without clinical intervention, there is substantial normal variation in their rate of acquisition and communication skills across all components of the language system (see Gilkerson et al., 2017; Hoff, 2013; Huttenlocher et al., 2010) and individual differences (IDs) in L1 oral skills are strongly related to students' later acquisition of L1 literacy skills (Kendeou et al., 2009). In their extensive review of the literature, Kidd et al. (2018) concluded that large and stable IDs are pervasive across all domains of language in both children and typically developing adults in their ultimate attainment (see also Kidd & Donnelly, 2020).

In a groundbreaking study, Skehan and Ducroquet (1988) assessed students from the Bristol Language Project (Wells, 1985) that had explored and quantified individual differences in the children's L1 attainment prior to 5 years of age and found that those scores were strongly correlated with their L2 aptitude and L2 achievement at ages 13–16. L2 aptitude scores continued to be the strongest single predictor of success in L2 achievement. Prediction of L2 achievement by L2 aptitude tests was augmented by the inclusion of L1 achievement measures, notably language comprehension and vocabulary growth. Skehan and Ducroquet speculated 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). L2 aptitude tests may measure an underlying language learning capacity that is similar in L1 and L2. In addition, aptitude tests also likely function as a "measure of the ability to learn from decontextualized material" (p. 34). For example, the items on aptitude tests demand that students must be able to use their language analytic abilities to "think" about how language works rather than to simply use their language.

Following Skehan and Ducroquet (1988), Sparks and Ganschow (1991, 1993) found that (a) high, average, and low-achieving L2 learners will display IDs in their L1 skills; (b) IDs in L1 predict ultimate attainment in the L2; (c) students' L1 skills are related to and consistent with their aptitude for L2 learning; (d) L2 achievement is moderated by L1 achievement; and (e) L1 and L2 achievement have a common underlying foundation (see also Cummins, 1979). Their studies over many years have consistently found strong relationships among L1 achievement, L2 aptitude, and L2 achievement (Sparks, 2022a, 2022b). Even so, the overall MLAT has been shown to be the strongest single predictor of L2 achievement (Sparks et al., 2009). They explained these findings by proposing, like Skehan (2002), that the MLAT may preempt ("cut out") the variance explained by L1 skills because the MLAT measures an "underlying language learning capacity which is similar in first and second language learning settings" and also assesses the ability to learn from decontextualized material. Sparks et al. also proposed that the MLAT gained its predictive value from tapping into students' metalinguistic ability—that is, the ability to think about, reflect on, and manipulate language more generally (Ranta, 2002).

The MLAT as a Measure of L1 Skills

Evidence that early L1 achievement is a strong predictor of L2 aptitude on the MLAT but that the MLAT is the best single predictor of L2 achievement prompted studies to explore the relationship of L1 and MLAT scores as predictors, individually and collectively. Sparks et al. (2023) asked the extent to which the prediction from the MLAT for L2 achievement is due to the MLAT's assessment of underlying L1 abilities. In their first study with U.S. participants followed over 3 years of high school Spanish, they conducted a series of hierarchical regressions with a fixed order of entry in which

all L1 achievement skills were entered first followed by MLAT to predict L2 reading, writing, listening comprehension, and vocabulary achievement. The results showed that L1 achievement explained from 21%–50% of the variance in L2 achievement skills, and that the variance explained by L1 skills increased from 1st to 2nd to 3rd-year Spanish. IDs in L1 literacy skills—for example, word decoding–were the best predictors of all L2 written and oral achievement. The MLAT explained an additional 2%–14% of variance in L2 achievement, which suggested that the MLAT measures important aspects of language ability that are not tapped by the L1 measures.

In a second study, Sparks and Dale (2023a) proposed that mapping the quantitative relations among L1 and L2 achievement and L2 aptitude would benefit by distinguishing two aspects of prediction from the MLAT to L2 achievement: *uniqueness*, or the extent to which L2 aptitude uniquely adds to the prediction of L2 outcomes beyond that explained by L1 achievement alone, and *efficiency*, or the extent to which the total prediction from a specific L1 measure to a specific measure of L2 achievement is mediated by the MLAT. The prediction from MLAT scores to L2 achievement was significantly and primarily (59%–87%) due to variance in L1 abilities captured by the MLAT—that is, uniqueness was low and efficiency was moderate.

In summary, prediction from MLAT to L2 achievement appears to be largely due to MLAT's assessment of L1 abilities, even though a substantial amount of L2 prediction-relevant L1 variance is missed by MLAT. The findings suggest that Carroll's (1973) speculation about L2 aptitude as a "remnant" of L1 achievement may have validity but that the MLAT also measures "language analytic," or metalinguistic, abilities that are not captured by L1 achievement tests.

Purpose of study and research questions

Carroll had little interest in the individual MLAT tests for prediction, preferring to use the composite score. This preference largely reflected the substantial correlations among the subtests, along with the lack of a simple one-to-one relationship between the MLAT subtests and the four hypothesized components of L2 aptitude. Despite the MLAT's long history of use for prediction, very little research has focused on the five subtests. Wesche (1981) suggested that the subtests could be used for matching students' strengths and weaknesses with teaching methods. In studies with secondary and postsecondary L2 students, Sparks et al. (1992) found significant differences between high- and low-achieving learners on all MLAT subtests (see reviews by Sparks, 2022a, 2022b). Li (2016) reported that phonetic coding is a significant predictor of L2 vocabulary learning ($r =$ .38) and that two MLAT subtests, Number Learning and Spelling Clues, were significant predictors of L2 writing ($r = .42$).

Contemporary cognitive science has a strong focus on identifying and assessing individual cognitive processes and their interrelationships, and improved statistical methods are available for this purpose. There are three broad goals for this study. The first is to determine the extent to which using the five subtests rather than the composite score adds to the prediction of L2 achievement, either on their own or in combination with L1 measures. The second is to use RCA to identify the unique contribution to prediction made by each subtest as well as that made by the shared variance among each possible combination of two or more subtests. The third is to compare the prediction by the subtests with the full distribution of L2 achievement scores with the prediction to just the variance remaining after the prediction by L1 scores has been removed—that is, the extent to which the MLAT provides information beyond the L1 measures. These goals are addressed with the following four research questions:

Research Question 1: Do MLAT subtests significantly add to the prediction of each L2 achievement measure beyond that of the composite score alone?

Research Question 2: Do MLAT subtests significantly add to the prediction of each L2 achievement measure beyond that of the composite score alone in predicting the variance in each L2 measure that is *not* predicted by the L1 measures—that is, the "residuals of prediction"?

Research Question 3: Of the total set of variance components (five MLAT subtests uniquely, plus 26 shared variance components, total of 31 unique components), which are the most important predictors of the full variance of each L2 achievement measure?

Research Question 4: Of the total set of variance components (five MLAT subtests uniquely, plus 26 shared variance components, total of 31 unique components), which are the most important predictors of the variance of each L2 measure *not* predicted by the L1 measures (residuals of prediction)?

It should be noted that Research Questions 3 and 4 are essentially descriptive, rather than inferential in nature. Regression commonality analysis is a technique for partitioning variance in a predictive relationship that has already been established as significant, which Research Questions 1 and 2 will address.

Method

The participants, measures, and data set for this report have been drawn from the study reported in Sparks et al. (2019).

Participants

The initial sample for the study included 307 participants (154 male; 153 female) randomly chosen from students in first-year Spanish courses at one of four high schools in a large Midwestern suburban school district. The mean age was 15 years, 7 months (range 13;7 to 17;6). The students were enrolled in 9th, 10th, and 11th grades at the beginning of the study. The sample comprised 301 Caucasian, four African American, and two East Asian students. Two hundred and ninety-three (148 females and 145 males) of the 307 students completed at least the first-year Spanish course, and 267 students completed both first- and second-year courses. 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. Review and approval of the study was conducted by the university IRB committee. Parental consent 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 10 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 memory, L2 aptitude, and L2 (Spanish) achievement. Each of the measures is briefly

described here. As the L1 measures were used only in a preliminary analysis to estimate the variance in L2 achievement that was not predicted by them, they are simply listed here; a complete description of each L1 measure, including references, is provided in Supplementary Appendix SA2. A complete description of the L2 aptitude and L2 achievement measures is provided in Supplementary Appendix SA3. Reliability coefficients for the instruments are reported in the Appendices.

L1 (English) Achievement

L1 Word decoding. The two measures of word decoding were the Woodcock Reading Mastery Test-Revised Basic Skills Cluster and the Test of Word Reading Efficiency.

L1 Reading Comprehension. The measure of L1 reading comprehension was the Stanford Achievement Test 10.

L1 Vocabulary. The measure of L1 vocabulary was the Woodcock-Johnson-III/NU Picture Vocabulary subtest.

L1 Language Analysis. The measure of language analysis was the Test of Language Competence-Expanded Edition Figurative Language subtest.

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 Memory. The measure of phonological short-term memory was the Comprehensive Test of Phonological Processing, Phonological Memory Composite. The measure of working memory was the Woodcock-Johnson-III/NU Working Memory Cluster.

L2 Aptitude

The measure of L2 aptitude was the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959/2000). This standardized test uses a simulated format to provide an indication of the probable degree of success in learning an L2. It is a pen-and-paper test, with some prerecorded stimuli. 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, 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.

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 on the second year, students' oral proficiency in Spanish was assessed via a 10–15-min individual interview, using a researcher-designed measure (see Sparks et al., 2006). The interviews were conducted by two L2 educators, who had been trained to conduct oral interviews, and graduate students trained by them. The interviewers had no previous knowledge about the participants, who were assigned randomly to an interviewer. The interviewers used prompts similar to those in Sparks et al. (2006). Each student's interview was recorded for scoring later by the two L2 educators. The interview was scored for five criteria according to a rubric developed by the L2 educators adapted from the ACTFL Speaking Guidelines (1999) and the AAPPL Rating Criteria (2017): vocabulary and discourse range, comprehensibility (accent and pronunciation), language comprehension, language control (grammar, word choice, word order), and task completion (score of 0–4 for each part, maximum composite score = 20). Cronbach's alpha for the L2 oral proficiency measure was .89.

Procedure

The testing instruments were administered at different times over the course of the study. The MLAT was administered in groups of 25–30 students by the second author in the first 3–4 weeks of the first-year Spanish course. The L1 measures were administered individually by the second author, who was assisted by a university Spanish professor and graduate students trained by the second author. 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 participants at the end of the first- and second-year courses by the second author, the university 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-Muñoz* 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. For this study, participants' scores according to 9th-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 analysis

The data analyses were conducted within SPSS v29.0.0.0, except as specifically noted. Descriptive statistics for all study measures were computed, and the relevant zero-order correlations among L1 skills, L2 aptitude, and L2 achievement were also computed. Data imputation (a set of processes for replacing missing data with substituted values based on other information about the case and the relations among variables, to maximize sample size) was not used in this study, as its effects on RCA are not known.

For Research Questions 1 and 2, correlations and multiple regression analyses with specified entry order were calculated and comparisons made using confidence intervals obtained from SPSS (for bivariate correlations) or from the R^2 confidence interval calculator available at https://www.danielsoper.com/statcalc/calculator.aspx?id=28 (for multiple R values resulting from regression). For Research Questions 3 and 4, RCA was conducted using the SPSS script developed by Nimon (2010), and the variance components accounting for at least 6% of the prediction were identified. (This criterion is somewhat arbitrary; it was chosen because it selected the 4–6 variance components—out of 31—that made the largest relative contribution within an overall prediction that was highly significant.) For Research Questions 2 and 4, multiple regression analyses using all L1 measures as predictors of each L2 measure were first conducted, and the residuals—the variance not accounted for by L1—were saved and used as the dependent variable for a second set of regression (RQ2) and regression commonality analyses (RQ4). (We note for R users that information on conducting RCA in that package is provided in Nimon et al., 2008).

Results

Preliminary analyses

Table 1 presents descriptive statistics for all study measures. The MLAT Composite (Long Form) has norms that compare students against their grade level (9th, 10th, 11th) at the time they completed the test. The MLAT score is reported as a standard score ($M = 100$, $SD = 15$). The MLAT does not report norms for the individual subtests, so subtest scores are reported as raw scores. Students' scores on the L1 and L2 achievement tests are reported as standard scores. The L2 oral proficiency score is reported as a raw score.

Table 2 presents the correlations among the five MLAT subtests and the composite measure. With the exception of Paired Associates-Spelling Clues, all subtest correlations are significant, though weak to moderate in magnitude. The Number Learning subtest had the highest loading with the MLAT composite score.

Table 3 presents the prediction from the MLAT subtests and the MLAT composite score to all L2 achievement measures. Phonetic Script was consistently the strongest single predictor.

To estimate the stability of relative performance for students across the years, the Year-1 to Year-2 correlations for L2 Reading, Listening Comprehension, and Writing were computed as .62, .66, and .83 respectively.

Comparison of MLAT composite score with subtests as predictors of L2 achievement
(RQ 1 and 2)
Prediction of L2 achievement by MLAT composite versus subtests (RQ 1). Table 4 summarizes the comparison of the MLAT composite score with the five subtests in

Table 1. Descriptive statistics for all study measures

Measure	n	Minimum	Maximum	M	SD
L1 Vocabulary [b]	307	60	129	93.3	10.57
L1 Word decoding [b]	307	69.5	131.5	100.42	10.16
L1 Memory [b]	307	66.5	127	97.18	10.58
L1 Lang Analysis [c]	307	1	12	9.25	1.68
L1 Reading Comp [b]	307	53	135	107.99	12.32
L1Writing [d]	307	5	16	10.83	1.97
Number Learning [a]	307	3	43	26.36	10.01
Phonetic Script [a]	307	12	30	22.22	3.42
Spelling Clues [a]	307	1	37	11.84	5.59
Words in Sentences [a]	307	3	23	11.69	3.59
Paired Associates [a]	307	0	24	11.53	5.13
MLAT Composite SS [b]	307	65	128	97.39	11.69
L2 Reading Yr 1 [b]	293	3.0	68.5	28.75	12.97
L2 Reading Yr 2 [b]	267	1.0	69.0	38.56	13.62
L2 Listen Comp Yr 1 [b]	293	4.0	52.0	12.55	7.22
L2 Listen Comp Yr 2 [b]	267	4.0	40.0	20.16	8.31
L2 Writing Yr 1 [b]	293	19.5	79.0	35.92	9.99
L2 Writing Yr 2 [b]	267	19.5	89.5	46.59	12.66
Span 2 Oral Prof [a]	256	1.0	16.0	8.32	2.20
Valid n (listwise)	240				

[a]Raw scores.
[b]Standard scores (M = 100, SD = 15).
[c]Standard score (M = 10, SD = 3).
[d]Standard score (M = 10.1, SD = 2.9) (range 1–16).

Table 2. Correlations among MLAT subtests and MLAT composite (long form) score

Measure	1	2	3	4	5	6
1. Number Learning	–					
2. Phonetic Script	.356**	–				
3. Spelling Clues	.198**	.241**	–			
4. Words in Sentences	.228**	.274**	.332**	–		
5. Paired Associates	.195**	.270**	.104	.153**	–	
6. MLAT Composite SS	.771**	.565**	.539**	.499**	.464**	–

Note. All n = 307;
*p < .05;
**p < .01.

predicting L2 achievement. For all seven L2 measures, the confidence interval for the correlations overlaps with the R from the regression. That is, for none of the L2 measures examined individually is the multiple regression R greater than the correlation based on the MLAT composite. However, the finding that all seven analyses show differences in the same direction (prediction from subtests is greater than that from composite score), suggests that with a larger sample size, the differences for individual measures might be significantly different; even so, the differences are small. The predictions are also greater for the Year-2 L2 achievement scores than those for the Year-1 L2 achievement scores, although the confidence intervals again overlap.

Contribution of MLAT composite versus subtest scores in prediction of variance in L2 achievement not predicted by the L1 measures (RQ 2). Table 5 summarizes the comparison of the MLAT composite score with the set of five MLAT subtests as

Table 3. Prediction from MLAT subtests and composite to L2 measures

MLAT Measure	Reading-yr1 $n = 295$	Reading-yr2 $n = 267$	Listen Comp-yr1 $n = 295$	Listen Comp-yr2 $n = 267$	Writing-yr1 $n = 295$	Writing-yr2 $n = 267$	Oral Prof-yr2 $n = 256$
Number Learning	.279**	.261**	.279**	.312**	.336**	.345**	.195**
Phonetic Script	.374**	.459**	.337**	.402**	.394**	.432**	.282**
Spelling Clues	.211**	.306**	.245**	.216**	.224**	.278**	.214**
Words in Sentences	.172**	.263**	.255**	.298**	.182**	.262**	.253**
Paired Associates	.228**	.260**	.318**	.332**	.307**	.321**	.156*
MLAT Composite SS	.385**	.453**	.415**	.441**	.421**	.473**	.303**

*$p < .05$,
**$p < .01$

predictors of the variance in L2 achievement that is not predicted by the L1 measures—that is, the residual of prediction. The improvement in prediction from using the five subtests is modest, and the confidence intervals overlap.

Summary of regression analyses for RQs 1 and 2

Taken together, the very modest and nonsignificant results of these analyses addressing the first two research questions suggest that the correlation among the MLAT subtests renders conventional multiple regression limited in its ability to provide information about the role of specific MLAT subtests. Regression commonality analysis, which identifies and integrates the role of shared variance among correlated subtests, appears to be more promising.

Regression commonality analysis to identify the most important predictors of each L2 achievement measure (RQ 3)

Regression commonality analysis was conducted to examine the total set of variance components of prediction (the five MLAT subtests uniquely, plus 26 shared variance components, total of 31 components). The results are summarized in Table 6, which includes all components predicting at least 6% of the total prediction, in order of effect size. For L2 Reading outcomes, Phonetic Script was the largest contributor in both years. Paired Associates also made a contribution in both years. Number Learning had a substantial prediction in Year 1, whereas Paired Associates showed substantial prediction in Year 2. Spelling Clues made a substantial contribution in Year 2. Shared components for prediction of L2 reading are also included in Table 6. There was commonality between Phonetic Script and Paired Associates in both years and commonality between Phonetic Script and Number Learning in Year 1.

For L2 Writing outcomes, Phonetic Script was the largest contributor in both years. Paired Associates made a contribution in both years, along with the shared contribution between Phonetic Script and Paired Associates. Number Learning, both individually and in shared contributions with Phonetic Script, also made a substantial contribution. The results for Years 1 and 2 were similar.

For L2 Listening Comprehension outcomes, Phonetic Script, and Paired Associates, both individually and in shared components, made the largest contribution to prediction in both Years 1 and 2. Number Learning, either individually or in shared contributions with Phonetic Script, also contributed. The only notable change between Year 1 and Year 2 is that Words in Sentences made a contribution to prediction in Year 2.

Table 4. Prediction of L2 achievement by MLAT composite score versus entry of five subtests individually

L2 measure	n	Correlation r (CI) with MLAT composite	R (CI) from regression with five MLAT subtests
Reading–Yr 1	294	.385 (.284–.470)	.437 (.335–.521)
Reading–Yr 2	266	.453 (.352–.543)	.531 (.439–.609)
Writing–Yr 1	294	.421 (.322–.510)	.497 (.272–.469)
Writing–Yr 2	266	.473 (.375–.562)	.542 (452–.620)
Listening Comp–Yr 1	294	.415 (.316–.505)	.470 (.373–.550)
Listening Comp–Yr 2	266	.441 (.339–.533)	.519 (.424–.599)
Oral Proficiency Index	255	.303 (.188–.411)	.360 (.235–.453)

Table 5. Prediction of residual of prediction of L2 achievement after prediction from L1 measures, by MLAT composite score versus entry of five MLAT subtests individually

L2 measure	n	Correlation r (CI) of MLAT with residual of prediction from L1 measures	R (CI) from regression from five MLAT subtests predicting residual of prediction from L1 measures
Reading–Yr 1	294	.139 (.025–.249)	.236 (.077–.326)
Reading–Yr 2	266	.179 (.060–.262)	.280 (.134–.373)
Writing–Yr 1	294	.210 (.098–.317)	.313 (.187–.402)
Writing–Yr 2	266	.216 (.099–.328)	.310 (.173–402)
Listening Comp–Yr 1	294	.247 (.137–.351)	.300 (.170–389)
Listening Comp–Yr 2	266	.214 (.096–.326)	.326 (.195–.418)
Oral Proficiency Index	255	.136 (.013–.254)	.183 (0–.276)

For the Oral Proficiency Index (Year 2 only), Phonetic Script was the strongest predictor, but Words in Sentences played a considerably stronger role than it did for any of the other L2 outcomes. Words in Sentences made a contribution individually and in shared variance with Phonetic Script and Spelling Clues. Spelling Clues also made an individual contribution.

In summary, Phonetic Script was the largest contributor to all measures except Listening Comprehension–Yr1 where it was second. It was also frequently a contributor in shared variance with other subtests. Paired Associates was also a major contributor, especially for Writing and Listening Comprehension measures, as was Number Learning. Words in Sentences was a major contributor only to Oral Proficiency.

Regression commonality analysis to identify the most important predictors of each L2 measure not predicted by the L1 measures (RQ 4)

This research question asked how the individual MLAT subtests contribute to the uniqueness of MLAT—that is, the extent to which its prediction goes beyond the contribution of the L1 achievement measures. As in the analysis for RQ3 (Table 6), a regression was computed for each L2 measure using all L1 measures and the residual of that prediction was used as the target of the prediction. The difference here is that RCA was used to evaluate prediction rather than multiple regression. Given the strong Year 1–Year 2 stability of the L2 measures, the similar correlations of the MLAT Composite with Year 1 and Year 2 measures (slightly higher for Year 2), and the similar results for

Table 6. MLAT variance components with the largest contribution to the prediction of L2 measures

L2 measure–year obtained	Variance component	% Contribution to prediction
Reading–Yr 1	PS	30.1
	NL PS	11.7
	NL	9.0
	PS PA	7.1
	PA	6.2
Reading–Yr 2	PS	33.1
	SC	10.3
	PS PA	7.1
	PA	6.0
Writing–Yr 1	PS	20.7
	PA	13.1
	NL	12.6
	NL PS	11.1
	PS PA	8.2
Writing–Yr 2	PS	19.9
	PA	12.1
	NL PS	8.8
	PS PA	7.4
	NL	7.0
Listening Comp–Yr 1	PA	18.9
	PS	11.7
	PS PA	7.4
	NL	6.4
Listening Comp–Yr 2	PS	16.9
	PA	15.5
	PA PS	7.9
	NL PS	7.1
	WS	6.7
Oral Proficiency Index	PS	19.5
	WS	13.9
	PS WS	6.4
	SC WS	6.3
	SC	6.2

Note. PS = Phonetic Script; PA = Paired Associates; SC = Spelling Clues; NL = Number Learning; WS = Words in Sentences.

the regression commonality analyses for the two years shown above (RQ3), these analyses were conducted only for Year-2 L2 achievement measures. Table 7 summarizes the analyses. Similar to the analyses in Table 6, all components predicting at least 6% of the total prediction are listed in the far right column in order of effect size. Of particular interest are changes—that is, increases or decreases, in the contribution of the individual subtests when the dependent variable changes from the full variability in the L2 measure to predicting the residual of prediction—the portion of the variance that is not predicted by the L1 measures.

For L2 Reading, the strongest unique contribution of MLAT was from Phonetic Script, by itself or in shared variance with Paired Associates. The effect was much stronger for residual of prediction than for full distribution of Reading. Spelling Clues was no longer a substantial predictor.

For L2 Writing, Phonetic Script was similar in importance in both analyses, whereas Paired Associates became much more substantial for predicting the residuals than for the full distribution. Otherwise, there is little difference in results of the two analyses.

For L2 Listening Comprehension, the role of Paired Associates, by itself and in shared variance with Phonetic Script and Number Learning, became the most important contributors. The unique role of Phonetic Script decreased in the analysis for residuals. The role of Words in Sentences was diminished in this analysis.

The results for the L2 Oral Proficiency Index in this analysis was distinct from the other L2 outcomes. Words in Sentences, individually and in shared variance with Phonetic Script, became the most important contributor. The role of Spelling Clues diminished, but Paired Associates also became an important predictor.

Discussion

We proposed three broad goals for this study: (a) determine the extent to which using the five MLAT subtests rather than just the composite score adds to the prediction of L2 achievement; (b) use RCA to identify the unique contribution to prediction made by each MLAT subtest individually and that made by the shared variance among each possible combination of two or more subtests; and (c) compare the prediction by the MLAT subtests to the full distribution of L2 achievement scores with prediction to the variance remaining after the prediction by L1 scores has been removed—that is, the extent to which the MLAT subtests provide information beyond the L1 measures. These goals were addressed with four research questions, each of which is discussed before turning to reconsideration of the content of the MLAT subtests and the meaning of the results for a theory of L2 aptitude.

MLAT subtest scores versus MLAT composite score for prediction of L2 achievement

Our first set of analyses generated estimates for the improvement in prediction of L2 achievement when the five MLAT subtest scores were used rather than the one MLAT composite score. These analyses were conducted in two contexts of prediction: the MLAT predicting the full range of L2 achievement scores (RQ1) and the MLAT predicting variance in L2 achievement not accounted for by the L1 measures (RQ 2). The results of these 14 analyses shown in Tables 4–5 are essentially identical: although the prediction results are stronger in both cases when the five MLAT subtests are used, in none of the analyses is the difference significant. That is, prediction for L2 achievement increased with the MLAT subtests, but the confidence intervals overlapped in each case. These results imply that there is no difference between using the MLAT composite or using the five MLAT subtests for prediction of L2 achievement or that the difference in prediction is very small. Although this result is not surprising given previous research, to our knowledge it is the first direct comparison of the two approaches. The lack of significant improvement in prediction of L2 achievement using the MLAT subtests is, at least in part, a reflection of the intercorrelations of the five subtests.

Regression commonality analysis eliminates the ambiguity of results from conventional regression analysis by comprehensively decomposing the prediction (in terms of variance accounted for) into an exhaustive but completely distinct set of predictors. For this reason, we used RCA to identify the most important predictors of each L2 achievement measure from among the five MLAT subtests (Research Question 3). Overall, the main results for L2 reading, writing, and listening comprehension (Table 6) showed that Phonetic Script was the strongest or second-strongest predictor in each case, including for L2 Oral Proficiency, followed by Paired Associates and/or shared variance between Phonetic Script and Paired Associates. The Number Learning subtest

Table 7. MLAT variance components with the largest contribution to the prediction of L2 achievement not predicted by L1 achievement measures

L2 measure–year obtained	Prediction to full distribution of L2 achievement measure		Prediction to residuals of prediction from L1 achievement measures	
	Variance component	% Contribution to prediction	Variance component	% Contribution to prediction
Reading–Yr 2	PS	33.1	PS	61.3
	SC	10.3	PS PA	14.3
	PS PA	7.1	PA	13.5
	PA	6.0		
Writing–Yr 2	PS	19.9	PA	33.6
	PA	12.1	PS	20.2
	NL PS	8.8	PS PA	12.4
	PS PA	7.4	PS NL	8.6
	NL	7.0	PS PA NL	7.7
			NL	6.7
Listening Comp–Yr 2	PS	16.9	PA	43.3
	PA	15.5	PS	12.7
	PA PS	7.9	PA PS	11.5
	NL PS	7.1	NL PA PS	6.5
	WS	6.7		
Oral Proficiency Index	PS	19.5	WS	23.7
	WS	13.9	PS	22.3
	PS WS	6.4	PS WS	8.9
	SC WS	6.3	PA	7.4
	SC	6.2		

Note. PS = Phonetic Script; PA = Paired Associates; SC = Spelling Clues; NL = Number Learning; WS = Words in Sentences.

also occurred as an important predictor in several cases for L2 reading, writing, and listening comprehension. L2 Oral Proficiency presented a different pattern, with the Words in Sentences subtest playing an important role, both individually and in shared variance with Phonetic Script and Spelling Clues. Another notable result was of the 26 possible combinations of subtests that might have relevant shared variance, only four subtests occurred in the lists of top predictors: Phonetic Script–Paired Associates was the most frequent, Number Learning–Phonetic Script was also frequent, and Phonetic Script–Words in Sentences and Spelling Clues–Words in Sentences occurred only for Oral Proficiency.

Three general conclusions emerge from these overall results. First, Phonetic Script plays a central role for all outcome measures, even the presumably oral outcomes of L2 Listening Comprehension and L2 Oral Proficiency. Second, in nearly every case (a single exception of Spelling Clues–Words in Sentences for L2 Oral Proficiency) of shared-variance pairs, the Phonetic Script subtest is present as an important, and often the strongest, predictor even for L2 Listening Comprehension and L2 Oral Proficiency. Third, although Phonetic Script was the strongest unique predictor for Year-2 L2 Listening Comprehension and Oral Proficiency, the Words in Sentences subtest, which measures grammatical sensitivity, and the Spelling Clues subtest, which measures L1 vocabulary (after the examinee has cracked the phonetic code for each word), were also important predictors of the two oral language-focused L2 measures. The predominant role played by Phonetic Script (and its closely related cousin, literacy) may be relevant for understanding the nature of the other four MLAT subtests, which will be addressed later in the Discussion.

The strongest aspect of these results, the central role of Phonetic Script, is consistent with the conclusions of the meta-analysis of the construct validity of language aptitude conducted by Li (2016). One major focus of this very comprehensive review was the evaluation of the relation of specific components of language aptitude to general and specific aspects of L2 learning. The review drew on a wider range of aptitude measures than those assessed by the MLAT. Li concluded that phonetic coding was a stronger predictor of overall proficiency than other components of aptitude (analytic ability and rote memory); that it was a stronger predictor of vocabulary than other aspects of L2 learning, even more than rote learning which is often assumed to be central to vocabulary learning; and that it was least predictive of listening comprehension. Note that given the design and research questions of almost all the published research on this topic, the analyses have not taken into account multicollinearity among predictors, and thus they are less able to identify unique predictive effects than the present study.

Table 7 reports the results of the RCA focused specifically on what the MLAT adds to prediction of L2 achievement in Year 2 beyond the L1 measures used in this study (RQ4). We are not aware of the previous research that has explicitly addressed this question other than Sparks and Dale (2023a), despite its importance for theorizing about the relationships among L1 skills, language aptitude, and L2 learning. That study used only the MLAT composite score, but its results, notably the relatively small additional prediction added by the MLAT, motivated the present investigation. The most striking result with respect to this question is the greatly increased importance of Phonetic Script both uniquely and in combination with Paired Associates, especially for L2 Reading, and for L2 Writing and L2 Oral Proficiency. Put differently, a substantial portion of the contribution of the MLAT by itself in prediction of L2 achievement (Table 6; first column of Table 7) is due to its functioning as an indirect measure of L1 ability. Consequently, that portion of prediction by MLAT variance components is diminished when prediction by L1 has been done first. In contrast, when the contribution of a MLAT component is less tied to its role as an L1 measure, its role here will be relatively higher, which is the case for Phonetic Script. Thus, literacy-related subtests may play an even larger role in the unique contribution of MLAT to prediction of L2 achievement beyond L1 measures. A similar elevation of relative contribution to prediction is shown by Paired Associates as well, especially for the prediction of L2 Listening Comprehension.

L2 Oral Proficiency is, to some extent, an exception to the pattern described above for L2 Reading, Writing, and Listening Comprehension, as it is the Words in Sentences subtest that is important in the first RCA analysis (Table 6) and it is the component that becomes even more important here (Table 7). L2 Oral Proficiency differs from L2 Reading and Writing in the smaller role that literacy plays in the construct being assessed, and it also differs from L2 Listening Comprehension in its emphasis on language production of longer units—for example, sentences and longer discourse units. The Words in Sentences subtest appears to fill another gap missed by the L1 achievement measures. Nevertheless, the result is still surprising given the strongly metalinguistic requirements of the Words in Sentences task and the presumably more implicit nature of grammatical functioning in the Oral Proficiency task. It is possible that at this early stage of L2 learning, oral proficiency in the new language is still substantially based on metalinguistic awareness inculcated through direct instruction. Research examining the relation between implicit knowledge and metalinguistic awareness of specific aspects of sentence and text structure might illuminate this possibility.

The results demonstrating the central importance of Phonetic Script in both analyses, in combination with its shared variance with other MLAT subtests, suggests

that these RCA results along with Carroll's (1962) original insights, other research literature, and our own analyses, can be used for a fresh examination of the nature of each MLAT subtest. This is the focus of the next section.

A new look at the MLAT subtests

Over 30+ years, a chain of evidence has been generated that links students' L2 aptitude on the MLAT and subsequent L2 achievement to their L1 achievement, especially L1 literacy (see Sparks, 2022a, 2022b). L1 skills, particularly in literacy, in elementary school have been found to account for over half to almost three fourths of the variance in MLAT scores in high school (Sparks et al., 2006). In a recent longitudinal study, IDs in L1 achievement alone, especially literacy-related skills, accounted for the largest portion of predicted variance in oral and written L2 achievement, whereas the MLAT made a significant but modest additional contribution (Sparks et al., 2023). Two recent longitudinal studies have led to an even stronger conclusion: prediction from MLAT to L2 achievement is due primarily to its functioning as a measure of L1 abilities, particularly L1 literacy skills, although substantial L1 variance that predicts L2 achievement was not captured by MLAT (Sparks & Dale, 2023a, 2023b). The results of the present study, which distinguish unique contributions of individual subtests from the contribution of shared variance among them, provide a basis for some new insights into the skills the subtests are assessing and which skills they may be missing as well as a more specific characterization of the mechanism of prediction. We propose here a revised view of how and why the MLAT predicts achievement in instructed language learners.

Traditionally, most aptitude researchers have focused on what the MLAT measures—that is, Carroll's (1962) four components—and have classified the five subtests according to one (or more) of these components. But, the procedure by which each of the subtests assessed the skill in question—that is, *how* an examinee had to respond when completing the items—was not considered. In essence, an implicit assumption is made that a test measures only the nominal skills and knowledge that motivated its development, despite the well-acknowledged fact that all cognitive and behavioral tests have additional requirements that may affect scores. In the strictest sense, the presence of these requirements can be viewed as a defect of the test. But it is also the case that they can add to the predictive ability of the test and, if carefully analyzed, facilitate a deeper understanding of the predictors of the target skill, in this case L2 aptitude. In particular, it is well known that the MLAT itself is heavily dependent on learners' literacy skills. Indeed, students who achieve higher scores on L1 literacy measures (word decoding, reading comprehension, writing, spelling) have been found to achieve significantly stronger scores on the MLAT.

We have examined the MLAT subtests with respect to the literacy and literacy-related skills required for their completion. This analysis, summarized in Table 8, suggests some reasons for these findings concerning the central role of literacy. On the Phonetic Script subtest, the examinee must learn a sound–symbol system that has several letter symbols and phonetic sounds that are regular to English but also introduces nine new letter symbols—for example, *iy*, *æ*, *ə*, *θ*—not used in English orthography and four English letter symbols that have phonetic sounds that are different from those in regular English —for example, *a* = /ŏ/, *aw* = /ow/, *ay* = /ī/, *ey* = /ã/. On the Spelling Clues subtest, the examinee must read and comprehend the directions and complete the sample items on their own, without feedback as to correctness. Specifically, the examinee must decode (read) an incomplete English word (*mblm*), then read five English vocabulary words and

Table 8. The role of literacy skills in the responding to MLAT subtests

Subtest	Assessment of literacy skills inherent in task	Optional strategies
Part I Number Learning	1. Distinguish root word (*tij*) from suffixes 2. Link root word to numeral meaning; can be done to the full root or the first sound/letter only 3. Link suffix to unit meaning: with no suffix is 1-4, with one suffix (*tijnu*) is 10-40, with two suffixes (*tijkunum*) is 100-400 4. Student can use the full form of the suffix or the number of syllables in the word (one syllable is 1–4, two syllables is 10–40, three syllables is 100-400) to determine the units.	Optional writing strategy: During task presentation, student can choose to *write* (*spell*) the words (*tij, tijnu, tijkunum*, etc.) spoken by the examiner to aid memory and analysis
Part II Phonetic Script	5. Student learns a new sound–symbol system [taxing both phonological and orthographic skills] 6. Student reads (decodes) words in the new language 7. Student learns nine new letters not in the English language 8. Student learns that four known English symbols have different sounds than in English (e.g., *aw* = /ow/)	
Part III Spelling Clues	9. Student must read and comprehend the directions 10. Student must read (decode) stimulus items using English letter–sound correspondences *but* with some letters omitted (e.g., *betl* = *beetle*), taxing both phonological and orthographic skills] 11. Uses phonological recoding skills 12. Student must read all vocabulary and understand five word choices for each item.	Optional meaning-based strategy: Instead of phonologically encoding stimulus and attempting to determine its meaning first, student may activate meanings of response items that can assist in identifying stimulus word
Part IV Words in Sentences	13. Student must read and comprehend all directions 14. Student must complete written sample items 15. Student must read (and understand) two sentences for each of the 45 items	
Part V Paired Associates	16. Student must read and understand the directions	Optional verbal rehearsal strategy: During the study period,

(Continued)

Table 8. (*Continued*)

Subtest	Assessment of literacy skills inherent in task	Optional strategies
	17. Student must read the 24 items, each a pairing of an English word with a Kurdish word in a 2-min practice period	student phonologically recodes both Kurdish and English words using English orthographic rules, then student rehearses overtly or silently each pairing of words
	18. Student reads the Kurdish words and *writes* the English words that match in a 2-min practice period	
	19. In the test, student reads a Kurdish word and match it with one of five English words	

choose the word most nearly in meaning to the decoded word. The size of a student's vocabulary is an important resource for this task, which has a well-established link to literacy via print exposure (see review by Stanovich, 2000; see also Sparks et al., 2012a, 2012b). For the Words in Sentences subtest, the examinee must read and comprehend the directions and complete the sample items independently, without feedback. To complete the items, the examinee must read two or more sentences for each of the items. For the Paired Associates subtest, the examinee must read and understand the directions and complete the sample items, again independently and without feedback. Specifically, the examinee must read English words (*day*) that are paired with a Kurdish word (*ja*). The task confronting the examinee is similar to that for learning new words in "real" L2s —for example, *dog* is perro in Spanish, *hund* in German, and *chien* in French, all of which are as orthographically dissimilar from *dog* as are the Kurdish words on the Paired Associates subtest. Even the Number Learning subtest, characterized as "weakly" representing the phonetic coding component (see Stansfield & Reed, 2019, p. 19), may have a stronger relationship to literacy than previously thought. For example, as described in Figure 1, an examinee could use phonetic coding, syllabic awareness, or (perhaps) morphological awareness to successfully learn the 12 new numbers in a "fake" language. The task becomes easier if the examinee discovers the "unfortunate" correspondence between the numbers and the alphabetic correspondence of their names (Carroll, 1990, p. 13).

Thus, in our view, performance on all five MLAT subtests is more or less dependent on literacy skills. What each of the subtests with shared variance has in common with the Phonetic Script subtest may be variable, but it falls within the domain of literacy skills. This dependence on literacy is both methodological and content based. The methodological aspect is due to the MLAT's use of print for both test instructions and for the completion of test items, whereas the content aspect is due to the nature of the cognitive and linguistic processes that are the ostensible focus of the subtest. Students' performance on the MLAT will generally be consistent with their L1 achievement, especially L1 literacy, but they may display intraindividual differences on the MLAT subtests, depending on their L1 literacy achievement, the extent to which a subtest relies on literacy, and their use of literacy-related strategies. Because the MLAT is timed, we would add that performance on the test is also dependent on processing speed, in this case, how fluently a student can use his/her literacy and language skills when completing the test items.

L1 literacy, L2 aptitude, and L2 achievement: What is the link?

Our goal in this study has not been to propose a specific model of L2 aptitude. Rather, it is to identify and highlight some empirical facts about L2 aptitude measures such as the MLAT, which will need to be more fully incorporated in future models, as they have been unacknowledged or underestimated in much past work. Three are discussed here. The first is the high degree of overlap between L1 measures and L2 aptitude. This is not surprising given that both types of measures predict L2 achievement. What is relatively new is how small the unique contribution of L2 aptitude is after entry of L1 measures first in a regression (Sparks et al., 2023). The second is that a primary component of what L2 adds to L1 is metalinguistic awareness and processing. The third, and a major finding of the present study, is the central role of L1 literacy and literacy-supporting skills (e.g., phonological awareness) in L2 aptitude.

With respect to the first statement above, Sparks and Dale (2023a) found that prediction from the MLAT composite score is largely due to its functioning as a measure of L1 abilities, particularly L1 literacy. Why then is the MLAT, not L1 skills, the strongest predictor of L2 achievement? Sparks and Dale surmised that the reasons are (a) MLAT includes a diverse range of basic language tasks that measure the language skills necessary for L2 learning; (b) unlike L1 tests that measure language skills encountered in everyday life—that is, contextualized material—the MLAT also measures the ability to learn from decontextualized material (see Skehan, 1986, 1998); and finally (c) the MLAT is heavily dependent on a students' literacy skills—that is, much of the test requires reading ability or invites the use of strategies such as rehearsal that rely on reading. The results of the present study provide strong support for their speculation about the MLAT and literacy skills. Moreover, the findings suggest the importance of literacy for prediction of L2 achievement after 1–2 years of language study.

With respect to the second and third statements above, Sparks (2022a) has proposed that the major link between students' L1 literacy and language achievement for contextualized material and their ability to use and understand decontextualized material on the MLAT may be metalinguistic awareness skills. A bidirectional link between the metalinguistic awareness and the onset of literacy is well established (see Bialystok, 2001; Gombert, 1992; Roehr-Brackin, 2018). Prior to literacy development, some metalinguistic awareness—for example, rhyming and alliteration—can be observed or tested in oral language (Snow et al., 1998). But, it is the development of literacy that generates further development of metalinguistic awareness. Ke et al. (2023) maintain that learning to read is "fundamentally metalinguistic because learners need to understand how the internal elements of a spoken word relate to units of graphic symbols" (p. 1). They also emphasize reciprocal, developmental relations among oral language, metalinguistic awareness, and reading competence.

Sparks and Dale (2023a) speculated that because the development of L1 literacy and metalinguistic awareness go hand in hand, the development of metalinguistic awareness and language aptitude may be similarly connected. The development of metalinguistic awareness allows a learner to reflect on the nature of language, both oral and written. This connection is not completely new. Skehan (1998, 2002) proposed that the inductive-language-learning-ability and grammatical-sensitivity components of Carroll's (1962) language aptitude model be relabeled as language analytic ability. Other researchers have proposed that metalinguistic ability and language aptitude are partially overlapping constructs (Herdina & Jessner, 2002; Jessner, 2006). Roehr-Brackin (2019) suggested that language analytic ability can be linked to the idea of

metalinguistic awareness because both reflect "the ability to focus on and manipulate language form, as well as the ability to treat language as an object of introspection, reflection, and analysis" (p. 1115).

In summary, we propose that stronger L1 literacy and language skills lead to stronger development of metalinguistic awareness (language analytic ability), which in turn leads to stronger L2 aptitude and achievement. Literacy has often been taken for granted by L2 educators and researchers because of the focus on listening and speaking, and literacy has not always been viewed as different from oral skills. However, the evidence suggests that instructed language learners with stronger literacy skills may be those learners whose language skills are so robust and so automatized that they have the freedom to think metalinguistically when engaged in L2 learning.

Situating the present study in current theorizing about aptitude

In recent years, SLA/L2 researchers, motivated by work in other areas of cognitive psychology, have advanced the proposal that implicit aptitude/knowledge is also an essential component of L2 learning and potentially even more important than explicit aptitude/knowledge as assessed (it is presumed) by the MLAT. The MLAT is said to measure cognitive abilities in the explicit domain such as traditional language aptitude and working memory (Li & DeKeyser, 2021). The argument for implicit aptitude has been advanced by, among others, Granena (2019, 2020), who described the differences in aptitudes by characterizing explicit aptitude/knowledge as "conscious, analytical, effortful, and slower [to access]," and implicit aptitude/ knowledge as "nonconscious, holistic, effortless, and faster [to access]" (2020, p. 4). Ellis (2004) has proposed that whereas explicit knowledge is conscious and accessible only through controlled processing, implicit knowledge is intuitive and available quickly and effortlessly (see also Roehr-Brackin, 2018). Researchers have suggested that implicit language aptitude can be assessed by measures such as some of the Hi-LAB subtests (ALTM, SRT; Linck et al., 2013), the LLAMA_D subtest (Meara, 2005), and other measures developed by cognitive psychologists. As Skehan (2023) and others have noted, the actual content of these measures has little or no relationship with language skills; however, this is nonetheless consistent with the theoretical views of the implicit aptitude construct, which emphasize very broad, domain-general processes. However, this theoretical dichotomy is still a matter of investigation. For example, several researchers have found that measures such as LLAMA_D are in fact part of the general LLAMA factor more aligned with explicit aptitude (see Li & Qian, 2021; Zhao et al., 2023). Although the idea of implicit aptitude may be found to be an important contribution to SLA theory, this line of research is new and not yet well developed. Even if the construct of implicit aptitude exists, little is known about how it can be measured reliably and with construct validity (see Iizuka & DeKeyser, 2023, p. 2). Conversely, the MLAT has been found to be a well-validated tool for the prediction of language learning skills. The value of the present study with the MLAT subtests and recent studies with the MLAT composite (Sparks & Dale, 2023a, 2023b; Sparks et al., 2023) is to show that RCA can suggest a different interpretation from traditional explanations of explicit L2 aptitude, one that very substantially implicates literacy processes, a diverse set of processes that would appear to include both explicit and implicit aspects. The present study does not address these issues directly, but the results provide important insights that future theories will need to address.

Limitations and implications

The strengths of this study are a large and representative sample, a prospective design, and a set of comprehensive L1 measures. At the same time, there are some limitations that may limit the generalizability of the conclusions. Although RCA is not new, the present study is the first to apply it to the relationship of L2 aptitude to L2 achievement. RCA addresses a somewhat different research question than conventional regression, so there are no similar extant results to which the present results can be compared. The MLAT was the single aptitude measure, so results concerning this measure and conclusions about aptitude generally may not be clearly distinguished. But, on grounds of face validity and actual similarity of items across aptitude tests, it is likely that measures of language aptitude will have a strong "family resemblance" along with differences among them. Ultimately these issues can only be clarified by research comparing different aptitude measures in a unified or at least comparable design.

A limitation to the data analysis is that individual MLAT item data were not available to us. Item data would have enabled estimation of the internal-consistency facet of reliability (Cronbach's alpha or IRT-based statistics). Reliability can vary substantially across studies (McKay & Plonsky, 2021), and cross-study comparison can illuminate similarities and differences in results in those studies. Further, an internal-consistency analysis of individual MLAT subtests and of the entire test would complement the RCA analyses in clarifying the relationship among subtests.

Other potential limitations to generalization of findings concern the specific context of L2 learning in this study, which is pedagogically conventional classroom-based L2 learning rather than immersion or immigration-based learning. The instruction is occurring after the development of students' L1 literacy as opposed to early simultaneous bilingualism. The study includes 2 years of L2 learning, which is typical for U.S. L2 students. Likewise, the L2 for all students is Spanish, for which there is a typologically moderately close relationship with English. The degree of orthographic similarities and differences may also be relevant given the importance of phonological awareness and word decoding as predictors of literacy outcomes. Each of these limitations constitutes a recommendation for further research and investigations with other languages, in particular those with more typological distance.

Our examination of the MLAT subtests emphasized the important of L1 literacy for L2 learning. It also highlighted for Number Learning, Paired Associates, and possibly other subtests the potential role of optional more effortful strategic choices, typically involving literacy skills even more fully across the MLAT subtests. One fruitful next step would be to replicate this study with other groups of L2 learners who are engaged in learning different languages. If L1 literacy is easier due to a transparent orthography (e.g., Italian, Finnish), it might not be such a strong predictor of L2. Danish, with its uniquely opaque orthography reflecting very different properties from English (Bleses et al., 2011), would contrast in the opposite direction. If literacy isn't phonologically based (e.g., Chinese), it might be a reduced predictor, although recent research has found that phonological awareness, a precursor for efficient word decoding, significantly accounted for unique variance in Chinese word reading and reading of arithmetic story problems with Chinese kindergarteners (Yang & McBride, 2020). Another important investigation would be to use observation and/or posttest interviews to determine which strategies individual students may use and how their choices may be related to their L1 skills and L2 achievement.

We conclude with two broader implications of these and related results. First, our findings and the reexamination of MLAT tests they inspired suggest that a reanalysis of

Carroll's (1992) four components of L2 aptitude may be appropriate. Specifically, the roles of working memory, processing speed, and language analytic abilities all merit further elaboration and investigation, including the distinction between unique and shared effects. We agree with Skehan's recent view that "the field is in urgent need of re-evaluation of existing aptitude instruments … which can take into account more recent, acquisition-oriented research. … It seems timely to engage in some degree of re-evaluation rather than [following] the same limited paths, relying on out-of-date, unvalidated, or restricted test batteries" (Skehan, 2023, p. 214). We believe that analysis of existing instruments, as in this study, can provide a basis for improved aptitude measures. Second, given the crucial role of literacy and the print exposure (reading volume) that drives literacy for metalinguistic development, L2 educators should be concerned not only by how well but also how much their students read, in both L1 and L2.

Competing interest. The authors declare none.

References

Author (2017). *Assessment of Performance Toward Proficiency in Language (AAPPL)*. Tarrytown,d NY: Language Testing Institute.

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

Bleses, D., Basboll, H., & Vach, W. (2011). Is Danish difficult to acquire? Evidence from Nordic past-tense studies. *Language and Cognitive Processes*, *26*(8), 1193–231. https://doi.org/10.1080/01690965.2010.515107

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. (1973). Implications of aptitude test research and psycholinguistic theory for foreign language learning. *International Journal of Psycholinguistics*, *2*, 5–14.

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. (2000). *Modern Language Aptitude Test (MLAT): Manual*. Second Language Testing, Inc. (Original work published 1959)

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

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.

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

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.

Gombert, J. (1992). *Metalinguistic development*. University of Chicago Press.

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

Granena, G. (2020). *Implicit language aptitude*. Cambridge University Press.

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

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, *49*, 4–14.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*, 343–365.

Iizuka, T., & DeKeyser, R. (2023). Scrutinizing LLAMA D as a measure of implicit learning aptitude. *Studies in Second Language Acquisition*. Advance online publication. doi:10.1017/S0272263122000559

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

Ke, S., Zhang, D., & Koda, K. (2023). *Metalinguistic awareness in second language reading development*. Cambridge University 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.

Kidd, E., & Donnelly, S. (2020). Individual differences in first language acquisition. *Annual Review of Linguistics*, *6*, 319–340.

Kidd, E., Donnelly, S., & Christiansen, M. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*, 154–169.

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.

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

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). UK: Routledge.

Li, S., & DeKeyser, R. (2021). Implicit language aptitude: Conceptualizing the construct, validating the measures, and examining the evidence: Introduction to the special issue. *Studies in Second Language Acquisition*, *43*, 473–497.

Li, S., & Qian, J. (2021). Exploring syntactic priming as a measure of implicit language aptitude. *Studies in Second Language Acquisition*, *43*, 574–605.

Linck, J., Hughes, M., Campbell, S., Silvert, N., Tare, M., Jackson, S., Smith, B., Bunting M., & Doughty, C. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*, 530–566.

McKay, T., & Plonsky, L. (2021). Reliability analyses: Estimating error in L2 research. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 468–482). New York: Routledge.

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

Nimon, K. (2010). Commonality analysis: Demonstration of an SPSS solution for regression analysis. *Multiple Linear Regression Viewpoints*, *36*, 10–17.

Nimon, K., Lewis, M., Kane, R., & Hanes, R. M. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods*, *40*, 457–466.

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.

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

Skehan, P. (1986). *Where does language aptitude come from?* In P. Meara (Ed.), *Spoken language* (pp. 95–113). ERIC Clearinghouse.

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

Skehan, P. (2002). Theorizing and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–94). John Benjamins.

Skehan, P. (2023). Testing language aptitude. In E. Wen, Skehan, P., & Sparks, R., *Language aptitude theory and practice* (pp. 208–239). Cambridge University Press.

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.

Sparks, R. (2022a). Where does language aptitude come from? Redux. *Language Teaching Research Quarterly*, *31*, 19–40.

Sparks, R. (2022b). *Exploring L1-L2 relationships: The impact of individual differences*. Multilingual Matters.

Sparks, R., & Dale, P. (2023a). The prediction from MLAT to L2 achievement is primarily due to underlying L1 abilities. *Studies in Second Language Acquisition*. Advance online publication. doi:10.1017/S0272263123000037

Sparks, R., & Dale, P. (2023b). Quantifying the uniqueness and efficiency of the MLAT relative to L1 attainment as a predictor of L2 achievement. [Manuscript submitted for publication].

Sparks, R., Dale, P., & Patton, J. (2023). Individual differences in L1 attainment and language aptitude predict L2 achievement in instructed language learners. *Modern Language Journal*, *107* https://doi.org/10.1111/modl.12831.

Sparks, R., & Ganschow, L. (1991). Foreign language learning difficulties: Affective or native language aptitude differences? *Modern Language Journal*, *75*, 3–16.

Sparks, R., Ganschow, L., Javorsky, J., Pohlman, J., & Patton, J. (1992). Test comparisons among students identified as high-risk, low-risk, and learning disabled students enrolled in high school foreign language courses. *Modern Language Journal*, *76*, 142–159. https://doi.org/10.2307/329769.

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

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.

Sparks, R., Patton, J., Ganschow, L., & Humbach, N. (2011). Subcomponents of second-language aptitude and second-language proficiency. *Modern Language Journal*, *95*, 1–21.

Sparks, R., Patton, J., Ganschow, L., & Humbach, N. (2012a). Relationships among L1 print exposure and early L1 literacy skills, L2 aptitude, and L2 proficiency. *Reading and Writing*, *25*, 1599–1634.

Sparks, R., Patton, J., Ganschow, L., & Humbach, N. (2012b). Do L1 reading achievement and L1 print exposure contribute to the prediction of L2 proficiency? *Language Learning*, *62*, 473–505.

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.

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.

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.

Stanovich, K. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. Guilford.

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

Wesche, M. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K.C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp.119–154). Newbury House.

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

Yang, X., & McBride, C. (2020). How do phonological processing abilities contribute to early Chinese reading and mathematics? *Educational Psychology*, *40*, 893–911.

Zhao, H., Luo, S. S., & Fu, A. (2023). Testing language aptitude for Chinese learners of foreign languages. In E. Wen, P. Skehan, & R. Sparks (Eds.), *Language aptitude theory and practice* (pp. 119–149). Cambridge University Press.