

REPLICATION RESEARCH ARTICLE

Quantifying the uniqueness and efficiency of the MLAT relative to L1 attainment as a predictor of L2 achievement: A conceptual replication

Richard L. Sparks¹  and Philip S. Dale²

¹Education Department, Mt. St. Joseph University, Cincinnati, OH, USA and ²Speech & Hearing Sciences, University of New Mexico, Albuquerque, NM, USA

Corresponding author: Richard L. Sparks; Email: Richard_Sparks@mail.msju.edu

(Received 14 June 2023; revised 26 September 2023; accepted 15 October 2023)

Abstract

In this conceptual replication of Sparks and Dale ([2023]. The prediction from MLAT to L2 achievement is largely due to MLAT assessment of underlying L1 abilities. Studies in Second Language Acquisition, 1–25) utilizing a dataset previously reported by Sparks et al. ([2009]. Long-term relationships among early L1 skills, L2 aptitude, L2 affect, and later L2 proficiency. *Applied Psycholinguistics*, 30, 725–755.), L1 achievement scores over 1st–5th grades and L2 aptitude scores from the Modern Language Aptitude Test (MLAT) in 9th grade were examined as predictors of L2 achievement for U.S. secondary students completing L2 courses in 9th and 10th grades. The study's focus was on the uniqueness and efficiency of MLAT with respect to measuring L1 achievement in predicting L2 achievement. All L1 measures and MLAT predicted L2 literacy and language, and L1 measures predicted MLAT scores. Word decoding was the strongest overall L1 predictor, though there was variation across the L2 measures. The unique contribution of MLAT was modest, as the majority of total prediction (77–86%) was due to L1 measures. The efficiency of MLAT in capturing predictive variance from L1 abilities was moderately high (median ~73%) but variable across the L1 and L2 measures. Findings are generally consistent with those of Sparks and Dale (2023) showing that prediction from MLAT to L2 is largely due to MLAT's assessment of L1 abilities, even though a substantial amount of L2 prediction-relevant L1 variance is missed by MLAT.

Keywords: L2 aptitude; Modern Language Aptitude Test; L1 achievement; mediation analysis

Introduction

A substantial and increasing body of research has documented the close relationship between individual differences (IDs) in L1 and IDs in L2 achievement (e.g., Sparks et al., 2009; Sparks, 2022a, 2022b; and of particular relevance here, Sparks & Dale, 2023; hereafter, S&D (2023)). A common finding is that although an L2 aptitude

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

measure such as the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959, 2000; Carroll, 1990) is the strongest single predictor of L2 achievement, L1 measures taken collectively are stronger still. In an overview of his research, Carroll (1981) concluded that aptitude for L2 learning consists of several basic abilities and is largely independent of verbal intelligence, but he also speculated that individuals high in overall language ability are individuals who “. . . have been able to retain and maintain native-language acquisition skills well from their early years, while low aptitude persons have for some reason lost those skills” (p. 113). Subsequent research has shown that the role of language aptitude for the prediction of L2 achievement is important not only for formal, instructed settings but also under different learning contexts and in communicative classrooms (Wen et al., 2017). The aforementioned findings provide both insight and questions concerning the skills that underlie L2 achievement and the nature (and perhaps even existence) of an L2-specific aptitude.

In a pioneering study, Skehan (1986) followed students from the Bristol Project (Wells, 1985) into L2 classes when they were 13–16 years of age. Students’ L1 development and attainment before age 5 were strongly correlated with their L2 aptitude and L2 achievement many years later in secondary school (Skehan & Ducroquet, 1988). However, although composite L2 aptitude scores were more successful predictors of students’ L2 achievement than the individual L1 predictors, the prediction of students’ L2 achievement by their L2 aptitude scores was further improved by adding L1 vocabulary and L1 language comprehension scores. Skehan (1986) 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. 7). Skehan (1989) also suggested that L2 aptitude tests gain predictiveness due to their measurement of the ability to learn from “decontextualized material.”

One theory underlying aptitude research, 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; like L1, L2 aptitude is componential, and that IDs in L1 achievement developed *prior to* L2 exposure are reflected in IDs in L2 aptitude and L2 achievement (see Sparks, 2022a). The premises of the LCDH are similar to those developed by Cummins (1979) in his Linguistic Interdependence Hypothesis (L1 and L2 have a common underlying foundation) and Linguistic Threshold Hypothesis (L2 proficiency is moderated by one’s level of attainment in L1). Research over 30 years has confirmed that L2 learners who exhibit stronger L1 achievement, especially in L1 literacy, have stronger L2 aptitude on the MLAT and stronger L2 achievement and vice versa (see reviews by Sparks, 2012, 2022b; Sparks et al., 2019). For example, in a study with high school students of Spanish, French, or German, L1 language and literacy skills in elementary school (assessed in 1st–5th grades) accounted for 58–73% of the variance in L2 aptitude on the MLAT administered in 9th grade and 30–43% of the variance in L2 achievement in 10th grade after the second-year L2 course in high school (Sparks et al., 2006). These connections were also clear in the moderate to strong correlations (.49–.68) between L1 skills in elementary school and L2 achievement in high school. Nevertheless, even in the presence of L1 achievement variables, the MLAT was found to be the best single predictor of L2 achievement (Sparks et al., 2009).

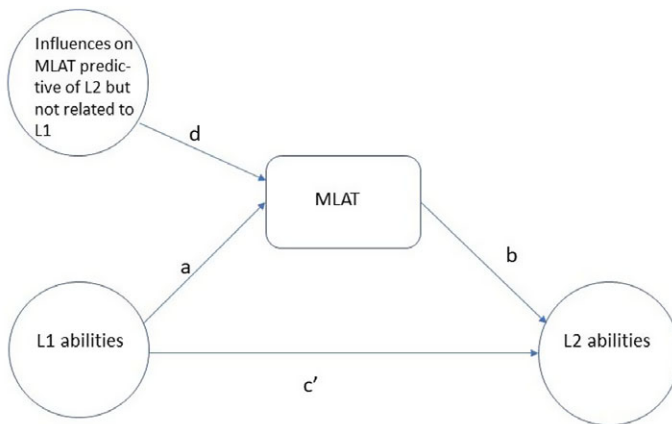
Why, then, is the MLAT the most important single predictor of L2 achievement? Like Skehan (1986), Sparks et al. (2009) have proposed that L2 aptitude tests such as the MLAT may preempt (cut out) the variance explained by L1 skills. The MLAT includes basic language tasks that measure skills necessary for language learning in both L1 and L2. But L2 aptitude tests also incorporate items that measure students' ability to process and learn from "decontextualized material" (Skehan, 1989, p. 34). For example, a student can read a word in the L1 without awareness of the letter-sound relationships or syllable types. But the MLAT Phonetic Script subtest requires the retention of new letter-sound relationships to read new words in a speeded format. Like Ranta (2002), Sparks et al. have speculated that aptitude tests may draw their predictive value from tapping into students' metalinguistic skills (see also Sparks, 2022b).

Relationship between L1 skills and MLAT for predicting L2 achievement: two indices

Based on their conjecture that the MLAT may preempt the variance explained by L1 skills for predicting L2 achievement, S&D (2023) suggested that a more comprehensive and integrated view of the relationships among L1 abilities, L2 aptitude, and L2 achievement would be facilitated by the separate assessment of *uniqueness* and *efficiency* for the prediction ability of L2 aptitude measures such as the MLAT. These comprehensive indices also have the potential to facilitate more quantitatively precise comparisons across studies. The present study is a conceptual replication of S&D (2023).

Their first index, *uniqueness*, is the extent to which the L2 aptitude measure (MLAT) adds to the prediction of L2 achievement beyond that predicted by L1 measures alone; this is measured by the ratio of variances (see Figure 1 for the conceptual model that underlies the data analysis for uniqueness and efficiency). Their second index, *efficiency*, is the extent to which the L2 aptitude measure captures the potential predictive potential of each L1 measure. Efficiency is measured as the degree of mediation by the aptitude measure of the total prediction from the L1 measures to the L2 measures. These definitions are, in principle, independent, that is, uniqueness and efficiency could both be high or low, or either one could be high or the other low. Assessing the role of L2 aptitude in this manner across diverse L2 measures can provide a more complete empirical basis for formulating hypotheses about the role of aptitude.

The mediation and regression analyses in S&D (2023) revealed that each L1 measure individually predicted all L2 scores and MLAT; the L1 measures collectively and substantially predicted MLAT scores; MLAT was a significant but only moderate mediator of prediction from L1 to L2 scores; and prediction from MLAT to L2 scores was significantly and substantially due to variance in L1 abilities captured by MLAT. Overall then, prediction from MLAT was due primarily to its functioning as a measure of L1 abilities, that is, with very limited uniqueness (12.6–40.4%) depending on the L2 outcome being predicted, although substantial L1 variance that predicts L2 achievement was not captured by the MLAT, i.e., only modest efficiency (highly variable with outcome, but median approximately 50%). The results also suggested a much more important role for L1 literacy, especially L1



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 1. Conceptual model for assessment of the relative role of L1 abilities and the MLAT in predicting L2 abilities.

word decoding, in explaining L2 aptitude and L2 achievement than has previously been acknowledged.

Purpose of study and research questions

The present paper reports a conceptual replication study of S&D (2023), drawing on the dataset analyzed by Sparks et al. (2009). Although the broad prediction design of the present study is similar to that of S&D (2023), i.e., the focus on L2 learning among U.S. high school L2 students, a similar range of L1 measures, and use of the MLAT as an aptitude measure, there are a number of important differences. First, the L1 measures were obtained much earlier in the students' development, i.e., from 1st to 5th grade rather than in 8th grade. Second, the study included students of French and German as well as Spanish, which was the only L2 studied in their previous study. Third, unlike the earlier study in which the L2 measures were standardized L2 assessments developed for native Spanish speakers, the L2 proficiency measures in the present study were educator-developed assessments based on the American Council on Teaching Foreign Languages (ACTFL) Proficiency Guidelines (1986, 1989) for reading, writing, speaking, and listening that were used in several previous studies (e.g., see Sparks et al., 2006, 2011). Fourth, and especially relevant given the central role played by L1 literacy, and especially L1 word decoding in the earlier study, the present study also included a measure of L1 phoneme awareness, an important precursor for L1 word decoding skill. These changes in L1 and L2 measures, including their timing, imply that this study constitutes a “conceptual replication” rather than a “direct replication,” following the guidelines proposed by Lebel et al. (2018).

In the present study, we addressed four research questions about the relationships among early L1 achievement, MLAT performance, and L2 achievement. The first two questions are essentially descriptive because they primarily serve to examine the correlations and regressions among the three kinds of measures for comparability to other research results: 1) How well do L1 scores predict MLAT scores and L2 achievement and 2) How well do MLAT scores predict L2 achievement? The following two questions address the core replication issues, based on the distinction between uniqueness and efficiency of MLAT prediction, utilizing multiple regression: 3) To what extent does MLAT add unique variance to the prediction of each L2 measure beyond the prediction from L1 measures alone, and 4) How much of the variance in each L1 skill that is predictive of each L2 skill is captured by the MLAT as a mediator of the correlation between the two measures?

Method

Data, analysis code, and study material information are available on the OSF website at <https://osf.io/zbef2/>

Participants

The dataset for the present study was that reported by Sparks et al. (2009). Sixty-five students attending a large, middle-class, rural public school district in the midwestern United States began the study in 9th grade. Fifty-four participants (29 females, 25 males) completed two years of Spanish ($n = 30$), French ($n = 14$), or German ($n = 10$) courses in the 9th and 10th grades. The remaining 11 participants (9 males, 2 females) did not complete the full 1st-year L2 course. The mean age of the 54 participants at the beginning of the 1st grade was 6 years, 9 months ($SD = 4.2$ months); at the end of the study ten years later, the mean age was 16 years, 5 months. Parental permission was obtained for each student.

All of the participants were monolingual, and their home background language was English. None of the participants had been exposed to a L2 or had received L2 instruction prior to 9th grade. To our knowledge, no participants were exposed to the L2 in which they were enrolled outside of the classroom. Each participant was exposed to similar learning conditions and experienced one hour per day of L2 instruction in the classroom over two years using a combination of textbook-based instruction and multiple daily opportunities to use the L2 in productive ways for listening, reading, writing, and speaking in presentational, interpretive, and interpersonal communication.

Instruments

There were several types of testing measures used in this study: L1 achievement, L2 aptitude, and L2 achievement. Each of the measures is only briefly described here, as more information is provided in Sparks et al. (2009) and in the Supplementary Online Material (Appendices S1 and S2).

L1 Achievement measures (1st through 5th grades)

Five measures of L1 achievement were administered in 1st–5th grades: (a) Woodcock Reading Mastery Test-Revised (Woodcock, 1987, 1998) Basic Skills Cluster (word decoding) and Reading Comprehension Cluster, (b) Test of Written Spelling-2 (Larsen & Hammill, 1986, 2013), (c) Peabody Picture Vocabulary Test-Revised (Dunn & Dunn, 1981, 2019), (d) Lindamood Auditory Conceptualization Test, a measure of phoneme awareness (Lindamood & Lindamood, 1979, 2004); and (e) listening comprehension assessed by the alternate form of the Woodcock Reading Mastery Test-Revised, Passage Comprehension Subtest. (See Aaron, 1991 for a complete description of the listening comprehension testing procedure).

L2 Aptitude measure (9th grade)

The MLAT-Long Form (Carroll & Sapon, 1959, 2000) was used to measure L2 aptitude. The test uses a simulated format to provide an indication of the probable degree of success in learning a L2.

L2 Achievement measures

The L2 achievement measures used in this study tested students' skills in the four areas identified by the American Council on Teaching Foreign Language (ACTFL) Proficiency Guidelines (1986, 1989) as essential for L2 acquisition: reading, writing, speaking, and oral language proficiency. They were designed by three L2 educators, who were formally trained in ACTFL guidelines, with the goal of ensuring uniformity across the three languages (i.e., Spanish, French, and German), e.g., the directions were the same for each of the three languages. The only differences in the three reading comprehension tests were those specific to a particular language.

For the present study, the L2 word decoding and the L2 reading comprehension measures were combined in a composite L2 reading measure (an average of the standardized scores for the two subtests). Likewise, the L2 writing and the L2 spelling scores were combined for a composite L2 writing score (again, an average of the standardized scores). The L2 oral proficiency score was obtained by a single test, a measure of oral language production and oral language comprehension. This process yielded three broad L2 measures: L2 reading, L2 writing, and L2 oral proficiency.

Procedure

The Woodcock Reading Mastery Test Basic Skills Cluster for L1 word decoding and reading comprehension, the Test of Written Spelling for L1 spelling, and the Peabody Picture Vocabulary Test for L1 vocabulary were administered in elementary school at five points: the beginning of 1st grade and the end of 1st, 2nd, 3rd, and 5th grades. The Lindamood test for L1 phoneme awareness was administered at the beginning of 1st grade and the end of 1st, 2nd, and 3rd grades. The Woodcock Reading Mastery Passage Comprehension subtest for L1 listening comprehension test (Aaron, 1991) was administered at the end of the 3rd and 5th grades. For each of these L1 measures, a student's scores were combined across

Table 1. Descriptive statistics for L1, L2, and MLAT measures

Measure	Minimum	Maximum	Mean	Std. Deviation
L1 Vocabulary ^a	84.0	136.2	106.7	11.2
L1 Listening Comp ^a	83.0	124.0	103.9	10.4
L1 Phon Aware ^b	-1.93	1.61	.054	.804
L1 Word Decode ^a	68.4	131.6	102.0	12.5
L1 Read Comp ^a	68.4	127.8	98.8	11.6
L1 Spelling ^a	75.2	124.2	95.8	11.6
L2 Reading ^b	-2.85	1.85	.00	.90
L2 Writing ^b	-1.50	2.14	.00	.90
L2 Oral Prof ^b	-1.42	2.86	.00	1.00
MLAT ^a	72.0	126.0	100.9	15.1

Note: $n = 54$ for all measures.

^aStandard scores ($M = 100$, $SD = 15$).

^b z -scores.

years, that is, 1st-5th, 1st-3rd, or 3rd-5th grades, to obtain a mean score on each measure. The L1 measures were administered by the first author with assistance from undergraduate and graduate students trained prior to the administration of the tests.

The MLAT was administered in small groups by the first author at the beginning of the 9th grade when the participants began the first-year L2 course. The L2 word decoding measure was administered individually to each student at the end of the 9th and 10th grades by the first author of this study and graduate students trained by him. The L2 spelling measure was administered at the end of the 10th grade by a L2 university professor who was fluent and literate in all three L2s. The L2 reading comprehension and L2 writing measures were administered in groups by the first author at the end of the students' second-year L2 course, that is, at the end of 10th grade. The L2 university professor scored the writing assessment. The L2 oral proficiency measure was administered individually by the L2 university professor at the end of the students' second-year L2 course, i.e., end of 10th grade. The oral interviews in the three L2s were taped for later scoring.

Results

Descriptive statistics for all measures are provided in Table 1. Five of the L1 measures—vocabulary, listening comprehension, word decoding, reading comprehension, and spelling—are standardized measures with $M = 100$, $SD = 15$, as is the MLAT. The students' raw scores for the Lindamood test were transformed to z -scores. Overall, the performance of these students is in the average range. The raw scores for the non-standardized L2 measures were transformed to z -scores. A complete list of correlations among the study variables is provided in Table S1 of

Table 2. L1-L2 predictive correlations (CIs)

L1 Measure	L2 Reading	L2 Writing	L2 Oral
L1 Vocabulary	.382 ** (.128–.590)	.473 ** (.235–.658)	.459 ** (.218–.647)
L1 Listening Comprehension	.426 ** (.178–.623)	.419 ** (.170–.617)	.398 ** (.145–.601)
L1 Phonological Awareness	.507 ** (.277–.682)	.591 ** (.385–.742)	.309 * (.045–.533)
L1 Word Decoding	.606 ** (.403–.752)	.565 ** (.350–.723)	.399 ** (.147–.602)
L1 Reading Comprehension	.545 ** (.325–.709)	.575 ** (.364–.731)	.470 ** (.231–.655)
L1 Spelling	.624 ** (.428–.764)	.631 ** (.437–.769)	.447 ** (.203–.638)

Note: $n = 54$. * $p < .05$; ** $p < .01$.

the Supplementary Online Material. The selected subsets relevant to the research questions are presented below.

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

Table 2 presents the predictive correlations from the L1 to L2 measures. All correlations are positive and significant. Overall, the correlations do not differ greatly, but the predictions of L2 literacy outcomes of reading and writing were strongest from L1 word decoding, L1 spelling, L1 reading comprehension, and L1 phoneme awareness. In contrast, the predictions of L2 oral proficiency were strongest from L1 spelling, L1 vocabulary, and L1 reading comprehension.

As shown in Table 3, all L1 measures significantly predict MLAT performance, with correlations ranging from .541 to .784. L1 word decoding, L1 spelling, and L1 reading comprehension are the strongest predictors of performance on the MLAT. Table 4 summarizes a multiple regression analysis utilizing all six L1 measures to predict MLAT. The multiple $R = .842$ reflects a significant contribution from L1 vocabulary, L1 spelling, and (marginally) L1 phoneme awareness.

RQ2: How well do MLAT scores predict L2 achievement?

Table 3 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 of L2 literacy skills (reading and writing) is stronger than that of L2 oral proficiency (i.e., speaking and listening comprehension).

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-L2 predictions due to the inclusion of L1 variance in MLAT. A separate regression analysis was conducted for each of the three L2 measures. Table 5 summarizes the results of these regressions. In each analysis, all L1 measures were entered first, followed by MLAT, to determine how much additional prediction of L2 achievement was gained by using MLAT. As shown in

Table 3. Correlations (CIs) of MLAT with L1 predictors and L2 outcomes

Measure	MLAT
L1 Vocabulary	.675 ** (.497–.798)
L1 Listening Comprehension	.613 ** (.413–.757)
L1 Phonological Awareness	.661 ** (.478–.789)
L1 Word Decoding	.738 ** (.585–.840)
L1 Reading Comprehension	.774 ** (.638–.863)
L1 Spelling	.784 ** (.653–.869)
L2 Reading	.683 ** (.509–.804)
L2 Writing	.723 ** (.564–.830)
L2 Oral Proficiency	.541 ** (.320–.707)

Note: $n = 54$. * $p < .05$; ** $p < .01$.

Table 4. Multiple Regression Prediction from L1 measures to MLAT

Model		Coefficients ^a				
		Unstandardized coefficients		Standardized coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.906	16.644		.054	.957
	L1 Decod	-.233	.344	-.194	-.677	.502
	L1RComp	.131	.369	.101	.356	.723
	L1Vocab	.366	.182	.272	2.007	.050
	L1Spell	.731	.307	.564	2.377	.022
	L1PhonAw	4.042	2.044	.216	1.978	.054
	L1LisCmp	.015	.202	.010	.073	.942

Note: Total prediction is $R = .842$; adjusted $R^2 = .67$.

^aDependent variable: MLATh.

Table 5, the majority of the prediction (77.0–83.6%) of the variance in L2 achievement is due to L1, with only a small proportion of the variance 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?

Mediation analysis was used to address the question of how efficiently (thoroughly) the MLAT extracts (or estimates) information about the variance in L1 scores that predict L2 scores. The question can be addressed for each combination of the six L1 and three L2 measures, for a total of 18 analyses. The analyses were conducted using

Table 5. Proportion of MLAT-L2 predictions which are due to inclusion of L1 variance in MLAT

L2 measure	Prediction from MLAT to L2		% variance for “direct” MLAT prediction <i>after</i> all L1 measures are 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 actually new, not predicted by L1 measure 1–col 5
	<i>r</i>	% variance			
Reading	.683	46.6	10.7	77.0	23.0
Writing	.723	52.3	8.6	83.6	16.4
Oral Prof.	.541	29.3	5.2	82.3	17.7

the PROCESS Macro v4.0 (Hayes, 2022; software available at <https://processmacro.org/download.html>) within SPSS. Mediation analysis is a specific application of regression analysis. The relevant output from this analysis is the proportion of the total prediction effect from the L1 measure which is mediated by MLAT (the indirect effect shown in Table 6). As illustrated in Figure 2, the results are variable across analyses, ranging from 63.4% to 100% (median ~73%). There is no trend for the MLAT to be more efficient with respect to predictions from some L1 measures more than others, in part because this efficiency measure varies substantially with the L2 outcome.

Assumption checking

To check some of the most important assumptions for multiple regression, diagnostic tests were conducted for the four most important analyses reported in Tables 4 (predicting MLAT from the L1 measures) and 5 (predicting L2 measures from the L1 measures and MLAT). They are summarized here, and the actual SPSS output is presented in the Supplementary Online Materials (Appendix S3). Durbin-Watson tests of independence of error yielded statistics in the normal range, all <2.5. Inspection of histograms and normal P-P plots for the distribution of residuals suggested that for all but L2 Writing as a dependent variable, the results were relatively normal. Plots of standardized residuals against predicted values did not show evidence for curvilinearity in any analysis, which would have indicated a non-linearity. Visual inspection of scattergrams did indicate some heteroscedasticity for L2 Writing and L2 Oral Proficiency. This does not affect the estimates (betas) for predictors, only their standard errors, which can influence significance testing. However, the present study is not focused on the comparison of effect sizes for individual L1 measures, so this finding has little relevance to our research questions. The same can be said for multicollinearity among predictors, which is shown in the Variance Inflation Factor results, particularly for L1 Decoding and L1 Reading Comprehension. Multicollinearity does affect effect size estimates, but again, the comparison of individual predictors is not part of this study. L1 measures are always entered as a group.

Table 6. Test of MLAT mediation of L1 predictions to L2 measures

L1 measure	L2 measure	Total effect	Indirect (mediated by MLAT) effect (% of total effect)	Direct effect from L1 measure
Vocabulary	Reading	.031	.042 (100%)	-.012
	Writing	.038	.040 (100%)	-.002
	Oral	.041	.026 (63.4%)	.015
Listening Comp	Reading	.037	.036 (94.7%)	.001
	Writing	.036	.040 (100%)	-.003
	Oral	.038	.028 (73.7%)	.010
Phonological Awareness	Reading	.568	.458 (80.6%)	.110
	Writing	.663	.436 (65.8%)	.226
	Oral	.385	.492 (100%)	-.107
Word Decoding	Reading	.044	.028 (63.6%)	.016
	Writing	.041	.036 (87.8%)	.005
	Oral	.032	.032 (100%)	.000
Reading Comprehension	Reading	.043	.039 (90.7%)	.003
	Writing	.045	.042 (93.3%)	.003
	Oral	.041	.030 (73.2%)	.011
Spelling	Reading	.048	.031 (64.6%)	.018
	Writing	.049	.036 (73.5%)	.013
	Oral	.038	.033 (86.8%)	.005

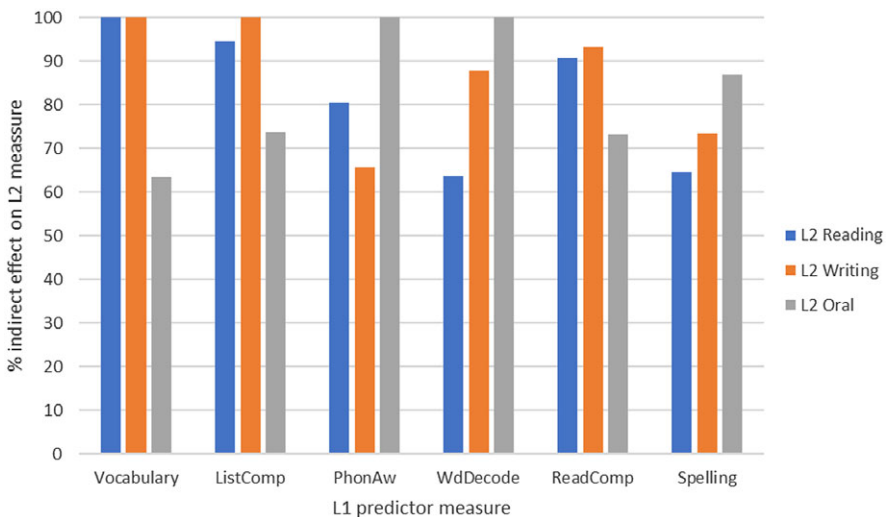


Figure 2. Degree of mediation of L1 prediction of L2 measures by MLAT.

Discussion

We asked four research questions, each of which is discussed here, including a comparison with the results of S&D (2023), before turning to more general implications.

Our first research question asked how well L1 achievement scores predict L2 achievement and MLAT scores. Table 2 shows that the predictive correlations from the L1 to L2 achievement measures are all positive and significant. The strongest L1 predictors for L2 reading and L2 writing were L1 spelling, word decoding, reading comprehension, and phoneme awareness. The finding that measures of L1 literacy are strong predictors of L2 literacy is similar to those in S&D (2023), although the correlations are somewhat stronger in the present study. The strongest predictors for L2 oral proficiency in the present study are L1 reading comprehension, vocabulary, and spelling. Overall, the findings from this 10-year replication study are quite similar to those of the three-year study of S&D (2023) and demonstrate the importance of L1 literacy for L2 achievement.

Table 3 also shows that all L1 measures significantly predicted MLAT performance with correlations ranging from .613 to .784. This pattern of results is also similar to that of S&D (2023). But in the present study, the correlations are much stronger than in S&D (2023) (.223 to .443). In both studies, L1 word decoding was the strongest predictor of MLAT performance, while correlations between MLAT and the remaining L1 variables ranged from .223 to .365. In the present study, *all* of the L1 variables were strongly correlated with MLAT scores, including L1 spelling, a skill strongly associated with word decoding. The difference in the strength of the correlations in the two studies might be explained by the time at which the L1 achievement measures were administered, i.e., 1st to 5th grades vs. 8th and 9th grades. This finding suggests that the L1 skills measured in elementary school are either a better measure of L1 achievement in general, and/or the L1 skills measured in elementary school are more closely related to the language skills measured by the MLAT. In any case, the findings highlight the importance of L1 literacy for the development of L2 aptitude.

Table 4 presents the results of a multiple regression analysis that used the L1 measures to predict MLAT scores. L1 spelling (a skill closely related to L1 word decoding), L1 vocabulary, and (marginally) L1 phoneme awareness each contributed to the prediction of MLAT scores ($R = .842$). Although the overall strength of the correlations is stronger, the findings are similar to those in S&D (2023) in which L1 word decoding and L1 vocabulary contributed significantly to the prediction of MLAT ($R = .524$). Here again, we speculate that the earlier time at which the L1 achievement measures were administered in the present study may be related to the differences in the strength of the prediction.

Our second research question asked how well MLAT scores predict students' L2 achievement. Table 3 shows that all correlations between the MLAT and L2 achievement were positive, substantial, and significant, suggesting that the MLAT is a strong predictor of L2 reading, writing, and oral proficiency. In fact, the correlations between the MLAT and L2 achievement were even stronger than those found in S&D (2023).

A new finding from the present study is that the correlations between three L1 literacy measures (word decoding, reading comprehension, and spelling) and the MLAT were stronger than those between the MLAT and the three L2 achievement measures. Similarly, the correlations between the remaining, primarily oral, L1 measures (vocabulary, listening comprehension, phoneme awareness) and the MLAT were stronger than those between the MLAT and L2 oral proficiency. As mentioned earlier, some likely explanations for the stronger predictions from L1 in the present study are the time at which the L1 achievement measures were administered, i.e., 1st–5th grades vs. 8th–9th grades, and/or the differences in the measures used to test L2 achievement in the two studies, i.e., standardized vs. researcher-developed, non-standardized measures. For example, in the present study L2 writing was measured, in part, by writing a letter rather than writing single sentences; and L2 oral proficiency was measured, in part, by having a face-to-face conversation in the L2 rather than listening to a sentence and filling in a missing word. Another potential explanation could be that the L1 measures in the present study, some of which assessed different skills than those in their previous study, e.g., spelling, phoneme awareness, and listening comprehension, are more closely aligned with the language skills measured by the MLAT.

Our third research question asked the extent to which MLAT uniquely adds to the prediction of each L2 measure beyond the prediction from the L1 measures alone. The results in Table 5 (last column) indicate that this contribution (uniqueness) is modest, i.e., $\leq 23.0\%$ for all three L2 measures. Thus, despite the stronger correlations between the L2 measures and MLAT in the present study, the results are similar to those in S&D (2023); specifically, the majority of the prediction for L2 achievement was due to the L1 measures. Why then is the unique variance accounted for the MLAT so modest even though the test has been found to be the best single predictor of L2 achievement? The most plausible answer is that there is considerable overlap between the language skills measured by the MLAT and the assessed L1 language skills, which minimizes the opportunity for MLAT to add unique variance. For example, there is an obvious overlap between phonetic coding ability measured by the L1 word decoding tests and the MLAT Phonetic Script subtest, as well as the overlap between the L1 spelling and vocabulary measures and the MLAT Spelling Clues subtest.

Our fourth research question asked how much of the variance in L1 abilities that is predictive of L2 skills is captured by MLAT mediation. In other words, how efficiently does the MLAT extract (estimate) information about L1 achievement while serving as a mediator? The relevant result from this analysis is the proportion of the total prediction effect from each specific L1 measure to each specific L2 measure that is mediated by the MLAT (indirect effect). As shown in Table 6 and Figure 2, the efficiency measure (column 4) is relatively high (median = 87%) but also quite variable. These measures are notably higher than those in S&D (2023) (median = 47%). How are such high-efficiency measures possible, that is, how can a single MLAT measure mediate the prediction from six different L1 measures? In our view, it is likely due to the substantial intercorrelations *among* the L1 measures; in effect, the MLAT is mediating the common variance among these measures.

In sum, the results of this conceptual replication study are largely consistent with the findings of S&D (2023); in that (a) all L1 measures collectively (and

substantially) predicted MLAT scores and L2 language and literacy achievement; (b) L1 word decoding was the best overall L1 predictor; and (c) prediction from MLAT to L2 achievement was significantly and substantially due to the variance in L1 skills captured by the MLAT, i.e., uniqueness was quite low. The one notable difference in the results is that efficiency, the degree of mediation of L1 prediction by MLAT, was higher in the present study, which is likely due to either higher reliability of L1 measures, or closer alignment of those measures with MLAT. Taken together, the results confirm that the prediction of L2 achievement from MLAT is due primarily, but not exclusively, to its functioning as a measure of L1 abilities although substantial L1 variance that predicts L2 achievement is not captured by MLAT.

Why does the MLAT so strongly preempt L1 skills for the prediction of L2 achievement?

The results of this conceptual replication study and those of S&D (2023) prompt the following questions: Given the strong relationship between L1 skills and L2 achievement, why has the MLAT, not L1 skills, been found to be the most important predictor of L2 achievement, and why does the MLAT appear to preempt the variance for predicting L2 achievement that might be explained by L1 skills? S&D (2023) proposed a three-part answer to these questions. First, the MLAT is comprised of basic language tasks that measure the language skills necessary for learning a L2, i.e., phonetic coding, grammar, etc. Second, unlike L1 tests that measure language skills encountered in everyday life, i.e., contextualized material, the MLAT also measures the ability to learn from decontextualized language. For example, while a student can speak a sentence in L1 without knowing parts of speech, the MLAT explicitly measures this skill (Part IV, Words in Sentences). Third, the MLAT is heavily dependent on a learner's literacy skills, i.e., much of the test requires reading ability.

Sparks and Dale proposed an additional factor for the comprehensive connections among these skills, namely, metalinguistic awareness. Several researchers have suggested that metalinguistic awareness and language aptitude (and the MLAT specifically) may be partially overlapping constructs (Herdina & Jessner, 2002; Jessner, 2006; Ellis, 2004). In L1 reading research, learning to read (literacy) has been found to be parasitic on speech and language development (Snowling & Hulme, 2012). Prior to literacy development, some metalinguistic awareness can be drawn from oral L1, e.g., rhyming (Snow, Burns, & Griffin, 1998). However, it is the development of literacy that leads to enhanced metalinguistic awareness, which further improves literacy skills. This research suggests that the global concept of L1 ability (achievement), should be “subdivided” into two components, L1 oral language and L1 literacy. If the development of metalinguistic awareness and the development of L1 literacy go hand in hand, is it possible that metalinguistic awareness and language aptitude have the same type of relationship? (See also Bialystok, 2001; Gombert, 1992; Roehr-Brackin, 2018; Yelland, Pollard, & Mercuri, 1993). A similar approach has been taken by other researchers who have been more specific in focusing on language analytic ability, which is similar to metalinguistic ability, i.e., “. . . ability to handle language in a decontextualized manner” (Roehr-Brackin, 2018, p. 87).

In our view, the necessary “connecting links” in a conceptualization of the relationships between L1 achievement and L2 achievement are 1) separation of L1 oral language from L1 literacy, 2) metalinguistic awareness as a consequence of L1 literacy development, and 3) metalinguistic awareness as a connection between L1 literacy and L2 aptitude. As noted above, some metalinguistic awareness is derived from oral L1 development, which provides the foundation for L1 literacy. The development of literacy leads to more developed metalinguistic awareness, which further enhances L1 literacy skills. L1 literacy and metalinguistic awareness are linked to stronger L2 aptitude, which in turn, both predicts and allows for stronger L2 achievement. Research also suggests that studying a L2 can, in turn, heighten language awareness, leading to improvements in L1 (Murphy et al., 2015) and higher scores on the MLAT (Sparks & Ganschow, 1993b; Sparks et al., 1998).

Our examination of the MLAT revealed the importance of L1 literacy for L2 achievement. The next step in the research agenda would be to replicate this study with other groups of L2 learners engaged in learning a more diverse range of languages and orthographies. If L1 literacy is easier due to a transparent orthography (e.g., Italian, Finnish), it might not be such a strong predictor of L2 achievement. Danish, with its uniquely opaque orthography that reflects different properties from English (Bleses et al., 2016), may present a unique challenge to the importance of L1 literacy. If literacy is not phonologically based (e.g., Chinese), it might be a reduced predictor, although research has found that phonological awareness, a precursor for efficient word decoding, significantly accounts for unique variance in Chinese word reading by Chinese kindergarteners (Yang & McBride, 2020). In another study, prereading phonological skills significantly predicted children’s reading performance in Chinese two and three years later even after controlling for the effects of age, IQ, and mother’s education (Ho & Bryant, 1997). The primary reason for this relationship is that phonological knowledge appears to assist children in using the limited phonetic component in Chinese characters. Similar to orthography, the role of L1 skills in the prediction of L2 may be influenced—and most likely heightened—by the degree of relationship between the two languages. The present study is a “close” replication of D&S (2023) involving similar populations of U.S. secondary-level learners study L2s with similar linguistic structures, namely languages from the Germanic and Romance families. Further research under more diverse circumstances, especially linguistic relationships, is needed.

A limitation of the present study design, but also characteristic of the literature, is the inability to address the relations among diverse L2 skills in the process of L2 learning. This question will require longitudinal designs with a larger number of time points and with more specific measures of L2 skills, so that, e.g., cross-lagged panel analyses can be used to assess possible causal relations. We also acknowledge the one area where the results of D&S (2023) were not strongly replicated: the higher efficiency of MLAT in the present study. We interpret this variation across studies as due to differences in measures, and more specifically, to the lack of fully parallel measures in L1 and L2. Every test reflects a distinct mixture of specific skills, and these differences will affect the correlations among tests, as well as correlations with an aptitude measure such as the MLAT. A highly desirable design feature for future prediction studies would be the use of parallel measures for the two languages.

How might the conclusion of S&D (2023) and the present study that most of the variance explained by MLAT is due to L1 ability influence how researchers and

educators approach the assessment of individual L2 learners? In their previous study, S&D (2023) suggested that it would be both prudent and cost-effective to include a standardized measure of L1 word decoding, such as the Woodcock Basic Skills Cluster used in the present investigation, as it added substantially to the prediction of L2 Reading and L2 Listening Comprehension beyond prediction by MLAT alone. Likewise, the assessment could include a standardized measure of L1 spelling, which was strongly correlated with students' L2 aptitude in the present study. Both of these measures can be administered in 5–7 minutes. Others have suggested the use of native-language course grades as predictors of L2 achievement, but caution is necessary here. In an early study, Sparks et al. (1995) found that 8th-grade L1 English grade was an important predictor of their first-year L2 grades. However, a problem with using L1 grades as predictors of L2 performance is that, at least in the U.S., there has been significant grade inflation at both the secondary (Sanchez & Moore, 2022) and postsecondary (www.gradeinflation.com) levels of education, but without corresponding increases in achievement. A similar problem exists for L2 learning assessment. In a recent study, Brown et al. (2018) found a “tenuous relationship” between students' L2 course grades and their performance on measures of L2 skills and cautioned against their use as valid measures of L2 performance.

Conclusion

In sum, we posit that the connections between L1 achievement and L2 achievement begin with oral L1 ability followed by the development of L1 literacy, which leads to the development of metalinguistic awareness. L1 literacy and metalinguistic awareness together form a foundation for L2 aptitude, which predicts L2 achievement. While the MLAT may be the best single predictor of L2 achievement because it measures both contextualized language ability and the ability to think about and handle decontextualized material, our results suggest a much more important role for L1 skills, especially literacy, in explaining learners' L2 aptitude and achievement than has previously been acknowledged.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/S0142716423000450>

Replication package. Data, analysis code, and study material information are available on the OSF website at <https://osf.io/zbef2/>.

Competing interests. The authors declare no competing interests.

References

- Aaron, P. G. (1991). Can reading disabilities be diagnosed without using intelligence tests? *Journal of Learning Disabilities*, 24, 178–186. <https://doi.org/10.1177/002221949102400306>
- American Council on the Teaching of Foreign Languages (1986). *ACTFL Proficiency Guidelines*. Yonkers, NY: ACTFL.
- American Council on the Teaching of Foreign Languages (ACTFL) (1989). *American Council on the Teaching of Foreign Languages Proficiency Guidelines*. Hastings-on-Hudson, NY: ACTFL.

- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. UK: Cambridge University Press.
- Bleses, D., Makransky, G., Dale, P., Hojen, A., & Ari, B. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, 37(6), 1461–1476. <https://doi.org/10.1017/S0142716416000060>
- Brown, A., Plonsky, L., & Teimouri, Y. (2018). The use of course grades as metrics in L2 research: A systematic review. *Foreign Language Annals*, 51(4), 763–778. <https://doi.org/10.1111/flan.12370>
- Carroll, J. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–117). Rowley, MA: Newbury House.
- Carroll, J. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. Parry & C. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 1–29). Englewood Cliffs, NJ: Prentice-Hall.
- Carroll, J., & Sapon, S. (1959, 2000). *Modern Language Aptitude Test (MLAT): Manual*. San Antonio, TX: 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(2), 222–251. <https://doi.org/10.3102/00346543049002222>
- Dunn, D., & Dunn, L. (1981, 2019). *Peabody Picture Vocabulary Test (PPVT)*. New York: Pearson.
- Ellis, R. (2004). Individual differences in second language learning. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 525–551). Malden, MA: Blackwell.
- Gombert, J. (1992). *Metalinguistic development*. Chicago, IL: University of Chicago Press.
- Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd Ed.). New York: Guilford.
- Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism: Perspectives of change in psycholinguistics* (Vol. 121). UK: Multilingual Matters.
- Ho, C. S. H., & Bryant, P. (1997). Phonological skills are important in learning to read Chinese. *Developmental Psychology*, 33(6), 946–951. <https://doi.org/10.1037/0012-1649.33.6.946>
- Jessner, U. (2006). *Linguistic awareness in multilinguals: English as a third language*. UK: Edinburgh University Press.
- LeBel, E., McCarthy, R., Earp, B., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, 1(3), 389–402. <https://doi.org/10.1177/25152459187874>
- Larsen, S., & Hammill, D. (1986, 2013). *Test of written spelling-2 (TWS)*. Austin, TX: PRO-ED.
- Lindamood, C., & Lindamood, P. (1979, 2004). *Lindamood auditory conceptualization test*. Austin, TX: PRO-ED.
- Murphy, V., Macaro, E., Alba, S., & Cipolla, C. (2015). The influence of learning a second language in primary school on developing first language literacy skills. *Applied Psycholinguistics*, 36(5), 1133–1153. <https://doi.org/10.1017/S0142716414000095>
- 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). Amsterdam: John Benjamins.
- Roehr-Brackin, K. (2018). *Metalinguistic awareness and second language acquisition*. New York: Routledge.
- Sanchez, E., & Moore, R. (2022). *Grade inflation continues to grow in the past decade (ED621326)*. Iowa City, IA: ACT.
- Skehan, P. (1986). The role of foreign language aptitude in a model of school learning. *Language Testing*, 3(2), 188–221. <https://doi.org/10.1177/026553228600300207>
- Skehan, P. (1989). *Individual differences in second language learning*. London: Arnold.
- 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*. Washington, DC: National Academy Press.
- Snowling, M., & Hulme, C. (2012). Interventions for children's language and literacy difficulties. *International Journal of Language & Communication Disorders*, 47(1), 27–34. <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(1), 187–214. <https://doi.org/10.1007/BF02648218>

- Sparks, R.** (2012). Individual Differences in L2 Learning and Long-Term L1-L2 Relationships. *Language Learning*, *62*(Supp. 2), 5–27. <https://doi.org/10.1111/j.1467-9922.2012.00704.x>
- Sparks, R.** (2022a). *Exploring L1-L2 relationships: The impact of individual differences*. UK: Multilingual Matters.
- Sparks, R.** (2022b). Where does language aptitude come from? Redux. *Language Teaching Research Quarterly*, *31*, 19–40.
- 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., & Dale, P.** (2023). The prediction from MLAT to L2 achievement is primarily due to underlying L1 abilities. *Studies in Second Language Acquisition*, 1–25. <https://doi.org/10.1017/S0272263123000037>.
- 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. L., Ganschow, L., & Patton, J.** (1995). Prediction of performance in first-year foreign language courses: Connections between native and foreign language learning. *Journal of Educational Psychology*, *87*(4), 638–655. <https://doi.org/10.1037/0022-0663.87.4.638>
- 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*(2), 253–273. <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., & Luebbbers, 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*(2), 255–283. <https://doi.org/10.1111/flan.12390>
- Wells, G.** (1985). *Language development in the pre-school years (Vol. 2)*. Cambridge, UK: Cambridge University Press.
- Wen, Z., Biedroń, A., & Skehan, P.** (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, *50*(1), 1–31. <https://doi.org/10.1017/S0261444816000276>
- Woodcock, R.** (1987, 1998). *Woodcock Reading Mastery Test-R (WRMT-R)*. Circle Pines, MN: AGS.
- Yang, X., & McBride, C.** (2020). How do phonological processing abilities contribute to early Chinese reading and mathematics? *Educational Psychology*, *40*(7), 893–911. <https://doi.org/10.1080/01443410.2020.1771679>
- Yelland, G., Pollard, J., & Mercuri, A.** (1993). The metalinguistic benefits of limited contact with a second language. *Applied Psycholinguistics*, *14*(4), 423–444. <https://doi.org/10.1017/S0142716400010687>

Cite this article: Sparks, R. L. & Dale, P. S. (2023). Quantifying the uniqueness and efficiency of the MLAT relative to L1 attainment as a predictor of L2 achievement: A conceptual replication. *Applied Psycholinguistics* *44*, 1143–1160. <https://doi.org/10.1017/S0142716423000450>