REPLICATION STUDY

Robust evidence for the simple view of second language reading: Secondary meta-analysis of Jeon and Yamashita (2022)

Akira Hamada[1] , Haruka Shimizu[2], Yuko Hoshino[3], Shuichi Takaki[4] and Yuji Ushiro[5]

[1]Kobe City University of Foreign Studies; [2]Tohoku Gakuin University; [3]Chiba University; [4]Fukushima University; [5]University of Tsukuba
Corresponding author: Akira Hamada; Email: hamada.akira@inst.kobe-cufs.ac.jp

Abstract

This paper reports a complete secondary analysis of Jeon and Yamashita's (2022) systematic review to build the second language (L2) model of the simple view of reading (SVR). The same meta-analytic methodologies were maintained, with the exception of applying meta-analytic structural equation modeling (MASEM). This study successfully replicated some of the aggregated correlations but not others, owing to (a) the recoding of the original raw data to recreate a dataset and (b) the motivated change in sample selection from a longitudinal study for MASEM. The MASEM results extended previous findings that L2 comprehension skills contribute more to L2 reading comprehension than L2 decoding skills, and together explain a large amount of variance in L2 reading comprehension. The SVR model with metalinguistic skills showed their contribution to L2 decoding and comprehension skills, but no direct impact on L2 reading comprehension, supporting the parsimonious structure of SVR in L2.

Introduction

Replicating meta-analysis

Reproducibility of meta-analysis is important (Plonsky, 2012; van IJzendoorn, 1994; Weissgerber et al., 2021), given that meta-analytically derived results have a massive impact on the relevant research domain and practice. Examining reproducibility is at the core of replication research, which assesses the extent to which study findings can be reproduced in replication attempts (Marsden et al., 2018). Because meta-analysis involves various methodological choices in the process (Borenstein et al., 2021), different analysts may produce different results even from the same raw data (van IJzendoorn, 1994). Therefore, Weissgerber et al. (2021) proposed the importance of methods' reproducibility to improve the quality of meta-analysis. They claimed that if

the description of meta-analytical approaches is sufficient, an independent researcher can reproduce the analysis results within an acceptable margin of error. The accessibility of the raw data, analysis code, and statistical output also contributes significantly to enhancing the reproducibility of meta-analysis.

Given the importance of meta-analysis reproducibility, the first goal of this study was to reproduce the findings of Jeon and Yamashita (2022; hereafter, "the 2022 study"), an updated version of "L2 Reading Comprehension and its Correlates: A Meta-Analysis" published in *Language Learning* (Jeon & Yamashita, 2014; hereafter, "the 2014 study"), which systematically summarized bivariate correlations between L2 reading comprehension and its underlying components. The 2014 study has significantly contributed to L2 reading research by being cited more than 500 times at the time of this writing, in accordance with the importance of understanding L2 proficiency (Jeon & In'nami, 2022). Moreover, both studies argued the relative importance of reading components for L2 reading comprehension, which serves as a valuable resource for L2 teachers to determine the degree to which each component should be emphasized in their reading instruction (Hamada, 2020).

Another motivation for replicating meta-analysis lies in its methodological progress that enables the exploration of questions left unresolved in previous meta-analyses (Plonsky, 2012). This aligns with the stringent definition of replication, or a series of modified repetitions of an original study (Porte & McManus, 2019). The 2014 and 2022 studies concluded that their unresolved issue was the establishment of a comprehensive model of L2 reading comprehension. Therefore, the second goal of this study was to extend the findings of the 2022 study to the modeling of the Simple View of Reading (SVR) in L2.

SVR is a leading theory that the process of learning to read entails mapping written words to spoken language; consequently, reading comprehension is determined by the interaction between word decoding and oral language comprehension skills (Hoover & Gough, 1990). Word decoding is the cognitive process of deciphering written words into their corresponding spoken language forms, requiring knowledge of letter-sound correspondence. Oral language comprehension represents linguistic knowledge and processes used to understand spoken language, such as vocabulary knowledge and listening comprehension. Moreover, word decoding skills explain more variance in reading comprehension in the early stages of learning to read, while the contribution of oral language comprehension to reading comprehension increases as the word decoding processes become automatized (see Sparks, 2021 for a more comprehensive review). The potential of this theory to characterize strengths and weaknesses in L2 readers has also been discussed both theoretically (Sparks, 2021) and empirically (Verhoeven & van Leeuwe, 2012). While the traditional meta-analysis, as employed in the 2014 and 2022 studies, is prevalent in the field of L2 research, it only provides aggregated results of bivariate relationships (e.g., L2 reading comprehension and vocabulary, grammar, and metacognitive knowledge) and small-scale interactions among them (Raeisi-Vanani et al., 2022). This limitation impedes the investigation of questions concerning the extent to which interactions among these components predict variance in L2 reading comprehension, aligning with the SVR framework.

Identifying the structural relationships among correlated components of L2 proficiency using meta-analytic structural equation modeling (MASEM) has received increasing attention in L2 research (In'nami, Cheung, et al., 2022). The MASEM approach combines meta-analysis and structural equation modeling methodologies (e.g., Cheung, 2015b), allowing for formulation of a hypothesized psychological model with multiple observed and latent variables using a synthesized/pooled correlation
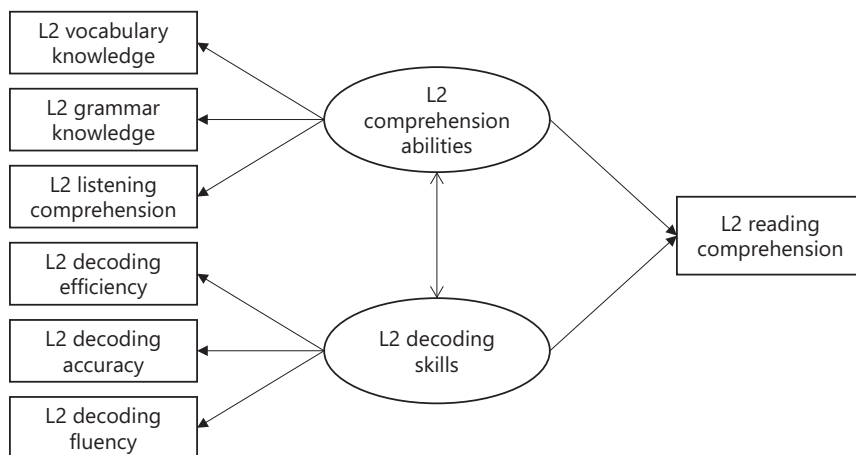
Figure 1. The SVR model in L2 proposed by Lee et al. (2022).

matrix. For example, Lee et al. (2022) applied the SVR model to L2 reading in their MASEM using 81 independent samples from 67 primary studies (n = 10,526). As depicted in Figure 1, rather than examining a direct relationship between L2 reading comprehension and each component, their model revealed an association between the two latent variables of L2 decoding and comprehension skills with reading comprehension. Collectively, these factors accounted for over 60% of the variance in L2 reading comprehension, with L2 comprehension abilities contributing more significantly. In comparison to Lee et al. (2022), the 2022 study synthesized a broader array of variables from a larger pool of samples. Consequently, the application of MASEM has potential to extend their findings by providing a comprehensive and robust structure between the correlated components of L2 reading comprehension. To achieve this, we will review the 2022 study and describe our replication approach.

The original study

The purpose of both the 2014 and 2022 studies was to examine the strength of the correlation between L2 reading comprehension and 10 components recognized as significant predictors of L2 reading variance. These components include L2 decoding, L2 phonological awareness, L2 orthographic knowledge, L2 morphological knowledge, L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension, L2 listening comprehension, working memory, and metacognition. Guided by the theoretical framework of the multi-component view of reading, these studies sought to investigate the relative magnitudes of the correlations, aiming to identify which components exhibit stronger associations with L2 reading comprehension.

As a meta-analytic procedure, the 2014 and 2022 studies systematically searched for primary studies that met specific criteria: (a) reporting both sample size and the original correlation coefficient between passage-level L2 reading comprehension and one or more of the identified components, (b) including only participants without language-related disabilities, and (c) being published in English. Consequently, the 2014 study identified 67 independent samples from 58 studies published between 1979 and 2011. The 2022 study expanded this database by adding research published between 2011 and

2017, comprising a total of 107 independent samples from 88 studies. Study characteristics were coded according to the following scheme: (a) if studies appeared to have reported on duplicate samples, only one study for each was included in the meta-analysis; (b) for longitudinal studies, only data collected at one time point were included by selecting the one most comparable with the rest of the studies in the analysis pool (e.g., "if there were 31 samples in the pool and 30 of them involved adolescent/adult participants, and one last sample was from a longitudinal study following a group of sixth graders for four years, we used the data taken in the tenth grade," Jeon & Yamashita, 2014, p. 177); and (c) if studies reported multiple measures for a construct that could be defined as unitary, the average value of the reported correlations was used.

Correlations, weighted for sample size and corrected for attenuation, were synthesized using the random-effects model with Comprehensive Meta-Analysis Version 2 (Borenstein et al., 2021). As a result, the 2022 study concluded that the relationship between L2 linguistic knowledge, such as vocabulary and grammar, and L2 reading comprehension was robust. L2 processing skills, such as L2 decoding, phonological, orthographic, and morphological skills, also showed moderate-to-strong correlations, suggesting their significance as predictors of L2 reading comprehension. Notably, L2 listening comprehension emerged as the strongest factor in achieving higher levels of L2 reading comprehension, aligning with the SVR model for L2 (Lee et al., 2022). In addition, the modest correlations of language-general variables, such as working memory and metacognition, suggest the relative importance of language processing skills based on sufficient linguistic knowledge in L2 reading comprehension (Nassaji, 2014).

The scope and approaches of replication

As noted, the scope of this replication is twofold: (a) to ascertain whether the meta-analytic findings of the 2022 study can be reproduced with respect to the strength of correlations between L2 reading comprehension and the 10 components and (b) to explore the extension of their findings to the L2 SVR model using the MASEM approach. Methodologically, this replication is a complete secondary analysis, according to a process model of different types of replications (van IJzendoorn, 1994). Regardless of the type of research, this approach involves the recoding and reanalyzing of the original data to recreate a dataset. Illustrating a complete secondary meta-analysis, Weissgerber et al. (2021) examined outcome reproducibility, entailing the re-creation of the dataset used in the original study, to inspect whether and to what extent the reanalyzed results vary when applying the same analytical approach (i.e., computational reproducibility and verification) or different analytical choices (i.e., analysis reproducibility). As the 2014 and 2022 studies presented a comprehensive list of publicly available literature and a detailed coding scheme with clearly described analytical methods, we can confirm the reproducibility of their outcomes.

As a modified repetition of the original study, we also employ the MASEM approach instead of the correlation-based analysis adopted in the 2022 study. Although the 2014 and 2022 studies estimated individual effects only, MASEM allows the investigation of more than just one effect simultaneously (In'nami, Cheung, et al., 2022), such as the simple effects of L2 vocabulary and grammar knowledge, along with their interaction effect on L2 reading comprehension. Another advantage is that MASEM can manage latent variables (e.g., L2 decoding) underlying several observed variables (e.g., word-
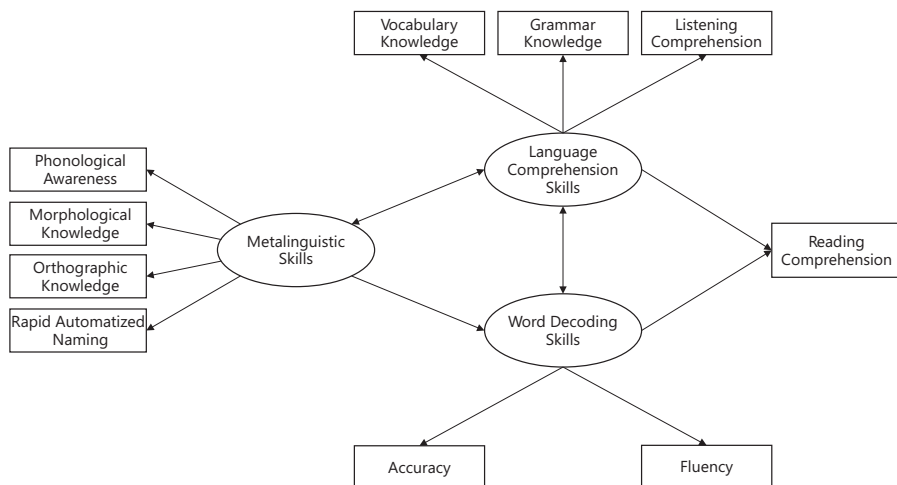
Figure 2. The SVR model in L1 proposed by Peng et al. (2021).

reading accuracy and fluency). Considering that the SEM approach allows for extending the findings of original studies without SEM (Hamada & Takaki, 2021), it is essential to investigate the robustness of specified models with meta-analyzed data. This minimal yet theoretically motivated change facilitates an extension of the 2022 study as well as a fair comparison of its findings with this replication (Porte & McManus, 2019).

The re-creation of the 2022 study dataset has the advantage of extending the SVR model developed by Lee et al. (2022). As shown in Figure 1, while their SVR model consists of six components (i.e., L2 vocabulary and grammar knowledge, listening comprehension, and L2 decoding accuracy, fluency, and efficiency), the 10 components commonly synthesized in the 2014 and 2022 studies can be applied to Peng et al.'s (2021) SVR model for L1 readers. As shown in Figure 2, this model incorporates the three components of vocabulary and grammar knowledge and listening comprehension for a latent variable of language comprehension skills; two components of decoding accuracy and fluency for a latent variable of word decoding skills; and four components of phonological awareness, morphological knowledge, orthographic knowledge, and rapid automatized naming for a latent variable of metalinguistic skills. Peng et al. added metalinguistic skills (defined as the ability to think about language with awareness by manipulating structural features of language, such as phonemes, words, and sentences; Tunmer et al., 1988) to the SVR model because they are robust predictors of developing word decoding and language comprehension skills. Their MASEM results showed that decoding and language comprehension together explained 53% of the variance in reading comprehension. In addition, although metalinguistic skills significantly contributed to decoding and showed a strong relationship with language comprehension, they did not directly contribute to reading comprehension. As previous studies have found that cognitive skills in L2 comprehension are also associated with the development of L2 metalinguistic skills (e.g., Siu & Ho, 2015; Wang et al., 2006), this study examined the SVR model established by Peng et al. for L2 readers.

Table 1. Methodological differences between the 2022 study and the present replication

Methodology	The 2022 study	The present replication
Study coding - Components - Treating data dependency	- 10 correlates - Multiple correlations for a pair of variables were averaged. - When a study repeated the same assessments, the data included changed based on compatibility with the other pooled data.	- 10 correlates - Multiple correlations for a pair of variables were averaged. - When a study repeated the same assessments, only the first data were recorded.**
Analytic procedure - Software - Model - Moderators - Analysis	- Comprehensive Meta-Analysis Ver.2 - Random-effects model - Age, language setting, L2 proficiency, L1-L2 language distance, L2-L2 script distance - Bivariate meta-analysis	- R package {meta} Ver.6.5-0** - Random-effects model - Not applicable due to insufficient samples** - Multivariate MASEM*

Note. Asterisks indicate motivated (*) and unmotivated (**) changes that potentially affected the replicated results. The methods used in the 2022 study were identical to those used in the 2014 study.

Method

Design of the current meta-analysis

Table 1 summarizes the differences in meta-analytic choices between the 2022 and present studies. As described in the Study Coding section, the change in the handling of data dependency from the 2022 study to perform MASEM was practically motivated, not theoretically (i.e., unmotivated change; see Marsden et al., 2018). In the Meta-Analytic Procedure section, we explain why we motivated the change in the methodology for this secondary meta-analysis. The use of different software for statistical computing would not result in a difference between the two studies (Schwarzer et al., 2015).

Study coding

The identical 88 studies used in the 2022 study were included in the present meta-analysis (see Supplementary Material 1). Two independent raters coded the study characteristics, such as study names, sample sizes, and correlations between L2 reading comprehension and each of the 10 components. Appendix A of the 2022 study provides detailed information on acceptable measures of both L2 reading comprehension and its components, which were referenced to determine which correlation coefficients reported in the primary studies should be recorded.

To satisfy the meta-analytic assumption of no more than one effect size per construct per study, studies by the same (co)authors were examined for data duplication (see Supplementary Material 2). In cases where multiple assessments were conducted for a single construct, they were averaged to control for data dependency. Similarly, if a longitudinal study reported multiple correlations from the same sample, only one set of correlations was included in the analysis. Although the 2022 study changed the data that were included based on comparability with other pooled studies per construct, we only used the first correlation to combine all variables into one pooled correlation for the MASEM approach.

To use the correlation matrices in the MASEM, correlations between the 10 components were also recorded. In addition, the measures of L2 word-decoding skills were further divided into decoding accuracy and fluency (tasks that measure the ability to read words aloud accurately in an untimed or timed format, respectively) to construct MASEM models in the same manner as Peng et al. (2021). As the intercoder agreement was sufficiently high (90%), any disagreements were resolved through discussion.

Meta-analytic procedure

Meta-analytic calculations were conducted following the procedures reported in the 2022 study. First, information regarding the sample sizes and correlations was entered into the {meta} package of R (Schwarzer et al., 2015). Correlations attenuated by measurement error were corrected using the relevant reliability indices reported in the primary studies. In cases where statistical information was not provided, the mean reliability for each measure was employed for calculations. Then weighted average correlations between L2 reading comprehension and the 10 components were computed along with their 95% confidence intervals using the random-effects model. Based on the Open Science Collaboration (2015), the criteria for replication success were set as a combination of significance and small telescopes. Traditionally, replication was considered successful if it produced (non)significant *p*-values, as in the original study. However, Simonsohn (2015) proposed small telescopes that combined null hypothesis significance testing with effect size estimation to examine whether the replicated effect size was significantly smaller than that of the original study. Therefore, in the present study, the success of replicating the previous findings was determined by considering that the average correlations between L2 reading comprehension and each of the 10 components did not differ significantly, with small standardized mean differences between the two studies (i.e., Cohen's *d* < 0.40; see Plonsky & Oswald, 2014).

For the MASEM, we used correlation matrices collected from primary studies to analyze Peng et al.'s (2021) SVR model. We followed a two-stage MASEM approach using the {metaSEM} package (Cheung, 2014, 2015a); 78 correlation matrices from 68 studies were combined to construct a pooled correlation matrix in Stage 1, and the pooled matrix was used to test the theoretical model in Stage 2. The primary studies did not include all the correlations between L2 reading comprehension and the components related to these models, and thus maximum likelihood estimation was used for such missing correlations. Among the four components of metalinguistic skills, L2 orthographic knowledge and rapid automatized naming were removed because of a lack of sufficient samples. The theoretical model was assessed using goodness-of-fit indices (see Kline, 2023, for review), including the comparative fit index (CFI ≥ .95), the Tucker-Lewis index (TLI ≥ .95), the root mean square error of approximation (RMSEA ≤ .06), and the standardized root mean square residuals (SRMR ≤ .05).

We built three theoretical models (see Figure 3) to determine whether (a) metalinguistic skills could be included in the SVR model and (b) metalinguistic skills made a direct contribution to L2 reading comprehension. Based on Peng et al. (2021), we examined whether in Model A, L2 metalinguistic skills directly impacted L2 decoding skills and were correlated with L2 comprehension skills. Model B integrated a direct effect of L2 metalinguistic skills on L2 reading into Model A. In Model C, any contribution from L2 metalinguistic skills was eliminated by fixing the relevant parameters to zero. Considering that the three models were nested, a likelihood ratio test was performed to determine the best-fitting model. We chose a complex model if it was significantly more accurate than a nested model. Otherwise, a nested model was
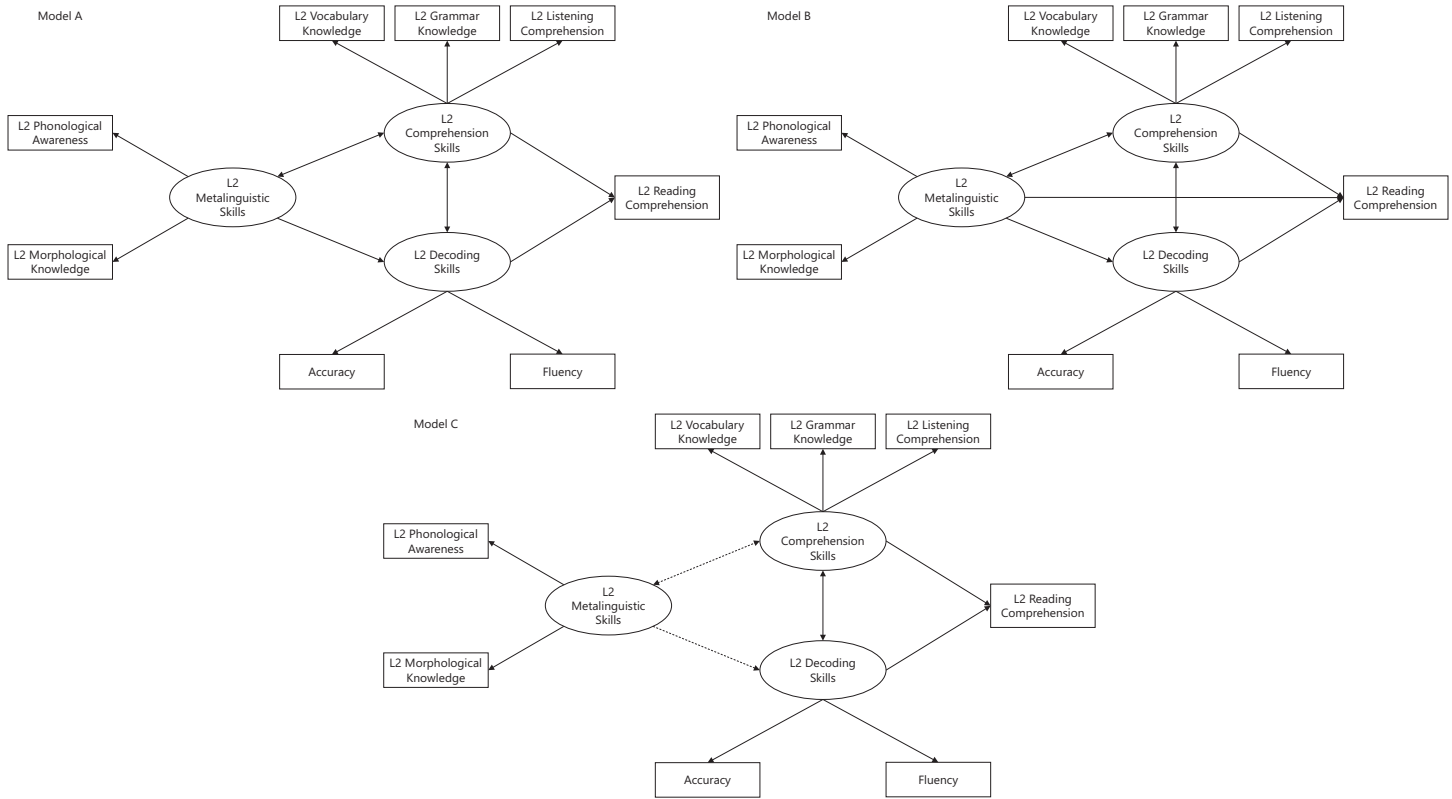
Figure 3. Three hypothesized SVR models for the current MASEM approach.
Note. Model A = SVR without direct effects of metalinguistic skills; Model B = SVR with direct effects of metalinguistic skills; Model C = SVR without the components of metalinguistic skills.

selected for model parsimony (Peng et al., 2021). Supplementary Material 3 contains the raw data, analysis codes, and statistical output (see also https://osf.io/jksp8/) in the manner proposed by In'nami, Mizumoto, et al. (2022).

Finally, we performed exploratory publication bias testing and confirmed that there was no unignorable influence of the compiled data on the present results. Supplementary Material 4 provides the details of the publication bias analyses and results.

Results and discussion

Replication of the 2022 study

Table 2 compares the number of samples used and the aggregated correlations between the 2022 and the present studies. Supplementary Material 5 provides forest plots for each component. Successfully reproduced effect sizes were observed for decoding, orthographic knowledge, morphological knowledge, grammar knowledge, listening comprehension, working memory, and metacognition, with no significant differences between the two studies. We first considered cases where there were no significant differences in effect sizes and sample selection (i.e., L2 orthographic knowledge, L2 listening comprehension, working memory, and metacognition). For L2 orthographic knowledge and metacognition, we combined the two groups sampled by Abu-Rabia and Sanitsky (2010). An additional sample was included for L2 orthographic knowledge (Noonan et al., 1997) and working memory (Shiotsu, 2010) because they each administered a spelling and reading span test. Although not significant, the effect of L2 listening comprehension was reduced in terms of the standardized mean difference. We added one sample that used a standardized L2 listening test (Woodcock Language

Table 2. Aggregated correlations between L2 reading comprehension and each of the components in the 2022 and present studies

L2 reading components	The 2022 study		The present study		Welch's t	df	p	d
	k	r [95% CI]	k	r [95% CI]				
L2 decoding	29	.59 [.45, .69]	38	.58 [.51, .64]	0.29	58.62	.774	0.07
L2 phonological awareness*	20	.61 [.52, .69]	19	.49 [.39, .59]	2.24	37.00	.031	0.72
L2 orthographic knowledge	6	.59 [.34, .76]	6	.57 [.33, .81]	0.38	10.00	.712	0.22
L2 morphological knowledge	14	.64 [.56, .70]	10	.53 [.41, .65]	1.39	18.34	.182	0.58
L2 vocabulary knowledge*	51	.72 [.64, .79]	59	.57 [.53, .62]	3.32	95.30	.001	0.64
L2 grammar knowledge	26	.70 [.52, .82]	33	.60 [.51, .68]	1.03	48.00	.310	0.28
L1 reading comprehension*	34	.48 [.36, .59]	38	.39 [.32, .46]	2.10	67.86	.039	0.50
L2 listening comprehension	20	.81 [.64, .91]	20	.62 [.53, .71]	1.74	33.95	.091	0.55
Working memory	19	.33 [.23, .43]	20	.31 [.20, .42]	0.96	36.99	.341	0.31
Metacognition	11	.33 [.08, .54]	10	.31 [.10, .53]	0.01	18.98	.991	0.00

Note. The original data were adopted from Table 2 of the 2022 study. Correlation coefficients were corrected for attenuation. For the components marked with an asterisk, there were significant differences in effect size between the two studies.

Proficiency Battery: August et al., 2006) and removed the sample in Crosson and Lesaux (2010) due to sample overlap with Lesaux et al. (2010).

Second, although no significant difference in the aggregated effect size for L2 decoding was observed, the number of samples included in the meta-analysis differed considerably between the 2022 and the present studies. We examined the types of L2 decoding tasks used for the independent samples that were not included in the 2022 study. Of the nine added samples, six used standardized tests, such as the Letter-Word Identification subset of Woodcock Language Proficiency Battery (Chen et al., 2012; Francis et al., 2006), Woodcock Reading Mastery Test-Revised (Jared et al., 2011; Jia et al., 2014; Xue & Jiang, 2017), Woodcock-Munuz Language Survey-Revised (Leider et al., 2013), Wide Range Achievement Test-Revised (Wang et al., 2006), and Würzburg Silent Reading Test (Limbird et al., 2014) for decoding accuracy, and the Test of Word Reading Efficiency for decoding fluency (Jared et al., 2011; Jia et al., 2014; Jiang et al., 2012; Lesaux et al., 2010). The other samples used tailor-made tests that measured the ability to decode L2 graphemes into their corresponding phonemes (Koda, 1998; Li et al., 2012; Siu & Ho, 2015). For these reasons, including these samples was reasonable and improved the robustness of the findings regarding the effects of L2 decoding skills on reading comprehension.

Similarly, we sampled seven additional correlations for L2 grammar knowledge, wherein Swanson et al. (2011) used a standardized test (the Morphological Closure subtest of the Illinois Test of Psycholinguistic Ability III), and the other samples employed tailor-made tests such as multiple-choice sentence completion (Brisbois, 1995; Larson, 1983; Shiotsu, 2010; Shiotsu & Weir, 2007) and grammaticality judgment (Mecartty, 2000). However, the difference in the effect size for L2 grammar knowledge were small and did not reach significance. In contrast, four samples were removed from the analysis of L2 morphological knowledge due to data duplication (Chen et al., 2012; Lam et al., 2012). They showed higher correlations with L2 reading comprehension than this study (average $r = .62$), resulting in a lower correlation coefficient.

The effect sizes that were not successfully replicated were for L2 phonological awareness, L2 vocabulary knowledge, and L1 reading comprehension. First, two samples were added to L2 phonological awareness (Francis et al., 2006; Kieffer & Vukovic, 2013), in which the Comprehensive Test of Phonological Processing was used as a standardized measure. Meanwhile, three samples were removed for the same reason as L2 morphological knowledge. While the added samples yielded smaller correlation coefficients than in the 2022 study (average $r = .50$), the removed samples exhibited larger coefficients (average $r = .71$), reducing the overall impact of L2 phonological awareness. Second, regarding L2 vocabulary knowledge, we sampled eight more correlations than in the 2022 study. Among them, six samples used the Peabody Picture Vocabulary Test (Grant et al., 2012; Jia et al., 2014; Li et al., 2012) and the Wechsler Abbreviated Scale of Intelligence vocabulary subtest (Kim, 2012). Four samples employed tailor-made tests (e.g., cloze and meaning-recall formats) to measure L2 receptive vocabulary knowledge (Brisbois, 1995; Sang et al., 1986; Shiotsu, 2010). In cases with multiple samples in one study, the other three samples were not included in the 2022 study (Edele & Stanat, 2016: either Russian or Turkish sample; Shiotsu & Weir, 2007). These samples produced lower correlations (average $r = .61$) than those synthesized in the 2022 study, reducing the strength of the correlation between L2 vocabulary knowledge and reading comprehension in this replication. Finally, the effect size for L1 reading comprehension was also significantly smaller than that in the 2022 study. We included four additional samples that used standardized L1 reading comprehension tests: the Woodcock-Muñoz Language Survey-Revised (Goodwin et al.,

2015; Swanson et al., 2011), the Gray Oral Reading Test (Jared et al., 2011), and the Neilson Reading Test (adapted for Chinese L1 reading; Pasquarella et al., 2011). As these samples exhibited lower correlations (average $r = .14$) compared to the 2022 study, the correlation aggregated in this study was accordingly reduced.

Possible reasons for the failure to reproduce the previous results are (a) the recoding of the original raw data to recreate a dataset and (b) the motivated change in sample selection from a study that repeated the same assessment. Consistently, additional samples for each variable showed a lower correlation with L2 reading comprehension in the current meta-analysis than in the 2022 study. Some of these samples included a significantly higher number of participants than the median sample size ($Mdn = 91$) of pooled studies (e.g., $n = 12,252$ for L2 vocabulary knowledge: Sang et al., 1986; $n = 471$ for L1 reading comprehension: Swanson et al., 2011). This suggests that these large-scale studies affected the pooled effect sizes, as they were weighted by sample size (Schwarzer et al., 2015). The synthesis of effect sizes may also have been affected by the motivated change in sample selection. However, given that a longitudinal design was used across the 10 components, the addition of the studies reporting small effect sizes may have led to failure to replicate the findings of the 2022 study.

Secondary analysis of the 2022 study

By using the restructured dataset of the 2022 study, pooled correlations among the variables were estimated (78 samples from 68 studies with 12,062 participants). Table 3 shows that the pooled correlation coefficients ranged from. 30 to. 63, with all correlations being statistically significant ($p < .05$), aligning closely with the results of Lee et al. (2022; $r = .21$–.62).

Based on the pooled correlation matrix, we built Models A, B, and C (see Figure 3), in which L2 reading comprehension was the dependent variable and the three latent variables were the independent variables. The likelihood ratio test showed no significant difference between Models A and B, $\chi^2(1) = 2.51, p = .113$, suggesting that Model A was more parsimonious than Model B regarding the structural complexity of SVR. The SVR model without metalinguistic skills (Model C) was not supported because the other two models were significantly more accurate, $\chi^2(2) = 1163.79, p < .001$ (vs. Model A), $\chi^2(3) = 1166.30, p < .001$ (vs. Model B). Therefore, based on Model A, we will discuss the relative contributions of L2 decoding and comprehension skills to L2 reading comprehension.

As depicted in Figure 4, the three latent variables were composed of significant and strong factor-loading coefficients (standardized path coefficient βs >. 60). The observed data fit the model with sufficient goodness-of-fit: $\chi^2(16) = 22.15, p = .138$, CFI =

Table 3. A pooled correlation matrix and the number of samples for each correlation

Variables	1	2	3	4	5	6	7	8
1. L2 reading comprehension	—	30	19	60	33	21	20	11
2. L2 decoding accuracy	.56	—	9	23	6	8	16	9
3. L2 decoding fluency	.46	.63	—	12	6	4	5	3
4. L2 vocabulary knowledge	.51	.49	.36	—	29	16	12	9
5. L2 grammar knowledge	.52	.56	.35	.53	—	12	3	3
6. L2 listening comprehension	.56	.42	.30	.52	.48	—	3	3
7. L2 phonological awareness	.46	.53	.42	.43	.39	.37	—	5
8. L2 morphological knowledge	.49	.56	.34	.53	.58	.52	.46	—

Note. Values above the diagonal indicate the number of samples for each pooled correlation.
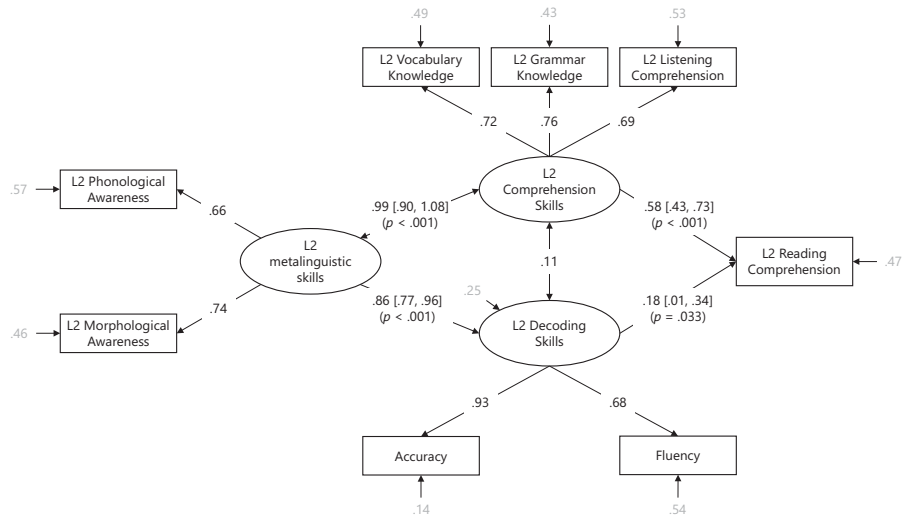
Figure 4. MASEM model A for the SVR in L2 with standardized estimates.
Note. Values in gray and brackets indicate the error variables and 95% CIs of each standardized estimate, respectively.

.998, TLI = .997, RMSEA = .006 (95% CI [.000,. 011]), SRMR = .043. Both L2 decoding (β = .18, 95% CI [.01,. 34], *p* = .033) and comprehension skills (β = .58, 95% CI [.43,. 73], *p* < .001) significantly contributed to L2 reading comprehension. There was a significant difference in their relative importance in L2 reading comprehension as the 95% CIs of the two path coefficients did not overlap. This result suggests that L2 comprehension skills contribute more to L2 reading comprehension than L2 decoding skills. Consistent with Peng et al. (2021), metalinguistic skills made a direct contribution to L2 decoding skills (β = .86, 95% CI [.77,. 96], *p* < .001) and were strongly correlated with L2 comprehension skills (β = .99, 95% CI [.90, 1.08], *p* < .001). In the SVR model with metalinguistic skills, 53% of the variance in L2 reading comprehension was explained.

Regarding the primary finding of the MASEM approach, a larger amount of variance in L2 reading comprehension could be attributed to both L2 decoding and comprehension skills in the SVR model with the indirect effects of metalinguistic skills compared to the model without metalinguistic skills. This coefficient of determination was comparable to that found in previous MASEM studies on L1 (53%; Peng et al., 2021) and L2 (61%; Lee et al., 2022) reading comprehension. As suggested by the 2014 and 2022 studies, the present MASEM results reinforce a more substantial contribution of L2 comprehension skills to L2 reading comprehension.

Limitations and future directions

The present replication includes a few limitations. First, it is necessary to investigate the complicated causes of the failure to replicate certain results of the 2022 study. The inherent flexibility in dataset creation can pose challenges in replicating meta-analytic results (Norouzian, 2021). Moreover, both the present and 2022 studies may oversimplify the intricate outcomes by either discarding or averaging correlations from primary studies with longitudinal designs or multiple measurements (see also Norouzian & Bui,

2023). While ideal solutions may involve leveraging modern approaches to meta-analysis and MASEM, such capabilities are not currently available. A promising avenue for improvement is to strive for the transparency of meta-analytic approaches and open-science practices in generating and reporting results, facilitating secondary analysis (Marsden & Plonsky, 2018). Although In'nami, Mizumoto, et al. (2022) advocate for computationally reproducible research that involves sharing supplementary information such as raw data, analysis code, and output files for reanalysis and evaluation of statistical results, it is essential to consider the other three aspects of reproducibility proposed by Weissgerber et al. (2021) for the best reporting practice in meta-analyses.

In addition, for the MASEM results, including a larger number of samples is desirable to establish a stable and precise SVR model for L2. The use of the 2022 study dataset not only supports the applicability of SVR to L2 readers as in Lee et al. (2022), but also expands our understanding of the structural relationships between cognitive and metacognitive skills in L2 reading proficiency. Unfortunately, due to the limited availability of samples, the current MASEM model could not incorporate the effects of moderator variables (e.g., age, language setting, L1-L2 language distance, L1-L2 script distance, and L2 proficiency). While transcribing correlation coefficients from primary studies, it was observed that in some cases, only the correlation between L2 reading comprehension and target variables was reported, omitting correlations among the variables. Future studies should present a correlation matrix between L2 reading comprehension and its components, even if some components are not the primary focus theoretically or practically. Specifically, as shown in Table 3, the number of correlations between L2 decoding fluency and listening comprehension was relatively limited, despite being central to the SVR model. It is also necessary to explore correlations between L2 listening comprehension and L2 phonological awareness and morphological knowledge when applying Peng et al.'s (2021) SVR model. Moreover, more research on L2 orthographic knowledge is urgently required to examine the moderating effects of L1-L2 script distance on L2 reading comprehension.

Lastly, the SVR model examined in this study does not represent the whole picture of L2 reading. The latest meta-analysis by Lee and Lee (2023) has proposed an extended simple view of L2 reading that includes intelligence and working memory as cognitive abilities in addition to the Peng et al.'s (2021) model. Consistent with the findings of the present study, they have also shown the parsimoniousness of the SVR model by confirming that cognitive variables do not uniquely contribute to L2 reading comprehension. However, given the complex factors that influence L2 reading difficulties, future studies need to be systematic in establishing a comprehensive model of L2 reading that can explain reading problems not identified by the SVR model alone (Sparks, 2021). For example, the component model of reading (Joshi & Aaron, 2000) incorporates affective and instructional factors that have long been studied in the context of L2 reading (e.g., Hamada & Takaki, 2021). Testing the applicability of such reading models to L2 contexts will contribute to the science of the multi-component view of reading.

Implications

Despite these limitations, this study represents the first attempt to replicate the meta-analytic results of L2 research. Our replication makes unique contributions to L2 research in the following ways:

1. Highlighting the importance of secondary analyses for evaluating research reproducibility, which aligns with the open science movement and methodological reforms in L2 research (Marsden & Plonsky, 2018).
2. Calling for replicating meta-analyses in L2 research because different meta-analysts may propose different procedures to extend and generalize the findings of original meta-analyses.
3. Extending the insights of the SVR model accumulated by individual studies with a large dataset, which will help develop a comprehensive model of L2 reading comprehension, affording a better understanding of the complicated relationships among reading components.

This study also has several theoretical and practical implications for L2 reading. From a theoretical perspective, our findings are consistent with the argument that the SVR model can be applied to L2 as an influential reading model (e.g., Lee et al., 2022; Lee and Lee, 2023; Sparks, 2021; Verhoeven & van Leeuwe, 2012). They also support the idea that L2 reading proficiency can be characterized in a way that L2 comprehension skills are stronger predictors than L2 decoding skills. More importantly, L2 metalinguistic skills have an indirect effect on L2 comprehension through L2 decoding and comprehension skills that function as possible mediators. These findings suggest that although the SVR model may be as parsimonious as possible for L2, L2 metalinguistic skills play a significant role in enhancing L2 decoding and comprehension skills for L2 reading comprehension.

For L2 reading instruction, the SVR model is helpful for obtaining information about struggling learners (Sparks, 2021). In the current SVR model (see Figure 4), L2 vocabulary and grammar knowledge, along with L2 listening comprehension, equally explained a large amount of the variance in L2 comprehension skills. These components are crucial for distinguishing successful and struggling L2 readers and should be the focal point of L2 reading instruction. When compared to improving word-reading fluency, improving L2 word-reading accuracy is expected to lead to being a stronger L2 reader (see also Lee et al., 2022). As L2 metalinguistic skills underlie L2 decoding skills and are associated with L2 comprehension skills, phonological and morphological (and orthographic) knowledge can be taught explicitly and learned implicitly through L2 comprehension activities for efficient word reading (Nassaji, 2014).

Conclusions

We replicated Jeon and Yamashita's (2022) meta-analytic findings regarding the bivariate relationships between L2 reading comprehension and its components. The previous results that were successfully reproduced in terms of correlational strength included L2 decoding, L2 morphological, orthographic, and grammar knowledge, L2 listening comprehension, L1 reading comprehension, working memory, and metacognition. Through a complete secondary analysis, this replication improved the statistical robustness concerning the significance of these correlates in L2 reading comprehension. The MASEM revealed that metalinguistic skills were associated with L2 decoding and comprehension skills but did not directly contribute to L2 reading comprehension, supporting the parsimonious structure of SVR, even in the context of L2 reading. The finding that L2 comprehension skills were stronger predictors of L2 reading comprehension than L2 decoding skills aligns with systematic (Jeon & Yamashita, 2014, 2022)

and narrative (Sparks, 2021) inquiries, as well as previous MASEM findings (Lee et al., 2022, Lee and Lee, 2023).

References

Abu-Rabia, S., & Sanitsky, E. (2010). Advantages of bilinguals over monolinguals in learning a third language. *Bilingual Research Journal*, *33*, 173–199. https://doi.org/10.1080/15235882.2010.502797

August, D., Francis, D. J., Hsu, H. Y. A., & Snow, C. E. (2006). Assessing reading comprehension in bilinguals. *The Elementary School Journal*, *107*, 221–238. https://doi.org/10.1086/510656

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (2nd ed.). John Wiley & Sons.

Brisbois, J. E. (1995). Connections between first- and second-language reading. *Journal of Reading Behavior*, *27*, 565–584. https://doi.org/10.1080/10862969509547899

Chen, X., Ramirez, G., Luo, Y. C., Geva, E., & Ku, Y.-M. (2012). Comparing vocabulary development in Spanish- and Chinese-speaking ELLs: The effects of metalinguistic and sociocultural factors. *Reading and Writing*, *25*, 1991–2020. https://doi.org/10.1007/s11145-011-9318-7

Cheung, M. W.-L. (2014). Fixed- and random-effects meta-analytic structural equation modeling: Examples and analyses in R. *Behavior Research Methods*, *46*, 29–40. https://doi.org/10.3758/s13428-013-0361-y

Cheung, M. W.-L. (2015a). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, *5*, 1521–1528. https://doi.org/10.3389/fpsyg.2014.01521

Cheung, M. W.-L. (2015b). *Meta-analysis: A structural equation modeling approach*. Springer.

Crosson, A. C., & Lesaux, N. K. (2010). Revisiting assumptions about the relationship of fluent reading to comprehension: Spanish-speakers' text-reading fluency in English. *Reading and Writing*, *23*, 475–494. https://doi.org/10.1007/s11145-009-9168-8

Edele, A., & Stanat, P. (2016). The role of first-language listening comprehension in second-language reading comprehension. *Journal of Educational Psychology*, *108*, 163–180. https://doi.org/10.1037/edu0000060

Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading*, *10*, 301–322. https://doi.org/10.1207/s1532799xssr1003_6

Goodwin, A. P., August, D., & Calderon, M. (2015). Reading in multiple orthographies: Differences and similarities in reading in Spanish and English for English learners. *Language Learning*, *65*, 596–630. https://doi.org/10.1111/lang.12127

Grant, A., Gottardo, A., & Geva, E. (2012). Measures of reading comprehension: Do they measure different skills for children learning English as a second language? *Reading and Writing*, *25*, 1899–1928. https://doi.org/10.1007/s11145-012-9370-y

Hamada, A. (2020). Using meta-analysis and propensity score methods to assess treatment effects toward evidence-based practice in extensive reading. *Frontiers in Psychology*, *11*, 617. https://doi.org/10.3389/fpsyg.2020.00617

Hamada, A., & Takaki, S. (2021). Approximate replication of Matsuda and Gobel (2004) for psychometric validation of Foreign Language Reading Anxiety Scale. *Language Teaching*, *54*, 535–551. https://doi.org/10.1017/S0261444819000296

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*, 127–160. https://doi.org/10.1007/BF00401799

In'nami, Y., Cheung, M. W. L., Koizumi, R., & Wallace, M. P. (2022). Examining second language listening and metacognitive awareness: A meta-analytic structural equation modeling approach. *Language Learning*, *73*, 759–798. https://doi.org/10.1111/lang.12548

In'nami, Y., Mizumoto, A., Plonsky, L., & Koizumi, R. (2022). Promoting computationally reproducible research in applied linguistics: Recommended practices and considerations. *Research Methods in Applied Linguistics*, *1*, 100030. https://doi.org/10.1016/j.rmal.2022.100030

Jared, D., Cormier, P., Levy, B. A., & Wade-Woolley, L. (2011). Early predictors of biliteracy development in children in French immersion: A 4-year longitudinal study. *Journal of Educational Psychology*, *103*, 119–139. https://doi.org/10.1037/a0021284

Jeon, E. H., & In'nami, Y. (2022). *Understanding L2 proficiency: Theoretical and meta-analytic investigations.* John Benjamins. https://doi.org/10.1075/bpa.13

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*, 160–212. https://doi.org/10.1111/lang.12034

Jeon, E. H., & Yamashita, J. (2022). L2 reading comprehension and its correlates: An updated meta-analysis. In Y. In'nami & E. H. Jeon (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 29–86). John Benjamins. https://doi.org/10.1075/bpa.13.03jeo

Jia, F., Gottardo, A., Koh, P. W., Chen, X., & Pasquarella, A. (2014). The role of acculturation in reading a second language: Its relation to English literacy skills in immigrant Chinese adolescents. *Reading Research Quarterly*, *49*, 251–261. https://doi.org/10.1002/rrq.69

Jiang, X., Sawaki, Y., & Sabatini, J. (2012). Word reading efficiency, text reading fluency, and reading comprehension among Chinese learners of English. *Reading Psychology*, *33*, 323–349. https://doi.org/10.1080/02702711.2010.526051

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology*, *21*, 85–97. https://doi.org/10.1080/02702710050084428

Kieffer, M. J., & Vukovic, R. K. (2013). Growth in reading-related skills of language minority learners and their classmates: More evidence for early identification and intervention. *Reading and Writing*, *26*, 1159–1194. https://doi.org/10.1007/s11145-012-9410-7

Kim, Y.-S. (2012). The relations among L1 (Spanish) literacy skills, L2 (English) language, L2 text reading fluency, and L2 reading comprehension for Spanish-speaking ELL first grade students. *Learning and Individual Differences*, *22*, 690–700. https://doi.org/10.1016/j.lindif.2012.06.009

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Press.

Koda, K. (1998). The role of phonemic awareness in second language reading. *Second Language Research*, *14*, 194–215. https://doi.org/10.1191/026765898676398460

Lam, K., Chen, X., Geva, E., Luo, Y. C., & Li, H. (2012). The role of morphological awareness in reading achievement among young Chinese-speaking English language learners: A longitudinal study. *Reading and Writing*, *25*, 1847–1872. https://doi.org/10.1007/s11145-011-9329-4

Larson, J. W. (1983). Skills correlations: A study of three final examinations. *The Modern Language Journal*, *67*, 228–234. https://doi.org/10.2307/327079

Lee, H., Jung, G., & Lee, J. H. (2022). Simple view of second language reading: A meta-analytic structural equation modeling approach. *Scientific Studies of Reading*, *26*, 585–603. https://doi.org/10.1080/10888438.2022.2087526

Lee, H., & Lee, J. H. (2023). Extending the simple view of reading in second and foreign language learning: A meta-analytic structural equation modeling approach. *Review of Educational Research.* Advance online publication. https://doi.org/10.3102/00346543231186605

Leider, C. M., Proctor, C. P., Silverman, R. D., & Harring, J. R. (2013). Examining the role of vocabulary depth, cross-linguistic transfer, and types of reading measures on the reading comprehension of Latino bilinguals in elementary school. *Reading and Writing*, *26*, 1459–1485. https://doi.org/10.1007/s11145-013-9427-6

Lesaux, N. K., Crosson, A. C., Kieffer, M. J., & Pierce, M. (2010). Uneven profiles: Language minority learners' word reading, vocabulary, and reading comprehension skills. *Journal of Applied Developmental Psychology*, *31*, 475–483. https://doi.org/10.1016/j.appdev.2010.09.004

Li, T., McBride-Chang, C., Wong, A., & Shu, H. (2012). Longitudinal predictors of spelling and reading comprehension in Chinese as an L1 and English as an L2 in Hong Kong Chinese children. *Journal of Educational Psychology*, *104*, 286–301. https://doi.org/10.1037/a0026445

Limbird, C. K., Maluch, J. T., Rjosk, C., Stanat, P., & Merkens, H. (2014). Differential growth patterns in emerging reading skills of Turkish-German bilingual and German monolingual primary school students. *Reading and Writing*, *27*, 945–968. https://doi.org/10.1007/s11145-013-9477-9

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning*, 68, 321–391. https://doi.org/10.1111/lang.12286

Marsden, E., & Plonsky, L. (2018). Data, open science, and methodological reform in second language acquisition research. In A. Gudmestad & A. Edmonds (Eds.), *Critical reflections on data in second language acquisition* (pp. 219–228). John Benjamins.

Mecartty, F. H. (2000). Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish. *Applied Language Learning*, 11, 323–348.

Nassaji, H. (2014). The role and importance of lower-level processes in second language reading. *Language Teaching*, 47, 1–37. https://doi.org/10.1017/S0261444813000396

Noonan, B., Colleaux, J., & Yackulic, R. A. (1997). Two approaches to beginning reading in early French immersion. *Canadian Modern Language Review*, 53, 729–742. https://doi.org/10.3138/cmlr.53.4.729

Norouzian, R. (2021). Interrater reliability in second language meta-analyses: The case of categorical moderators. *Studies in Second Language Acquisition*, 43, 896–915. https://doi.org/10.1017/S0272263121000061

Norouzian, R., & Bui, G. (2023). Meta-analysis of second language research with complex research designs. *Studies in Second Language Acquisition*. Advance online publication. https://doi.org/10.1017/S0272263123000311

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716. https://doi.org/10.1126/science.aac4716

Pasquarella, A., Chen, X., Lam, K., Luo, Y. C., & Ramirez, G. (2011). Cross-language transfer of morphological awareness in Chinese–English bilinguals. *Journal of Research in Reading*, 34, 23–42. https://doi.org/10.1111/j.1467-9817.2010.01484.x

Peng, P., Lee, K., Luo, J., Li, S., Joshi, R. M., & Tao, S. (2021). Simple view of reading in Chinese: A one-stage meta-analytic structural equation modeling. *Review of Educational Research*, 91, 3–33. https://doi.org/10.3102/0034654320964198

Plonsky, L. (2012). Replication, meta-analysis, and generalizability. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 116–132). Cambridge University Press.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878–912. https://doi.org/10.1111/lang.12079

Porte, G., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.

Raeisi-Vanani, A., Plonsky, L., Wang, W., Lee, K., & Peng, P. (2022). Applying meta-analytic structural equation modeling to second language research: An introduction. *Research Methods in Applied Linguistics*, 1, 100018. https://doi.org/10.1016/j.rmal.2022.100018

Sang, F., Schmitz, B., Vollmer, H. J., Baumert, J., & Roeder, P. M. (1986). Models of second language competence: A structural equation approach. *Language Testing*, 3, 54–79. https://doi.org/10.1177/026553228600300103

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R*. Springer. https://doi.org/10.1007/978-3-319-21416-0

Shiotsu, T. (2010). *Components of L2 reading: Linguistic and processing factors in the reading test performances of Japanese EFL learners*. Cambridge University Press.

Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing*, 24, 99–128. https://doi.org/10.1177/0265532207071513

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, 26, 559–569. https://doi.org/10.1177/0956797614567341

Siu, C. T.-S., & Ho, C. S.-H. (2015). Cross-language transfer of syntactic skills and reading comprehension among young Cantonese-English bilingual students. *Reading Research Quarterly*, 50, 313–336. https://doi.org/10.1002/rrq.101

Sparks, R. L. (2021). Identification and characteristics of strong, average, and weak foreign language readers: The Simple View of Reading model. *The Modern Language Journal*, 105, 507–525. https://doi.org/10.1111/modl.12711

Swanson, H. L., Orosco, M. J., Lussier, C. M., Gerber, M. M., & Guzman-Orth, D. A. (2011). The influence of working memory and phonological processing on English language learner children's bilingual reading

and language acquisition. *Journal of Educational Psychology*, *103*, 838–856. https://doi.org/10.1037/a0024578

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

van IJzendoorn, M. H. (1994). A process model of replication studies: On the relation between different types of replication. In R. van der Veer, M. H. van IJzendoorn, & J. Valsiner (Eds.), *Reconstructing the mind: Replicability in research on human development* (pp. 57–70). Ablex Publishing.

Verhoeven, L., & van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and Writing*, *25*, 1805–1818. https://doi.org/10.1007/s11145-011-9346-3

Wang, M., Cheng, C., & Chen, S.-W. (2006). Contribution of morphological awareness to Chinese-English biliteracy acquisition. *Journal of Educational Psychology*, *98*, 542–553. https://doi.org/10.1037/0022-0663.98.3.542

Weissgerber, S. C., Brunmair, M., & Rummer, R. (2021). Null and void? Errors in meta-analysis on perceptual disfluency and recommendations to improve meta-analytical reproducibility. *Educational Psychology Review*, *33*, 1221–1247. https://doi.org/10.1007/s10648-020-09579-1

Xue, J., & Jiang, X. (2017). The developmental relationship between bilingual morphological awareness and reading for Chinese EFL adult learners: A longitudinal study. *Reading and Writing*, *30*, 417–438. https://doi.org/10.1007/s11145-016-9683-3