Cross-linguistic Sharing of Morphological Awareness in Biliteracy Development: A Systematic Review and Meta-analysis of Correlation Coefficients

Sihui Ke, University of Kentucky, sihui.ke@uky.edu

Ryan Miller, Kent State University, rmill129@kent.edu

Dongbo Zhang, University of Exeter, d.zhang4@exeter.ac.uk

Keiko Koda, Carnegie Mellon University, kkoda@andrew.cmu.edu

ABSTRACT

This study examined the cross-linguistic sharing of morphological awareness (MA) in biliteracy development. The analysis included 34 correlational studies with 40 independent samples (N = 4,056). Correlational coefficients were meta-analyzed, yielding four main findings: (1) the correlation between first language (L1) and second language (L2) MA was small ($r = 0.30$). (2) The interlingual correlations between L1 MA and L2 word decoding and between L1 MA and L2 reading comprehension were both small (respective $r = 0.35, 0.39$). (3) The intralingual correlations between L2 MA and L2 word decoding and between L2 MA and L2 reading comprehension were both moderate ($r = 0.45, 0.52$). (4) MA measurement type and age were significant moderators. Our review suggested that there is a need for future research to align the definition and measurement of MA as a multifaceted construct, and pay equal attention to its contributions to both word decoding and reading comprehension.

Keywords: morphological awareness, cross-linguistic sharing, biliteracy/second language reading, meta-analysis

Authors' note

We are grateful for the feedback and comments provided by Prof. Emma Marsden and the anonymous reviewers. Special thanks go to Maria Carlo, Alexis Filippini and other authors of the primary studies who have responded to our request for data sharing for this meta-analysis.

Correspondence should be addressed to: Sihui Ke (Ph.D.), University of Kentucky, sihui.ke@uky.edu. Mailing address: 1055 Patterson Tower, Department of Modern & Classical Languages, Literature & Cultures, University of Kentucky, Lexington, KY 40506.

Introduction

Morphological awareness (MA) is generally understood as the ability to reflect upon, analyze, and manipulate morphemes and morphological structure of printed words[1] (Carlisle, 2000; Koda, 2000; Kuo & Anderson, 2006). It is a multi-dimensional, multi-faceted construct that entails diverse aspects and levels of insights into morphologically complex words. MA can range from (a) an ability to recognize that printed words can be segmented into smaller word parts, to (b) an understanding of rules that govern morpheme concatenations and functional constraints on those rules, and to (c) an ability to apply structural understandings for functional purposes such as word learning (Koda & Miller, 2018; Tyler & Nagy, 1989). While MA is sometimes used interchangeably with "morphological knowledge" or "morphological processing," Nagy, Carlisle, and Goodwin (2014, p.4) identified "morphological knowledge" as an umbrella term for "morphological awareness (explicit/strategic analysis and use of morphology)" and "morphological processing (implicit/tacit understanding and use of morphology)." In this research, following Nagy et al. (2014), we focus on MA, distinguishing it from morphological processing.

There is increasing correlation-based evidence that MA, in addition to predicting monolingual reading development across seventeen languages (Verhoeven and Perfetti, 2017), is also a critical predictor of second language (L2) / bilingual reading development (e.g., Deacon, Wade-Woolley, & Kirby, 2007; Tong & McBride-Chang, 2010; Wang, Ko, & Choi, 2009). In addition, some experimental evidence suggests that L2 readers, like monolingual readers, benefit from explicit morphological instruction (Goodwin & Ahn, 2010, 2013; Kirby & Bowers, 2017). Emerging evidence also seems to suggest that morphological instruction provided in learners' first/dominant language can enhance reading subskills development in their other language (e.g.,

D. Zhang, 2016). However, to date, our understanding of how MA functions as a cross-linguistically shareable resource in biliteracy development is still unclear. To what extent does MA correlate between readers' L1 and L2? Are the relationships similar between MA and different L2 reading skills such as word decoding and reading comprehension? Are these relationships subject to the influence of learner-internal or learner-external factors (e.g., word formation rules, L1-L2 writing system type, age, or MA and reading outcome measurement type)?

To answer these questions, this meta-analysis examined MA as a complex construct in child biliteracy development. It focuses on child readers because they differ qualitatively from adult L2 readers in term of literacy and linguistic profiles. Specifically, adult L2 readers are typically skilled readers in their L1 and thus tend to show little individual difference in L1 MA. In contrast, child L2 readers are developing readers in both L1 and L2 and hence constitute a unique case for studying cross-linguistic sharing of MA based on correlational associations, an important goal of the present research. This narrower focus without including adult learners also reflects the fact that there has been little research on MA in adult biliteracy or that has aimed to study MA and literacy development concurrently in L1 and L2 in adult populations.

The literature search yielded 34 studies and 40 independent samples for meta-analysis. Given the relatively small sample pool, we combine the meta-analysis with a critical review. The goals of this research were three-fold: first, we aimed to provide a scoping review of the theoretical conceptualizations and methodological designs of previous research in this area. Second, we set out to synthesize existing evidence by (1) examining to what extent L1 MA correlates with L2 MA, (2) investigating the intralingual and interlingual correlations between MA and L2 word decoding, and between MA and L2 reading comprehension, and (3) exploring linguistic-, learner- and task-related moderating effects on biliteracy development. Third, we aimed to, through the scoping review and the meta-analysis, expand the current understanding of *the extent to* which and *how* MA functions as a cross-linguistically shareable resource in child biliteracy development, and provide implications and future directions for research that examines this complex construct in (bi)literacy development.

Theoretical Background

Cross-linguistic Transfer of MA

MA pertains to a learner's sensitivity to the internal structure of printed words. Morphologically complex words are prevalent, particularly in written texts (Nagy & Anderson, 1984). MA is thus arguably an important underpinning of reading and its development. In L2 reading development, which involves two languages (and sometimes two writing systems as well), there are complex interactions among L1 MA, evolving L2 MA, L2 linguistic knowledge (e.g., vocabulary), and L2 reading skills that develop subsequently. Notably, the transfer facilitation model (Koda, 2005, 2007, 2008) contends that metalinguistic awareness, including MA, can be transferred from readers' L1 to facilitate their L2 metalinguistic awareness and reading development. In this research, cross-linguistic transfer is defined as "automatic activation of previously established L1 mapping patterns triggered by linguistic input in a later acquired language (i.e., L2)" (Koda, Lü, & Zhang, 2014, pp.145-146).

Different frameworks have been proposed that have informed the understanding of *transfer* in L2/bilingual research in general and L2 reading research in particular. They include, for example, (1) the contrastive and typological framework (Lado, 1957), which views transfer as interference on L2 stemming from L1 structural properties; (2) the linguistic interdependence hypothesis, which distinguishes between cognitively and conceptually demanding and less demanding knowledge and predicts the conditions under which the learner can display transfer of knowledge cross-linguistically (Cummins,1979, 1981); (3) the common underlying cognitive processes framework, which states that individual differences in reading skills in L1 and L2 can be predicted by a common set of underlying cognitive constructs (e.g., phonological awareness and decoding) (e.g., Geva & Ryan, 1993), and (4) the transfer facilitation model (Koda, 2005, 2007, 2008) briefly noted earlier.

Koda's transfer facilitation model, according to Chung, Chen, and Geva's (2019) recent review on cross-language transfer in bilingual reading, is "the most elaborate theory of transfer to date" (p. 158). The model specifies that what is transferred in L2 reading acquisition is not a set of L1 linguistic rules or a holistic construct like L1 reading ability or L1 proficiency but rather subskills, notably metalinguistic awareness (including phonological awareness and MA), that have been formed through L1 processing or literacy experiences. It highlights the *non-volitional and automatic* nature of transfer, and provides predictions on the effects of multiple factors on transfer or conditions for the transfer of a skill to happen, such as L1-L2 type, L1-L2 proficiency, and linguistic complexity. Since it was first proposed in Koda (2005), the model has been widely tested in empirical studies (see a review in Koda & Reddy, 2008); some conceptualizations underpinning the model, however, wait to be further explored. For one thing, the automatic nature of transfer might only account for L2 reading in older learners, who are typically skilled L1 readers; and the model thus has left unclear transfer occurring in younger learners acquiring L1 and L2 literacy simultaneously (see also Chung et al., 2019). For another, the model seems to only consider L1-to-L2 unidirectional transfer; however, there is emerging evidence that bidirectional transfer is possible (e.g., J. Zhang et al., 2010). Nonetheless, it should be pointed out that the central tenet of the model, which underscores a facilitative transfer effect of L1 metalinguistic awareness, is consistent with the broad conceptualizations in the literature of L1 as *resources* in L2/bilingual reading (Genesee et al., 2006); and that the association between L1 and L2 MA predicted by the model has been largely supported by some empirical studies. Consequently, we have adopted this model to guide the present study.

Intralingual and Interlingual Associations of MA with L2 Word Decoding and Reading Comprehension

MA plays a dual role in reading development through its capacity for enabling children to segment words into their morphological components (Koda, 2000, 2005; see also Nagy et al., 2014): (a) to help them learn how oral language components are mapped onto the written symbols that encode them (e.g., word decoding); and (b) to promote analytical and constructive approaches to word knowledge development

To our knowledge, there are only two meta-analyses that may shed some light on intralingual associations between MA and word decoding. Goodwin and Ahn (2010, 2013) were both meta-analyses that focused on morphological instruction and reading development; neither, however, specifically targeted L2 readers. Goodwin and Ahn (2013), for example, examined the effect of morphological instruction in mixed types of child readers of English (L1 English typical readers, L1 poor readers/spellers, L1 children with reading disabilities, and L2 English language learners), and found a moderate effect of morphological instruction on decoding ($d = 0.59$).

As to the intralingual relationship between MA and English reading comprehension, Goodwin and Ahn (2013) did not find any significant effect of morphological instruction on English reading comprehension. However, this result was again based on a mixed population of

L1 and L2 child readers. Jeon and Yamashita (2014) was among the first meta-analyses that included aggregated evidence specific to L2 reading (both child and adult readers were included, yet not distinguished in the analysis). In that study, *morphological knowledge* was used as an overarching term that included both MA and morphological processing (i.e., the tacit use of morphology; see Nagy et al., 2014), and it was found that the overall mean correlation between L2 *morphological knowledge* and L2 reading comprehension was significant and large[2] ($r = 0.61$, $k = 6$).

With respect to the interlingual association between L1 MA and L2 word decoding and that between L1 MA and L2 reading comprehension, recent critical reviews have suggested that L1 MA is related to L2 MA, which subsequently contributes to L2 word decoding and L2 (passage) reading comprehension (Ke & Xiao, 2015; Koda & Ke, 2018). However, no quantitative reviews or meta-analyses are currently available in the literature.

Moderating Effects of Linguistic-, Learner- and Task-related Factors

Many factors may influence the extent to which MA is shared across languages and contributes to biliteracy development. For example, MA is a complex construct that entails diverse aspects and levels of insights into morphologically complex words; previous studies have varied in their operational definitions of MA and used a wide range of MA measurement tasks. Those studies have also focused on learners of diverse biliteracy profiles (see Ke & Xiao, 2015). In Jeon and Yamashita's (2014) meta-analysis, moderator analysis was not conducted on the relationship between MA and L2 reading. In the present study, we sought to consider the potential moderating effects of some selected factors of theoretical and methodological interest, which represent linguistic-, learner- and task-related influences, respectively: word formation rules (i.e., inflection, derivation and compounding), L1-L2 writing system type, age, MA and reading outcome measurement characteristics. We define each moderator below and review pertinent literature. Further detail about the extent to which and how these factors have been investigated in the primary studies are presented in the Results section.

Word formation rules. Three oft-cited word formation rules are inflection, derivation, and compounding. Since the majority of the primary studies focused on English as the target L2, the examples described below are based on English. Inflection refers to a morphological process of modifying the grammatical properties of lexemes (e.g., *walked, walking*) while maintaining their basic lexical meaning (e.g., *walk*); derivation is characterized by an affixational process that forms new words by modifying the meaning, and often the word class as well, of the base lexeme (e.g., *walker*); compounding involves the combination of two or more lexemes (e.g., *walkover*). The shareability of MA in L2 reading subskill development may depend on linguistic-related factors such as word formation rules in the target L2 (e.g., Xue & Jiang, 2017; Yeh, Joshi, & Ji, 2015; D. Zhang, 2013). For example, D. Zhang (2013) compared the relative interrelationships of compound and derivational awareness in Chinese-English biliteracy development in sixth graders learning English as a foreign language in mainland China. The results suggested that the correlation between Chinese MA and English MA was larger for compound words than for derived words, and that Chinese compound awareness correlated significantly with English compound word meaning inferencing, whereas Chinese derivational awareness did not correlate significantly to English derived word meaning inferencing. D. Zhang (2013) held that the results might be due to the linguistic distance between Chinese and English, because compounding is a productive way of word formation in both languages whereas derivation is more productive in English word formation.

L1-L2 writing system type. Writing systems refers to the principles governing mapping between graphemes and various spoken language units such as phonemes, syllables, and morphemes (Chang, Plaut, & Perfetti, 2016). According to Chang et al. (2016), there are five writing systems for existing languages: alphabet (e.g., English), abjad (e.g., Arabic), alphasyllabary (e.g., Hindi), syllabary (e.g., Japanese), morphosyllabary (e.g., Chinese). MA plays a role in reading across writing systems; however, it also reflects the particular grapheme-phoneme-morpheme mappings of each writing system (Koda & Miller, 2018). The shareability of MA for L2 reading might depend in part on similarities between the L1 and L2 (e.g., in Ramirez, 2010, both L1 Spanish and L2 English are alphabetic).

Age. Developmentally speaking, age is an important factor in the relationship between MA and (bi)literacy development. Although L1 (English) reading research suggests that MA does not emerge until grade three and above (e.g., Berninger, Abbott, Nagy, & Carlisle, 2010), Goodwin and Ahn's (2013) meta-analysis of the effects of morphological instruction on English reading development suggested that MA's effect is greater for younger learners than learners at upper grade levels (though remember that this study combined typical monolingual readers, those with reading disabilities, and L2 readers). It should be noted that there are other pertinent learner-related factors (e.g., L1-L2 proficiency and educational contexts) (Chung et al., 2019; Koda, 2005). However, our initial coding results indicated that these factors were neither systematically examined nor coded transparently in prior studies. We thus did not include them in the moderator analysis. We discuss this further in the Limitations and Future Directions section.

MA and reading outcome measurement characteristics. A few reviews have included discussion of measurement of (L1) MA in different languages (e.g., Arabic in Tibi & Kirby, 2017; Chinese in Liu, Zhou, & McBride-Chang, 2010; and English in Apel, 2014). Apel (2014) summarized that, overall, quite a range of tasks, either norm-referenced or researcher-designed, have been used to assess students' MA, such as judgment tasks, production tasks, blending/segmentation tasks, analogy tasks, and affix identification tasks (see examples of all measurement tasks in Appendix S3 in the Supporting Information online). However, no single task adequately assesses all of the components of MA as a complex, multifaceted construct. The assessment of reading outcomes (i.e., word decoding and reading comprehension) in the primary studies was also coded and examined in this research.

Viewed collectively, although it is generally agreed that MA plays an important role in biliteracy development, there is a lack of consensus regarding the extent to which and how MA functions as a cross-linguistically shareable resource in biliteracy development across languages and writing systems. Since MA is a complex construct, it is crucial to adopt a componential view (Carr & Levy, 1990) and examine the cross-linguistic association of MA, and the extent to which MA associates with the development of different reading subskills (i.e., word decoding and reading comprehension) interlingually and intralingually in biliteracy development, as well as the influence of linguistic-, learner- and task-related factors on MA's shareability.

The Present Study: Goals and Research Questions

To recapitulate, the goals of this research were threefold. First, it sought to provide a scoping review of the research methods applied in previous studies to investigate the relationship between MA and child biliteracy development. Moe specifically, we explored the study design and sample characteristics in the primary studies, and how MA has been conceptualized and measured in the primary studies. Second, it sought to determine to what extent and how MA correlates with child L2 reading development, specifically the cross-linguistic shareability of MA and its relationships with L2 word decoding and reading comprehension, and the moderating effects of linguistic-, learner- and task-related factors. Last but not least, it sought to provide implications for future empirical research, meta-analyses and pedagogical practice. Specifically, to address the second goal, we identified the following four research questions (*RQ*s) to guide the meta-analysis of correlations.

RQ1. To what extent does L1 MA correlate with L2 MA?

RQ2. To what extent does L1 MA correlate with L2 word decoding and with L2 reading comprehension interlingually?

RQ3. To what extent does L2 MA correlate with L2 word decoding and with L2 reading comprehension intralingually?

RQ4. To what extent do the relations above (in *RQ*s1-3) vary as a function of word formation rules, L1-L2 writing system type, age, as well as MA and reading outcome measurement characteristics?

Method

Inclusion and Exclusion Criteria

Primary studies were selected based on a-priori criteria. This meta-analysis included primary research that (1) was written in English between 1975 and March 2019; (2) included participants aged between five and eighteen years old without a learning or reading disability; (3) presented empirical data based on direct testing of L1 MA and/or L2 MA; (4) reported measurement of MA as explicit awareness (as opposed to tacit/implicit knowledge) and other L2 reading outcomes (e.g., word decoding and passage-level reading comprehension); and (5) reported (or the authors shared upon contact) sample size and data of/related with effect size, including correlation coefficient *r*, and/or descriptive statistics with *mean* and *SD*. In addition, there were two exclusion criteria: (1) duplicated reports were excluded; (2) studies on adult L2 readers only were excluded. Since this study aimed to explore the associations between MA and L2 reading component subskills and to identify potential moderating effects, only correlational evidence was included in the analysis.

Literature Search

To locate pertinent primary studies, a literature search was conducted in four phases. First, four sets of key words, including “morphological awareness”, AND “second language reading OR L2 reading, OR transfer, OR reading skill/subskill”, were used in Boolean searches among five electronic databases: ERIC, LLBA, ProQuest Dissertations, PsycINFO, and SciVal. Second, the same key words were searched using Google Scholar and Web of Science. Third, manual searches were conducted among 14 scholarly journals which are highly regarded in L2 learning research (the list of L2 journals can be found in Li, Shintani, & Ellis, 2012), and some specific to reading research, including *Journal of Educational Psychology*, *Journal of Literacy Research*, *Journal of Research in Reading*, *Reading and Writing*, *Reading Research Quarterly*, *Reading Psychology*, and *Scientific Studies of Reading*. Finally, references in recent research syntheses, narrative reviews, and previous primary studies were examined (e.g., Bowers et al., 2010; Carlisle, McBride-Chang, Nagy & Nunes, 2010; Goodwin & Ahn, 2010, 2013; Jeon & Yamashita, 2014; Kirby & Bowers, 2017; Koda & Reddy, 2008; Lee, 2011; Nagy et al., 2014; Reed, 2008; Ruan et al., 2018). As of March 31, 2019, a total of 1348 studies, including duplicated reports, were identified. After screening using the a-priori criteria, 34 studies with 41 independent samples were included for initial analysis (as illustrated in Figure 1).

<Insert Figure 1 about here>

Figure 1. Literature Search and Study Screening

Coding of Effect Sizes and Moderator Variables

The complete coding is available at: https://osf.io/vrcm8/?view_only=4c6a9ddd968f4606a60e4aee3a4ab5cf. Since this study aimed to examine the correlations between MA and L2 reading outcomes, the coding was based on the effect size in a zero-order correlation matrix (i.e., Pearson's *r* in primary research). For primary studies of longitudinal design, only correlations recorded at the first time point were coded to avoid interdependence[3]. Because it is common to find multiple measures of MA and reading outcomes (e.g., real word reading and pseudoword reading for word decoding) in primary studies, we followed Branum-Martin et al. (2012), Lee (2011), and Ruan et al. (2018) and used the arithmetic mean of effect sizes for matched MA measure per reading outcome category. Ultimately, five outcomes were coded: the correlations between L1 and L2 MA (k = 20), between L1 MA and L2 word decoding (k = 19), between L1 MA and L2 reading comprehension (k = 6), between L2 MA and L2 word decoding (k = 31), and between L2 MA and L2 reading comprehension (k = 17). Given the small sample size for the relationship between L1 MA and L2 reading comprehension (k = 6), moderator analyses[4] were not run for that relationship, but were run for the other four relationships (i.e., between L1 MA and L2 MA, between L1 MA and L2 word decoding, between L2 MA and L2 word decoding, and between L2 MA and L2 reading comprehension).

For the moderator variables, the coding was conducted with attention to six factors: word formation rules (inflection, derivation, compounding), L1-L2 writing system type, age, MA measurement task type, and reading outcome measurement characteristics (i.e., word decoding tasks, and reading comprehension tasks) (see the complete coding scheme in Appendix S1 in the Supporting Information online). The descriptions of moderator coding are as follows.

Word formation rules. Four categories were coded with regard to the word formation rules in L1 and L2, including inflection only, derivation only, compounding only, and mixed. The distribution of target L2s (Arabic, Chinese, English, French, and Hebrew) and word formation rules in the primary studies can be found in Appendix S2 in the Supporting Information online.

L1-L2 writing system type. Based on the primary studies, eight different L1-L2 writing system combinations were coded: L1 abjad-L2 abjad (e.g., Eviatar et al., 2018), L1 abjad-L2 alphabet (e.g., Schiff & Calif, 2007), L1 alphabet-L2 abjad (e.g., Saiegh-Haddad & Geva, 2008), L1 alphabet-L2 alphabet (e.g., Kieffer & Lesaux, 2008), L1 alphabet-L2 morphosyllabary (e.g., Sun & Curdt-Christiansen, 2016), L1 alphasyllabary-L2 morphosyllabary (Zhou et al., 2018), L1 morphosyllabary-L2 alphabet (e.g., Tong et al., 2018), and mixed L1s and L2 alphabet (e.g., D'Angelo et al., 2017).

Age. We first sought to use mean age as the moderator variable, but not all primary studies provided enough information for calculating mean age. Often, only an age range was given, and attempts to contact the authors of the primary studies sometimes did not result in greater precision. Therefore, instead of using mean age, we used five age groups to indicate the participants' age characteristics, including kindergarten to grade two (K-G2), grades three to five (G3-5), grades six to eight (G6-8) and grades nine to twelve (G9-12), and mixed grade level (Mixed) (similar categories can be found in the meta-analyses by Goodwin & Ahn, 2010, 2013).

As a result of this categorization, we identified a greater number of studies that included child participants between kindergarten and fifth grade. Details can be found in Tables 3-6 in the Results section.

MA measurement type. A total of eight different MA tasks have been used in the primary studies: (1) affix choice (after Singson, Mahony, & Mann, 2000) (k = 1), (2) morphological structure awareness (e.g., McBride-Chang et al., 2005) (k = 4), (3) morphological decomposition (e.g., Saiegh-Haddad & Geva, 2008) (k = 2), (4) morphological relatedness (e.g., Saiegh-Haddad & Geva, 2008) (k = 3), (5) riddle guess (e.g., Berninger & Nagy, 1999) (k = 1), (6) sentence analogy (after Nunes et al.,1997) (k = 2), (7) the Test of Morphological Structure (after Carlisle, 2000) (k = 17), and (8) the Wug Test (e.g., Eviatar et al., 2018) (k = 3). Another eight samples used mixed measures (see examples of all measurement tasks in Appendix S3 in the Supporting Information online). A review of the most widely cited MA measurement task (i.e., the Test of Morphological Structure) and sample items are provided in the Discussion section.

Word decoding task. Three types of coding were included (after García & Cain, 2014): accuracy, speed, and efficiency (both accuracy and speed).

Reading comprehension task. Again, following García and Cain (2014), we coded this moderator with six levels: multiple choice, cloze, open-ended questions, picture matching, reading aloud, and mixed.

Intercoder Reliability

The first two authors coded 72% of the samples. The reliability for continuous effect size coding was .80 (Cronbach's alpha). For categorical variable coding, Cohen's kappa was .93. The agreement rate was 97.3%. Recoding continued until inconsistencies were resolved. The first author then coded the rest of the samples.

Meta-analytic Procedures

This study followed the suggestions by Borenstein, Hedges, Higgins, and Rothstein (2009) and Cooper (2010) for doing meta-analyses, as well as works on meta-analysis methods in second language acquisition research (e.g., Li et al., 2012; Norris & Ortega, 2006; Plonsky & Oswald, 2012; Teimouri, Goetze, & Plonsky, 2019). All data coding and analyses were conducted using Comprehensive Meta-Analysis (CMA) software Version 3.0 (https://www.meta-analysis.com/) and Microsoft Excel. Below, we describe the six major steps of the analysis.

First, following Teimouri et al.'s (2019) procedure to remove outliers (defined as standard residuals > 2.5), we examined the standard residuals in the primary studies, and removed[5] one outlier for the relationship between L2 MA and L2 reading comprehension (standard residual = 3.18 for the Vietnamese subgroup of L2 English learners in Kieffer & Lesaux, 2012), which resulted in 34 studies and 40 independent samples left for the final analysis. Second, data analysis was based on random-effects models, which, according to Borenstein et al. (2009), is a more plausible match than fixed-effect models when studies are selected from the published literature. Third, to examine the overall correlations, a weighted average of the correlations of all primary studies was calculated by an estimation of the inverse variance of each effect size (see Borenstein, Hedges, & Rothstein, 2007). Fourth, in order to investigate whether there was variability among the correlations across primary studies, a heterogeneity test (Q-test) (Hedges & Olkin, 1985) was conducted. A significant value on the heterogeneity test would provide evidence for variability among the correlations, and I^2 indicates the proportion of the heterogeneity that is real rather than due to chance. In addition, the parameter τ^2 indicates the variance in true effect sizes[6]. Fifth, regarding moderator effects, considering that all moderator variables in this study are categorical, a heterogeneity test (Q-test) was performed both within and across subgroups based on random-effects modeling. A statistically significant difference between subgroups would suggest that there is influence of a moderator on the mean correlation. The analysis of potential moderators indicates whether subgroup membership affects the correlational outcomes. Following one of the anonymous reviewers' suggestions, for cases where there was only one sample for a certain subgroup, we conducted a sensitivity analysis and removed one sample at a time to see if the results would be altered (all sensitivity analysis results can be found in Appendix S4 in the Supporting Information online). Last, following the suggestions of the anonymous reviewers and practices of previous meta-analyses, we examined possible effects of publication bias in these ways: (1) both published and unpublished research were selected in the meta-analysis, including 29 journal articles, one book chapter, and four unpublished doctoral dissertations. (2) Funnel plots for random-effects models were used to determine the presence of retrieval bias, for which the funnel would be asymmetric (see Appendix S5 in the Supporting Information online). (3) The results of Duval and Tweedie's Trim and Fill were reported in Table 2 when asymmetries were found in the funnel plots.

Results

In this section, we first provide an overview of the study and sample characteristics, and summarize the conceptualization and operationalization of MA in the primary studies. We then report the descriptive statistics with regard to the meta-correlation analysis results for the cross-linguistic association between L1 and L2 MA, as well as the intralingual and interlingual relationships of L1 MA and L2 MA with L2 word decoding and L2 reading comprehension. Finally, the results of moderator analysis are presented.

A Scoping Review of the Research Methodology in Selected Studies

This meta-analysis included 34 correlational studies with 40 independent samples (N = 4,056) of nine languages (Arabic, Chinese, English, French, Hebrew, Korean, Malay, Spanish, and Tagalog) and four writing systems (i.e., abjad, alphabet, alphasyllabary, and morphosyllabary) when sample participants' L1 and L2 were both considered. Of these, 26 samples (approximately 65.0%) focused on English as the L2; the other L2s included Arabic, Chinese, French and Hebrew. Half of the samples (k = 20) focused on derivation; six included compounding words; four examined inflectional words; and another 10 samples were of mixed word formation rules. There were eight possible combinations of L1 and L2 writing systems, yet in 22 out of 40 samples L1 and L2 were both alphabetic languages. Regarding age of participants, 13 samples (32.5%) focused on participants who were students below grade three; 15 samples (37.5 %) included participants in grades three through five; one in grades six to eight; one in grades nine to twelve; and 10 were of mixed grade levels[7]. Twenty-eight samples (70.0%) were from North America, eight from Asia, and four from the Middle East. In addition, the selected research included 17 cross-sectional and 17 longitudinal studies. We identified eight interventional studies in the literature search and were able to retrieve the correlational data from two studies only (even after attempting to contact the study authors), of which we included one (i.e., Filippini, 2007) and excluded one based on the a-priori criteria.

The coding of operational definitions and measures of MA in the selected studies, target L2s, and sample MA task items can be found in Appendix S3 in the Supporting Information online. Regarding the definition of MA, 18 (52.9%) out of the 34 studies cited Carlisle (1995, 2000, 2003) and referred to MA as “conscious awareness of the morphemic structure of words and [the] ability to reflect on and manipulate that structure;” five studies (14.7%) followed Kuo and Anderson (2006, p.161) and considered MA as “the ability to reflect upon and manipulate morphemes and employ word formation rules in one’s language;” another two (5.9%) cited both; the rest had different definitions by citing different sources or even without citing a source.

As to the tasks used to measure MA in the primary studies, as mentioned earlier, we have identified eight different tasks. The most widely adopted measure was the Test of Morphological Structure, with 16 (40.0%) out of the 40 samples using this measure. We further review this task and evaluate its validity in the Discussion section.

Last, with regard to outcome measurement characteristics, most of the samples measured word decoding accuracy only (k = 25), while the rest (k = 9) measured word decoding efficiency (including both accuracy and speed). As for reading comprehension, the majority of studies (k = 9) used a multiple-choice format, three used cloze tests, one used open-ended questions (Gottardo, Mirza, Koh, Ferreira, & Javier, 2018), one used picture matching (Ramirez, Chen, & Pasquarella, 2013), one used reading aloud (Lok, 2014), one used recall (Filippini, 2007), and for another two, mixed formats were used. The potential effects of outcome measurement characteristics are reported later.

For the purpose of evaluating the quality of the selected studies (see Marsden, Morgan-Short, Thompson, & Abugaber, 2018), we summarize the internal reliabilities (based on Cronbach’s alpha) of L1 and L2 MA, as well as L2 word decoding and reading comprehension measures reported in primary studies in Table 1. All primary studies reported the internal reliability (Cronbach's alpha) for at least one instrument. Regarding the respective numbers of studies that reported the reliability of L1 MA, L2 MA, L2 word decoding and L2 reading comprehension tasks, the respective proportions were 78.3%, 92.1%, 76.5%, and 80.0%. The L1 and L2 MA tasks seemed slightly less reliable (respective medians = 0.79, 0.82) when compared to the 25%, 50% and 75% percentiles of instrument reliability coefficients in Second Language Acquisition (SLA) research (i.e., 0.74, 0.82, 0.89) or the median of SLA research that focused on reading skills (i.e., 0.86) reported by Plonsky and Derrick (2016). L2 word decoding and reading comprehension tasks were relatively more reliable (respective medians were .94 and .86).

Table 1

Reliability Information (Cronbach's alpha) Reported in the Independent Samples

Tasks	Mean	Median	Range
L1 MA (*k* = 23)	.75	.79	.34, .93
L2 MA (*k* = 38)	.81	.82	.61, .95
L2 word decoding (*k* = 34)	.93	.94	.77, .99
L2 reading comprehension (*k* = 19)	.85	.86	.73, .94

In summary, in response to our first goal of conducting a scoping review, the results suggest that the majority of primary studies tended to (1) adopt an observational design (without intervention), (2) include participants between grades one and five, who learn to read in an alphabetic L1 and L2, and learn English as the target L2 in North America, (3) define MA as "conscious awareness of the morphemic structure of words and their ability to reflect on and manipulate that structure" after Carlisle (1995, 2000), (4) measure MA with the Test of Morphological Structure (Carlisle, 2000) or in a similar fashion, and (5) measure word decoding accuracy as the target outcome and pay less attention to reading comprehension. There was no notable difference in the numbers of cross-sectional versus longitudinal studies (17 and 17, respectively). Additionally, only one independent sample included in this research involved an MA intervention.

Cross-linguistic Association Between L1 and L2 MA

To recap, to address our second goal, we identified four research questions involving quantitative meta-analysis of correlations. First, regarding the correlation between L1 and L2 MA, 20 independent samples (N = 1,657) were analyzed (see Table 2 and Figure 2). The overall mean correlation between L1 and L2 MA was small (r = 0.30) yet significant (95% CI [0.22, 0.38], z (19) = 7.14, p < .001). In addition, there was significant and moderate heterogeneity in the correlations (Q (19) = 48.58, p < .001, I^2 = 60.89%, τ^2 = 0.02).

<Table 2 inserted about here>

<Figure 2 inserted about here>

Figure 2. Forest plot for the relationship between L1 and L2 MA

Interlingual Relationship between L1 MA and L2 Reading Outcomes

We first report the correlations between L1 MA and L2 word decoding (N = 1,761). The findings showed that there was a significant and small correlation (r = 0.35, k = 19, 95% CI [0.25, 0.43], z (18) = 6.95, p < .001) (as illustrated in Table 2). The heterogeneity analysis indicated that the correlations between L1 MA and L2 word decoding varied significantly (Q (18) = 73.13, p < .001, I^2 = 75.39%, τ^2 = 0.04) (see also Figure 3).

<Figure 3 inserted about here>

Figure 3. Forest plot for the relationship between L1 MA and L2 word decoding

Second, the mean correlation between L1 MA and L2 reading comprehension (N = 467) was also found to be significant and of small effect size (r = 0.39, k = 6, 95% CI [0.29, 0.48], z (5) = 7.08, p < .001) (see also Table 2). There was non-significant variation among the samples (Q (5) = 8.39, p = .136, I^2 = 40.43%, τ^2 = 0.01) (as illustrated by Figure 4).

<Figure 4 inserted about here>

Figure 4. Forest plot for the relationship between L1 MA and L2 reading comprehension

In sum, to answer *RQ2*, the results indicate that the interlingual correlations between L1 MA and L2 word decoding and between L1 MA and L2 reading comprehension were both small yet significant.

Intralingual Relationship between L2 MA and L2 Reading Outcomes

We first analyzed the relationship between L2 MA and L2 word decoding (N = 2,754), and found a medium and significant correlation (r = 0.45, k = 31, 95% CI [0.40, 0.50], z (30) = 15.94, p < .001). The heterogeneity test was significant (Q (29) = 74.53, p < .001, I^2 = 60.98%, τ^2 = 0.02) (see the forest plot in Figure 5).

<Figure 5 inserted about here>

Figure 5. Forest plot for the relationship between L2 MA and L2 word decoding

Lastly, the mean correlation between L2 MA and L2 reading comprehension was analyzed (N = 2,575), and we found a moderate and significant correlation (r = 0.52, k = 17, 95% CI [0.46, 0.57], z (16) = 13.90, $p < .001$). There was significant variation among the studies (Q (16) = 57.70, $p < .001$, I^2 = 77.27%, τ^2 = 0.02) (as illustrated in Figure 6).

<Figure 6 inserted about here>

Figure 6. Forest plot for the relationship between L2 MA and L2 reading comprehension

In response to *RQ3*, the findings suggest that the intralingual correlations between L2 MA and L2 word decoding and between L2 MA and L2 reading comprehension were both medium and significant.

Moderating Effects

As mentioned earlier, moderator analyses were conducted and reported for four relationships (i.e., between L1 MA and L2 MA, between L1 MA and L2 word decoding, between L2 MA and L2 word decoding, and between L2 MA and L2 reading comprehension). There are four major findings. First, only one factor, that is, MA measurement type, had a statistically significant effect on the correlation between L1 MA and L2 MA. As shown in Table 3, the correlation was moderate in studies that used the morphological structure awareness task (r = 0.55, k = 2, 95% CI [0.38, 0.68]), small in studies that used mixed tasks (r = 0.32, k = 4, 95% CI [0.21, 0.41]) or the morphological relation measure (r = 0.34, k = 3, 95% CI [-0.09, 0.66]), and minimal for studies that adopted such tasks as sentence analogy (r = 0.22, k = 2, 95% CI [0.05, 0.37]) and the Test of Morphological Structure (r = 0.23, k = 7, 95% CI [0.14, 0.32]). There was no overlap between the studies that used the morphological structure awareness task and studies that used the Test of Morphological Structure.

<Table 3 inserted about here>

Second, regarding the interlingual relationship between L1 MA and L2 word decoding, moderator analyses did not find any significant moderating effects (as shown in Table 4). The effect of MA measurement type approached a level of statistical significance ($p = 0.055$), yet there were overlaps in the 95% CIs among MA measurement types.

<Table 4 inserted about here>

Third, we did not find any significant moderator effects on the correlation between L2 MA and L2 word decoding. The respective effects of age and MA measurement type approached the significance level ($p = 0.056, 0.050$). Yet, there were overlaps in the 95% CIs between the two age groups (K-G2 and G3-5), and among the different measurement types (see Table 5).

<Table 5 inserted about here>

Last, age was the only significant moderator for the correlation between L2 MA and L2 reading comprehension. As shown in Table 6, the effect size was bigger when MA was tested in children at grades three to five ($r = 0.59$, $k = 6$, 95% CI [.50, .66]). There was no overlap regarding the 95% CIs between the two age groups (K-G2 and G3-5).

<Table 6 inserted about here>

In summary, two significant moderators were identified, namely, MA measurement type and age. MA measurement type had a significant moderating effect on the correlation between L1 and L2 MA. The effect size was moderate for studies that measured MA with the morphological structure awareness task, and minimal for studies that used the Test of Morphological Structure. Age was a significant moderator for the relationship between L2 MA and L2 reading comprehension. The effect size was bigger for children at the upper elementary grades. We did not find any statistically significant moderating effects of word formation rules, L1-L2 writing system type, or reading outcome measurement characteristics.

Discussion

Summary of Findings

To reiterate, this study included systematic efforts to (a) examine the cross-linguistic sharing of MA between L1 and L2, (b) investigate MA's interlingual and intralingual relationships with two different reading subskills (i.e., word decoding and reading comprehension) in L2, and (c) explore the influence of moderators of theoretical and methodological interest (i.e., word formation rules, L1-L2 writing system type, age, as well as MA and reading outcome measurement characteristics). There are three major findings. First, L1 and L2 MA correlated significantly across nine different languages and four writing systems. Second, the interlingual correlations between L1 MA and L2 word decoding and between L1 MA and L2 reading comprehension were small and significant, while the intralingual correlations between L2 MA and L2 word decoding and between L2 MA and L2 reading comprehension were both moderate and significant. Finally, MA measurement type was a significant moderator for the correlation between L1 and L2 MA; age was a significant moderator for the correlation between L2 MA and L2 reading comprehension.

The Shareability of MA in Child Biliteracy Development: What Does the Evidence Tell Us?

It is noteworthy that the majority of the studies selected in this synthesis included participants between grades one and five, who learned to read in two alphabetic languages with English as the target L2 in North America. The evidence present in primary studies and analyzed in the current meta-analysis is largely correlational, not causal (based on intervention). Whereas previous meta-analyses only examined the intralingual relationship between L2 MA and L2 reading (e.g., Jeon & Yamashita, 2014), our research is among the first to add aggregated evidence of the cross-linguistic associations between MA and L2 reading outcomes. Regarding the cross-linguistic sharing of MA between two languages in child biliteracy development, a small yet significant correlation was found. Another important finding was that L1 MA was significantly associated with both L2 word decoding and L2 reading comprehension, although both effects were small. These findings are consistent with the prediction of the transfer facilitation model (Koda, 2005, 2007 2008), namely that MA is a shareable and valuable resource for L2 readers. Note, however, that the correlation between L1 and L2 MA was affected by MA measurement type. This is elaborated later in this section in the discussion of moderator effects.

In contrast with the small effect sizes observed in the interlingual correlations, the intralingual correlations between L2 MA and L2 word decoding, and between L2 MA and L2 reading comprehension were both moderate. Notably, we identified more studies with word decoding as an outcome than passage-level reading comprehension. In Jeon and Yamashita's (2014) meta-analysis, they identified a large correlation between L2 MA and L2 reading comprehension ($r = 0.61$) with a small sample size ($k = 6$), whereas in this research, the effect size was slightly smaller ($r = 0.54$) and with significant variation in the 17 independent samples.

A third area that warrants further discussion is moderator effects. This research conducted various moderator analyses, including word formation rules, L1-L2 writing system type, age, as well as MA, word decoding and reading comprehension measurement characteristics, so as to expand our knowledge of the linguistic-, learner-, and task-related factors that might constrain or bootstrap MA's contributions to biliteracy development. With regard to the two linguistic-related moderators (word formation rules, L1-L2 writing system type), no significant moderating effects were found for the interlingual relationships between L1 and L2 MA, or between L1 MA and L2 word decoding. This seems to suggest that the shareability of MA between two languages is not subject to linguistic constraint.

A second moderator to be considered is age. As mentioned earlier, most of the primary studies focused on children from the kindergarten level up to grade five; and we categorized them into lower level (K-G2) and upper level (G3-5). Interestingly, there was no significant difference between the two groups in the correlations between L1 and L2 MA, or between L1 MA and L2 word decoding, or between L2 MA and L2 word decoding, yet there was a significant difference in the correlation between L2 MA and L2 reading comprehension. In other words, it seems that the transferability of MA is not constrained by age, and that L1 MA is associated with the development of L2 MA and L2 word decoding early on; yet, the role MA plays in meaning retrieval tasks (e.g., reading comprehension) might not surface until upper grade levels. Our findings are consistent with previous empirical studies that underscored a developmentally more important role of MA in monolingual reading in grade three and above (e.g., Berninger et al., 2010); yet they appear to contradict the finding of Goodwin and Ahn's (2013) study which included L2 as well as L1 English readers and where the effect of morphological instruction was found to decrease as grade level increases.

With respect to MA measurement type, this study observed a statistically significant effect on the correlation between L1 MA and L2 MA, but no significant effect on other correlations. As mentioned above, the correlation was large in studies that used the morphological structure awareness task, yet small for those that measured MA with the Test of

Morphological Structure. A follow-up analysis indicated that the two studies that used the morphological structure awareness task were both based on research with L1 Chinese (morphosyllabary) learners of L2 English (alphabet), whereas six out of the seven studies that adopted the Test of Morphological Structure focused on children with two alphabetic languages (see Appendix S2 in the Supporting Information online).We review the two tasks with more detail below.

The morphological structure awareness task was developed by McBride-Chang et al. (2005). It was first tested in studies of monolingual children in different cultures (China, Korea, and the U.S.), and aimed to assess “the awareness of and access to morphemes, reflected in the ability to apply morphemic knowledge to recognize and create new word forms that are morphologically complex and conform to the structure of a given language” (p.141). In this task, children are asked to come up with a novel compound word to describe the object or concept for a given scenario. There are two sections. An example for the first section is: “Early in the morning, we can see the sun rising. This is called a sunrise. At night, we might also see the moon rising. What could we call this? (Answer: moonrise)” (McBride-Chang et al., 2005, p.433). In the second section, children are asked to generate a novel compound word on their own. For instance, “what do we call a monster that only eats pizza?” (Answer: pizzaeating monster). There might be an issue with test validation (measuring what a test is intended to measure; Kerlinger, 1999; Messick, 1989). While the second section of the task might require syntactic manipulation, the first section can be completed with analogy reasoning only. In other words, children may simply substitute “sun” in “sunrise” with “moon” and then produce “moonrise,” without using any morphemic awareness (see also Liu, Zhou, & Mcbride-Chang, 2010). Also, when it comes to the predictability power of this task in readers of two alphabetic orthographies, it is unclear since it has not been empirically tested.

As mentioned earlier, a large number of primary studies cited Carlisle's (1995, p. 194) definition of MA as "conscious awareness of the morphemic structure of words and their ability to reflect on and manipulate that structure" and adapted the Test of Morphological Structure after Carlisle (2000) (an earlier version appeared in Carlisle, 1988). In a study with third and fifth graders in the U.S. in 2000, Carlisle explained the development of the Test of Morphological Structure, which consisted of two parts: one required the decomposition of derived words in order to complete sentences, the other required the production of a derived word in order to complete a sentence. Sample items for the two parts respectively are: "Driver. Children are too young to ____ (answer: drive)." (p.188); "Farm. My uncle is a ____ (answer: farmer)" (p.187). Carlisle (2000) also explained item inclusion criteria: (1) both transparent and shift words are used (e.g., "reason" and "reasonable;" and "produce" and "production"). The task was administered in the oral format. To our knowledge, Goodwin et al. (2012) was among the first to assess the internal structure validity and convergent validity of the Test of Morphological Structure in third through fifth grade English monolingual students and Spanish-speaking English language learners in the U.S. using the Rasch measurement framework. Their conclusion was that the Test of Morphological Structure produced valid and reliable scores that can be used in making literacy research and educational decisions with a few adjustments. However, based on the evidence from selected studies using the Test of Morphological Structure, our research has only found small to medium meta-correlations across the four focal interlingual and intralingual relationships (excluding the relationship between L1 MA and L2 reading comprehension). We urge more research to validate the morphological structure awareness task, the Test of Morphological Structure, and the six other MA tasks in the future.

Lastly, when it comes to potential moderating effects of reading outcome measurement characteristics, we did not find any significant differences. No conclusive judgement of word decoding and reading comprehension measurement characteristics could be made.

Limitations and Future Directions

The findings of this study should be interpreted with caution, and there are a few limitations to be acknowledged. First, the variation in one moderating factor (e.g., age) may be confounded with other factors (e.g., L1-L2 writing system type, MA measurement task characteristics, and L2 proficiency[8]). Since we averaged a number of studies, it is possible that the effect revealed in one or more individual studies might have been overlooked. Second, the sample size for the interlingual correlation between L1 MA and L2 reading comprehension was insufficient for moderator analysis. Also, we included small subgroups in the moderator analyses (a minimum number of two studies per subgroup). Replication of this meta-analysis is needed with a larger pool of studies in the future or with inclusion of both published and unpublished research (e.g., conference proceedings, which were not reviewed in this research). Meta-structural equation modeling based on a larger sample is another possible direction. Third, we examined age as a categorical variable rather than as a continuous variable. As mentioned earlier, the reasons were that the participant samples were often mixed in primary studies, and the age information of some studies was missing despite our attempts to contact the corresponding authors. Fourth, only one independent sample involving morphological intervention (i.e., Filippini, 2007) was included in the present meta-analysis of correlations because we did not retrieve enough correlational data even after contacting the study authors of some other intervention studies. Consequently, we focused mainly on correlation *r* in observational studies in this research; thus, the results could not and should not be interpreted to indicate any causal effect of MA in biliteracy development. For the purpose of examining any causal effect, primary intervention studies are needed (see a review by Kirby & Bowers, 2017). Fifth, the scope of this research is on unidirectional transfer from L1 MA to L2 reading subskills development. There has been emerging evidence of bidirectional transfer in L2 reading development in recent years (e.g., J. Zhang et al., 2010), which is an area that warrants further investigation. Last, the literature search involved a combination of four sets of key words only (i.e., "morphological awareness", AND "second language reading OR L2 reading, OR transfer, OR reading skill/subskill"). Future syntheses or systematic reviews should expand the search with other key words such as "morphological knowledge", "morphemic awareness/knowledge", and "biliteracy".

To date, different theoretical frameworks have been proposed to explain to what extent and how MA predicts L2 reading development intralingually and interlingually, including the transfer facilitation model by Koda (2005, 2007, 2008), and more recently, the interactive framework by Chung et al. (2019). Those frameworks have been supported by an emerging body of primary evidence. Our meta-analytic findings also support the transfer facilitation model. However, as shown in the present meta-analysis, the bulk of evidence is based on studies focusing on the relationships between MA and L2 word decoding among child participants between grades one and five, who learn to read in alphabetic L1 and L2, with English as the target L2 in North America. In what follows, we propose viable ways of addressing several cross-cutting themes, including *complexity*, *individual differences and development* and *context*, which were identified broadly by the National Literacy Panel (NLP) on Language Minority Children and Youth (August & Shanahan, 2006). Although the themes did not target toward MA specifically in the report of the NLP, we believe they serve as a good framework of reference for our discussion here on MA.

First, the issue of complexity involves critical (re)examination of *what* MA is and *how* it contributes to (bi)literacy development. According to our review, at least eight different measures of MA have been used in existing literature, with some being more language specific (e.g., using the morphological structure awareness task in Chinese-English bilingual readers and the Test of Morphological Structure Test in readers of two alphabetic languages). The assessment of MA as a complex, multi-faceted construct should be aligned with its conceptualization by including multiple measures for empirically distinguishable facets (see also Apel, 2014). For example, recent research has begun distinguish between a language-independent facet of MA that involves segmentation of words into sub-lexical components (referred to as a *basic facet* by D. Zhang & Koda, 2013, and as *structural awareness* by H. Zhang & Koda, 2018) and a language-specific facet that involves mapping morphological information onto graphic symbols (a *refined facet* in D. Zhang & Koda, 2013, and *functional awareness* in H. Zhang & Koda, 2018). Future research should further investigate these and other facets of MA using model-testing statistical procedures such as confirmatory factor analysis and the Rasch measurement framework. Also, construct validation should be conducted with typologically diverse languages and writing systems. Regarding the contribution of MA to L2 reading development, this study has identified that previous research has mainly focused on the intra-lingual contribution of L2 MA to L2 word decoding. Thus, more attention should be paid to both interlingual and intralingual contributions to different reading outcomes.

Second, given that the majority of data were collected from children between grades one and five, two questions are in need of further exploration with respect to *individual differences* in the *development* of MA in linguistically diverse groups of L2 learners. One is whether MA plays similar or different roles across ages and grade levels, that is, kindergarten to grade twelve (and even adulthood, though adult L2 reading is beyond the scope of this meta-analysis). The other is how MA facilitates younger learners' transition from using spoken language(s) only to learning to read (e.g., in kindergarten and grade one) (e.g., Kuo & Anderson, 2008; Pan et al., 2016).

Lastly, according to our scoping review, *context* or educational setting in previous studies has usually been taken as the background rather than the study focus; there are at least three gaps to be filled: (a) compared with second language and immersion settings, there are fewer studies in a foreign language setting, at least for child L2 reading, (b) for the purpose of examining any causal effect of MA on L2 reading development, more intervention studies in classrooms are needed for this line of inquiry, and (c) to date, to what extent and how L2 child learners' trilingual/multilingual and biliterate profiles can moderate the transfer facilitation effects of MA is still unclear.

Conclusions

This meta-analysis set out to expand our understanding of the extent to which and how MA functions as a cross-linguistically shareable resource in child biliteracy development. Through the analysis of 34 correlational studies with 40 independent samples (N = 4,056) of nine languages and four writing systems in children from kindergarten to grade 12, we found that MA was associated cross-linguistically between two different languages in child biliteracy development and it was significantly correlated with the acquisition of different L2 reading subskills (i.e., word decoding and reading comprehension). The correlation of MA between L1 and L2 was not constrained by word formation rules or L1-L2 writing system type, but was affected by MA measurement characteristics. Additionally, the intralingual relationship between L2 MA and L2 passage reading comprehension was significantly moderated by age.

Given the heterogeneous conceptualizations and operationalizations of MA in previous studies, we have urged for more research along three cross-cutting themes (*complexity*, *individual differences and development*, and *context*), and anticipate that it will contribute to further systematic inquiries in this area. In sum, three major implications can be drawn for future empirical research, meta-analyses and evidence-based practices: (a) There is a need for future research to align the definition and measurement of MA as a complex, multifaceted construct, and pay equal attention to its interlingual and intralingual contributions to both word decoding and reading comprehension. (b) Based on our finding that the correlation between L1 and L2 MA was not constrained by age, yet there was a higher correlation between L2 MA and L2 reading comprehension among child readers at grades three to five, before teachers introduce any instructional activities that can facilitate the use of children's L1 resources, they might want to wait until children are more cognitively mature and are ready to make use of transferred MA for L2 reading comprehension. (c) More interventional studies are needed to examine the causal effect of morphology on L2 reading development, especially for L2 child learners who are transitioning from using spoken language(s) only to learning to read.

NOTES

1. Although we describe MA in the context of printed words, we do acknowledge that MA is initially developed as an oral language competence. For instance, Pan et al. (2016) operationalized preliterate MA as younger children's (four to six years old) ability to combine known morphemes in new ways using lexical compounding orally in Chinese, and postliterate MA as older children's ability to produce words with the same grapheme unit that represents different morphemes (e.g., children are given the word 面粉 "flour (powder)" with the target character 面, and expected to produce two different words like 面包 "bread" and 面子 "face". 面包 is both orthographically and morphologically related with 面粉; whereas 面子 is only orthographically related with 面粉).
2. The interpretation of effect sizes is based on Plonsky and Oswald's (2014) benchmarks for small, medium, and large correlations (*r* = .25, .40, .60).
3. We have integrated the two types of studies in our final analyses because the exploratory moderating effect analysis did not show any differences between cross-sectional and longitudinal studies (see Appendix S6 in the Supporting Information online). For longitudinal studies, please note that we only included correlational data at the first time point. The main reason for not conducting two separate sets of analyses is that the inferences are different for concurrent and longitudinal studies along the research line of cross-linguistic transfer of L2 reading. Concurrent studies focus on correlational relationships, whereas longitudinal findings indicate possible causal effects (see also Chung et al., 2019). This meta-analysis focuses on the correlational relationships. As we

also note in the Limitations and Future Directions section, there is a need in the future to conduct meta-analyses of causal effects of MA with a larger sample pool.

4. Our meta-analysis procedures followed three previous meta-analysis studies – two pertinent to L2 reading acquisition (i.e., Jeon & Yamashita, 2014, $k = 67$, minimum subgroup sample size = 3, minimum subgroup analysis sample pool = 18; and Melby-Lervåg & Lervåg, 2011, $k = 52$, minimum subgroup sample size = 3, minimum subgroup analysis sample pool = 7), as well as one pertinent to the role of MA in reading development (Ruan et al., 2018, $k = 41$, minimum subgroup sample size = 1, minimum subgroup analysis sample pool = 43) – and conducted moderator/subgroup analyses for independent sample sizes (*k*s) bigger than 10 (see rule of thumb by Borenstein et al., 2009) with a minimum of two samples in each subgroup.

5. Removing the corresponding value from the sample pool decreased the overall result (from $r = 0.54$ to $r = 0.52$). However, excluding this outlier seemed to increase the precision of the averaging results (i.e., narrowing 95% CI from [0.47, 0.60] to [0.46, 0.57]), reducing sample errors indexed by Q from 77.85 to 57.70, and reducing I^2 from 78.16% to 72.27%).

6. Below is the formula for estimating a pooled τ^2 (adopted from Borenstein et al., 2009, p.171).

a. $$T^2 = \frac{Q - df}{C},$$

$$\mathrm{df} = \mathrm{k} - 1,$$

$$\mathrm{C} = \sum W_i - \frac{\sum W_i^2}{\sum W_i}.$$

7. Authors of the primary studies all reported exact grades or grade ranges. Therefore, age differences associated with grade levels in different countries should not be a concern.
8. Although L2 proficiency has been discussed in some transfer frameworks reviewed earlier, the relevance of L2 proficiency seems notably different between those studying child biliteracy (often scholars in educational psychology) and those studying adult L2 reading research (often applied linguists). To our knowledge, it is uncommon for studies of child biliteracy development to consider a so-called L2 proficiency measure. For theoretical reasons, they sometimes considered and measured oral vocabulary and general reading achievements with some standardized tests, such as the Peabody Picture Vocabulary Test by Dunn and Dunn (1997) (e.g., Lam et al. 2012), the Wide Range Achievement Test by Robertson (2001) (e.g., Wang, Cheng, & Chen, 2016), and the Woodcock Language Proficiency Battery-Revised Passage Comprehension subtest by Woodcock (1991) (e.g., Kieffer & Lesaux, 2008). Yet, those measures by no means aimed to document a specific level of L2 proficiency (e.g., intermediate or advanced) in child readers. Overall there lacks systematic and transparent reporting in the primary studies on what children's general L2 proficiency was. These factors have made it impossible to analyze L2 proficiency as a moderator.

REFERENCES

∗References marked with an asterisk indicate studies included in the meta-analysis.

Apel, K. (2014). A comprehensive definition of morphological awareness: Implications for assessment. *Topics in Language Disorders*, *34*, 197–209. https://doi.org/10.1097/TLD.0000000000000019

August, D., Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language minority children and youth*. Mahwah, NJ: Lawrence Erlbaum Associates.

Berninger, V., & Nagy, W. (1999). *University of Washington morphological awareness battery*. Unpublished experimental test battery, Seattle, WA.

Bowers, P. N., Kirby, J. R., & Deacon, S. H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research*, *80*, 144–179. https://doi.org/10.3102/0034654309359353

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley. https://doi.org/10.1002/9780470743386

Branum-Martin, L., Tao, S., Garnaat, S., Bunta, F., & Francis, D. J. (2012). Meta-analysis of bilingual phonological awareness: Language, age, and psycholinguistic grain size. *Journal of Educational Psychology*, *104*, 932–944. https://doi.org/10.1037/a0027755

Carlisle, J. F. (1988). Knowledge of derivational morphology and spelling ability in fourth, sixth, and eighth graders. *Applied Psycholinguistics*, *9*, 247–266. https://doi.org/10.1017/S0142716400007839

Carlisle, J. F. (1995). Morphological awareness and early reading achievement. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp.189-209). Hillsdale, NJ: Lawrence Erlbaum Associates.

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing*, *12*, 169–190. https://doi.org/10.1023/A:1008131926604

Carlisle, J. F. (2003). Morphology matters in learning to read: A commentary. *Reading Psychology*, *24*, 291-322. https://doi.org/10.1080/02702710390227369

Carlisle, J. F., McBride-Chang, C., Nagy, W., & Nunes, T. (2010). Effects of instruction in morphological awareness on literacy achievement: An integrative review. *Reading Research Quarterly*, *45*, 464–487. https://doi.org/10.1598/RRQ.45.4.5

Carr, T. H., & Levy, B. A. (Eds.). (1990). *Reading and its development: Component skills approaches*. San Diego: Academic Press.

Chang, L.Y., Plaut, D.C., Perfetti, C.A. (2016). Visual complexity in orthographic learning: Modeling learning across writing system variations. *Scientific Studies of Reading*, *20*, 64–85. https://doi.org/10.1080/10888438.2015.1104688

*Chen, X., Ramírez, 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

*Cho, J. R., Chiu, M. M., & McBride-Chang, C. (2011). Morphological awareness, phonological awareness, and literacy development in Korean and English: A 2-year longitudinal study. *Scientific Studies of Reading*, *15*, 383–408. https://doi.org/10.1080/10888438.2010.487143

Chow, B. W. Y., McBride-Chang, C., & Burgess, S. (2005). Phonological Processing Skills and Early Reading Abilities in Hong Kong Chinese Kindergarteners Learning to Read English as a Second Language. *Journal of Educational Psychology*, *97*, 81–87. https://doi.org/10.1037/0022-0663.97.1.81

Chung, S. C., Chen, X., & Geva, E. (2019). Deconstructing and reconstructing cross-language transfer in bilingual reading development: An interactive framework. *Journal of Neurolinguistics*, *50*, 149–161. https://doi.org/10.1016/j.jneuroling.2018.01.003

Cooper, H. (2010). *Applied social research methods series: Vol. 2. Research synthesis and meta-analysis: A step-by-step approach (4th ed.)*. Thousand Oaks, CA: Sage Publications.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*, 222–251. https://doi.org/10.3102/00346543049002222

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (Ed.). *Schooling and language minority students: A theoretical framework* (pp. 3–49). Los Angeles, CA: National Dissemination and Assessment Center.

*D'Angelo, N., Hipfner-Boucher, K., & Chen, X. (2017). Predicting growth in English and French vocabulary: The facilitating effects of morphological and cognate awareness. *Developmental Psychology*, *53*, 1242–1255. https://doi.org/10.1037/dev0000326

*Deacon, S. H., Wade-Woolley, L., & Kirby, J. (2007). Crossover: The role of morphological awareness in French immersion children's reading. *Developmental Psychology*, *43*, 732–746. https://doi.org/10.1037/0012-1649.43.3.732

*Deacon, S. H., Wade-Woolley, L., & Kirby, J. R. (2009). Flexibility in young second-language learners: examining the language specificity of orthographic processing. *Journal of Research in Reading*, *32*, 215–229. https://doi.org/10.1111/j.1467-9817.2009.01392.x

Durgunoğlu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology*, *85*, 453–465. https://doi.org/10.1037//0022-0663.85.3.453

*Eviatar, Z., Taha, H., & Shwartz, M. (2018). Metalinguistic awareness and literacy among semitic-bilingual learners: a cross-language perspective. *Reading and Writing*, *31*, 1869–1891. https://doi.org/10.1007/s11145-018-9850-9

*Farran, L. K., Bingham, G. E., & Matthews, M. W. (2012). The relationship between language and reading in bilingual English-Arabic children. *Reading and Writing*, *25*, 2153–2181. https://doi.org/10.1007/s11145-011-9352-5

*Filippini, A. L. (2007). *Effects of a vocabulary-added instructional intervention for at-risk English learners: Is efficient reading instruction more effective?* (Unpublished doctoral dissertation). University of California, Santa Barbara, USA.

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, *84*(1), 74–111. https://doi.org/10.3102/0034654313499616

Genesee, F., Geva, E., Dressler, C., & Kamil, M. (2006). Synthesis: Cross-Linguistic Relationships in working memory, phonological processes, and oral language. In D. August, & T. Shanahan (Eds.). *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth* (pp. 153-174). Mahwah, NJ: Lawrence Erlbaum Associates.

Geva, E., & Ryan, E. B. (1993). Linguistic and cognitive correlates of academic skills in first and second language. *Language Learning*, *43*, 5–42. https://doi.org/10.1111/j.1467-1770.1993.tb00171.x

*Goodwin, A. P. (2010). *Does meaning matter for reading achievement? Untangling the role of phonological recoding and morphological awareness in predicting word decoding, reading vocabulary, and reading comprehension achievement for Spanish-speaking English language learners* (Unpublished doctoral dissertation). University of Miami, USA.

Goodwin, A. P., & Ahn, S. (2010). A meta-analysis of morphological interventions: Effects on literacy achievement of children with literacy difficulties. *Annals of Dyslexia*, *60*, 183–208. https://doi.org/10.1007/s11881-010-0041-x

Goodwin, A. P., & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children. *Scientific Studies of Reading*, *17*, 257–285. https://doi.org/10.1080/10888438.2012.689791

Goodwin, A. P., Huggins, A. C., Carlo, M., Malabonga, V., Kenyon, D., Louguit, M., & August, D. (2012). Development and validation of extract the base: an English derivational morphology test for third through fifth grade monolingual students and Spanish-speaking

English language learners. *Language Testing*, *29*, 265–289.
https://doi.org/10.1177/0265532211419827

*Goodwin, A. P., Huggins, A. C., Carlo, M. S., August, D., & Calderon, M. (2013). Minding morphology: How morphological awareness relates to reading for English language learners. *Reading and Writing*, *26*, 1387–1415.
https://doi.org/10.1007/s11145-012-9412-5

*Gottardo, A., Mirza, A., Koh, P. W., Ferreira, A., & Javier, C. (2018). Unpacking listening comprehension: the role of vocabulary, morphological awareness, and syntactic knowledge in reading comprehension. *Reading and Writing*, *31*, 1741–1764.
https://doi.org/10.1007/s11145-017-9736-2

Hedges, L. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press. https://doi.org/10.1016/C2009-0-03396-0

*Hu, C. F. (2013). Predictors of reading in children with Chinese as a first language: A developmental and cross-linguistic perspective. *Reading and Writing*, *26*, 163–187.
https://doi.org/10.1007/s11145-012-9360-0

*Jeon, E. H. (2011). Contribution of morphological awareness to second-language reading comprehension. *The Modern Language Journal*, *95*, 217–235.
https://doi.org/10.1111/j.1540-4781.2011.01179.x

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

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

Ke, S., & Xiao, F. (2015). Cross-linguistic transfer of morphological awareness between Chinese and English. *Language Awareness*, 24, 355–380. https://doi.org/10.1080/09658416.2015.1114624

Kerlinger, F. N. (1999). *Foundations of behavioral research*. New York: Wadsworth Publishing.

*Kieffer, M. J., & Lesaux, N. K. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Reading and Writing*, *21*, 783–804. https://doi.org/10.1007/s11145-007-9092-8

*Kieffer, M. J., & Lesaux, N. K. (2012). Direct and indirect roles of morphological awareness in the English reading comprehension of native English, Spanish, Filipino, and Vietnamese speakers. *Language Learning*, *62*, 1170–1204. https://doi.org/10.1111/j.1467-9922.2012.00722.x

*Kieffer, M. J., Biancarosa, G., & Mancilla-Martinez, J. (2013). Roles of morphological awareness in the reading comprehension of Spanish-speaking language minority learners: Exploring partial mediation by vocabulary and reading fluency. *Applied Psycholinguistics*, *34*, 697–725. https://doi.org/10.1017/S0142716411000920

Kirby, R. J., & Bowers, P. N. (2017). Morphological instruction and literacy: Binding phonological, orthographic, and semantic features of words. In K. Cain, D. L. Compton & R. K. Parrila (eds.), *Theories of Reading Development* (pp. 437–461). Amsterdam: John Benjamins. https://doi.org/10.1075/swll.15.24kir

Koda, K. (2000). Cross-linguistic variations in L2 morphological awareness. A*pplied Psycholinguistics*, *21*, 297–320. https://doi.org/10.1017/S0142716400003015

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139524841

Koda, K. (2007). Reading and language learning: Cross-linguistic constraints on second language reading development. *Language Learning*, *57*(suppl.1), 1–44. https://doi.org/10.1111/0023-8333.101997010-i1

Koda, K. (2008). Impacts of prior literacy experience on second-language learning to read. In K. Koda & A. M. Zehler (Eds.), *Learning to read across languages: Cross-linguistic relationships in first- and second- language literacy development* (pp.68–96). New York: Routledge. https://doi.org/10.4324/9780203935668-11

Koda, K. & Ke, S. (2018). 一语在二语阅读发展中的促进作用 [L1-induced facilitation in L2 reading development]. *国际汉语教学研究 [Journal of International Chinese Teaching], 18*(2), 40–48.

Koda, K., Lü, C., & Zhang, D. (2014). L1-induced facilitation in biliteracy development in Chinese and English. In X. Chen, Q. Wang, and Y. Luo (Eds), *Reading Development in Reading Development and Difficulties in Monolingual and Bilingual Chinese Children* (pp. 141–170). New York: Springer. https://doi.org/10.1007/978-94-007-7380-6_8

Koda, K., & Miller, R.T. (2018). Cross-linguistic interaction in L2 word meaning inference in English as a foreign language. In H.K. Pae (Ed.), *Writing systems, reading processes, and cross-linguistic influences: Reflections from the Chinese, Japanese and Korean languages* (pp. 293–312). Philadelphia: John Benjamins. https://doi.org/10.1075/bpa.7.14kod

Koda, K., & Reddy, P. (2008). Cross-linguistic transfer in second language reading. *Language Teaching*, *41*, 497–508. https://doi.org/10.1017/S0261444808005211

Kuo, L. -J., & Anderson, R. C. (2006). Morphological awareness and learning to read: A cross-language perspective. *Educational Psychologist*, *41*, 161–180. https://doi.org/10.1207/s15326985ep4103_3

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.

*Lam, K., & Chen, X. (2018). The crossover effects of morphological awareness on vocabulary development among children in French immersion. *Reading and Writing*, *31*, 1893–1921. https://doi.org/10.1007/s11145-017-9809-2

Lee, S. (2011). *The relationship between morphological awareness and literacy outcomes of elementary students: A meta-analysis study* (Unpublished doctoral dissertation. University of Oregon, USA.

Li, S., Shintani, N., & Ellis, R. (2012). Doing meta-analysis in SLA: Practice, choices, and standards. *Contemporary Foreign Language Studies*, *384*, 1–17.

Liu, P. D., Zhou, Y., & McBride-Chang, C. (2010). Morphological awareness: Why and how to link it to Chinese literacy teaching and learning. In J. Chen, C. Wang, & J. Cai (Eds.), *Teaching and learning Chinese: Issues and perspectives* (pp. 237-256). Charlotte, NC: Information Age Publishing.

*Logan, B. (2010). *The role of morphological awareness in the reading development of English language learners* (Unpublished doctoral dissertation). University of Washington, USA.

*Lok, K. Y. Y. (2014). *The crossover effects of morphological awareness on literacy outcomes among children in French immersion* (Unpublished doctoral dissertation). University of Toronto, Canada.

*Luo, Y. C., Chen, X., & Geva, E. (2014). Concurrent and longitudinal cross-linguistic transfer of phonological awareness and morphological awareness in Chinese-English bilingual children. *Written Language & Literacy*, *17*, 89–115. https://doi.org/10.1075/wll.17.1.05luo

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

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

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13–103). New York: American Council on Education.

McBride-Chang, C., Cho, J. R., Liu, H., Wagner, R. K., Shu, H., Zhou, A., ... & Muse, A. (2005). Changing models across cultures: Associations of phonological awareness and morphological structure awareness with vocabulary and word recognition in second graders from Beijing, Hong Kong, Korea, and the United States. *Journal of Experimental Child Psychology*, *92*, 140–160. https://doi.org/10.1016/j.jecp.2005.03.009

Nagy, W. E., Carlisle, J. F., & Goodwin, A. P. (2014). Morphological knowledge and literacy acquisition. *Journal of Learning Disability*, *47*, 3–12. https://doi.org/10.1177/0022219413509967

*Neugebauer, S. R., Kieffer, M. J., & Howard, E. R. (2015). Exploring multidimensionality and mediation in the roles of lexical knowledge in reading comprehension for

Spanish-speaking language minority learners. *Learning and Individual Differences*, *39*, 24–38. https://doi.org/10.1016/j.lindif.2015.03.003

Norris, J. M. & Ortega, L. (Eds.). (2006). *Synthesizing research on language learning and teaching.* Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.13

Nunes, T., Bryant, P., & Bindman, M. (1997). Morphological spelling strategies: developmental stages and processes. *Developmental Psychology*, *33*, 637–649. https://doi.org/10.1037//0012-1649.33.4.637

Pan, J., Song, S., Su, M., McBride, C., Liu, H., Zhang, Y., Li, H., Shu, H. (2016). On the relationship between phonological awareness, morphological awareness and Chinese literacy skills: Evidence from an 8-year longitudinal study. *Developmental Science*, *19*, 982–991. https://doi.org/10.1111/desc.12356

*Park, E. C. (2004). *The relationship between morphological awareness and lexical inference skills for English language learning with Korean first-language background* (Unpublished doctoral dissertation). Carnegie Mellon University, USA.

Plonsky, L., & Oswald, F. L. (2012). How to do a meta-analysis. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 275–295). London: Basil Blackwell. https://doi.org/10.1002/9781444347340.ch14

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

*Ramírez, G., Chen, X., Geva, E., & Kiefer, H. (2010). Morphological awareness in Spanish-speaking English language learners: Within and cross-language effects on word reading. *Reading and Writing*, *23*, 337–358. https://doi.org/10.1017/S0142716411000233

*Ramírez, G., Chen, X., & Pasquarella, A. (2013). Cross-linguistic transfer of morphological awareness in Spanish-speaking English language learners: The facilitating effect of cognate knowledge. *Topics in Language Disorders*, *33*, 73–92. https://doi.org/10.1097/TLD.0b013e318280f55a

Reed, D. K. (2008). A synthesis of morphology interventions and effects on reading outcomes for students in grades K–12. *Learning Disabilities Research & Practice*, *23*, 36–49. https://doi.org/10.1111/j.1540-5826.2007.00261.x

Ruan, Y., Georgiou, G. K., Song, S., Li, Y., & Shu, H. (2018). Does writing system influence the associations between phonological awareness, morphological awareness, and reading? A meta-analysis. *Journal of Educational Psychology*, *110*, 180–202. https://doi.org/10.1037/edu0000216

*Saiegh-Haddad, E., & Geva, E. (2008). Morphological awareness, phonological awareness, and reading in English–Arabic bilingual children. *Reading and Writing*, *21*, 481–504. https://doi.org/10.1007/s11145-007-9074-x

*Schiff, R., & Calif, S. (2007). Role of phonological and morphological awareness in L2 oral word reading. *Language Learning*, *57*, 271–298. https://doi.org/10.1111/j.1467-9922.2007.00409.x

Singson, M., Mahony, D., & Mann, V. (2000). The relation between reading ability and morphological skills: Evidence from derivational suffixes. *Reading and Writing*, *12*, 219–252. https://doi.org/10.1023/A:1008196330239

*Sun, B., & Curdt-Christiansen, X. L. (2016). Morphological awareness and reading development in bilingual English-Chinese children in Singapore. In R.E. Silver, W.D.

Bokhorst-Heng (eds.), *Quadrilingual Education in Singapore* (pp. 83–101). Singapore: Springer. https://doi.org/10.1007/978-981-287-967-7

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, *41*, 363–387. https://doi.org/10.1017/S0272263118000311

Tibi, S., & Kirby, J. R. (2017). Morphological awareness: Construct and predictive validity in Arabic. *Applied Psycholinguistics*, *38*, 1019–1043. https://doi.org/10.1017/S0142716417000029

Tong, X., & McBride-Chang, C. (2010). Chinese-English biscriptal reading: Cognitive component skills across orthographies. *Reading and Writing*, *23*(3–4), 293–310. https://doi.org/10.1007/s11145-009-9211-9

*Tong, X., McBride, C., Ho, C. S. H., Waye, M. M. Y., Chung, K. K. H., Wong, S. W. L., & Chow, B. W. Y. (2018). Within-and cross-language contributions of morphological awareness to word reading and vocabulary in Chinese–English bilingual learners. *Reading and Writing*, *31*, 1765–1786. https://doi.org/10.1007/s11145-017-9812-7

Tyler, A., & Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of Memory and Language*, *28*, 649–667. https://doi.org/10.1016/0749-596X(89)90002-8

Verhoeven, L., & Perfetti, C. (Eds.). (2017). *Learning to read across languages and writing systems*. Cambridge University Press. https://doi.org/10.1017/9781316155752

*Wang, M., Ko, I. Y., & Choi, J. (2009). The importance of morphological awareness in Korean–English biliteracy acquisition. *Contemporary Educational Psychology*, *34*, 132–142. https://doi.org/10.1016/j.cedpsych.2008.12.002

*Wang, M., Yang, C., & Cheng, C. (2009). The contributions of phonology, orthography, and morphology in Chinese-English biliteracy acquisition. *Applied Psycholinguistics*, *20*, 291–314. https://doi.org/10.1017/S0142716409090122

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

Yeh, Y. F., Joshi, R. M., & Ji, X. R. (2015). The development of morphological awareness in Chinese ESL students. *Contemporary Educational Psychology*, *43*, 51–60. https://doi.org/10.1016/j.cedpsych.2015.10.002

Zhang, D. (2013). Linguistic distance effect on cross-linguistic transfer of morphological awareness. *Applied Psycholinguistics*, *34*, 917–942. https://doi.org/10.1017/S0142716412000070

Zhang, D. (2016). Morphology in Malay–English biliteracy acquisition: An intervention study. *International Journal of Bilingual Education and Bilingualism*, *19*, 546–562. https://doi.org/10.1080/13670050.2015.1026873

* Zhang, D., Chin, C. F., & Li, L. (2017). Metalinguistic awareness in bilingual children's word reading: A cross-lagged panel study on cross-linguistic transfer facilitation. *Applied Psycholinguistics*, *38*, 395–426. https://doi.org/10.1017/S0142716416000278

Zhang, D., & Koda, K. (2013). Morphological awareness and reading comprehension in a foreign language: A study of young Chinese EFL learners. *System*, *41*, 901–913. https://doi.org/10.1016/j.system.2013.09.009

*Zhang, D., Koda, K., Leong, C. K., & Pang, E. (2019). Cross-lagged panel analysis of reciprocal effects of morphological processing and reading in Chinese in a multilingual context. *Journal of Research in Reading*, *42*, 58–79. https://doi.org/10.1111/1467-9817.12135

Zhang, J., Anderson, R. C., Li, H., Dong, Q., Wu, X., & Zhang, Y. (2010). Cross-language transfer of insight into the structure of compound words. *Reading and Writing*, *23*, 311–336. https://doi.org/10.1007/s11145-009-9205-7

Zhang, H., & Koda, K. (2018). Vocabulary knowledge and morphological awareness in Chinese as a heritage language (CHL) reading comprehension ability. *Reading and Writing*, *31*, 53–74. https://doi.org/10.1007/s11145-017-9773-x

*Zhou, Y., McBride, C., Leung, J. S. M., Wang, Y., Joshi, M., & Farver, J. (2018). Chinese and English reading-related skills in L1 and L2 Chinese-speaking children in Hong Kong. *Language, Cognition and Neuroscience*, *33*, 300–312. https://doi.org/10.1080/23273798.2017.1342848

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher’s website:

Appendix S1. Study Details and Moderator Coding

Appendix S2. Distribution of Target L2s and Word Formation Rules in the Primary Studies

Appendix S3. MA Measure Coding

Appendix S4. Sensitivity Analyses Results

Appendix S5. Funnel Plots

Appendix S6. Comparing Effects of Study Design (i.e., Correlational Versus Longitudinal)

Table 2

Mean Correlations for the Five Correlations

Relationship	*k*	*r* [95% CI]	*Z*-value (*p*)	Significance test of different (*Q* test) (*p*)	I^2 (%)	τ^2	Adjusted effect estimate after trim and fill (Random-effects model)	Number of trimmed studies
L1MA-L2MA	20	0.30 [0.22, 0.38]	7.14 (<.001)	48.58 (<.001)	60.89	0.02	N/A	0
L1MA-L2 word decoding	19	0.35 [0.25, 0.43]	6.95 (<.001)	73.13 (<.001)	75.39	0.04	N/A	0

L1 MA-L2 reading comprehension	6	0.39 [0.29, 0.48]	7.08 (<.001)	8.39 (.136)	40.43	0.01	0.37[0.28, 0.46]	To left of mean (*k* = 1)
L2MA-L2 word decoding	31	0.45 [0.40, 0.50]	15.94 (<.001)	74.53 (<.001)	60.98	0.02	0.50[0.45,0.54]	To right of mean (*k* = 7)
L2 MA-L2 reading comprehension	17	0.52 [0.46, 0.57]	13.90 (<.001)	57.70 (<.001)	77.27	0.02	0.54[0.49,0.60]	To right of mean (*k* = 3)

Note. N/A, not applicable.

Table 3

Moderator Analysis Results for the Relationship Between L1 and L2 MA

	Moderator variable	Number of correlations (k)	Correlation (r)	95% CI	95%Difference in r (highest-lowest category) CI	Significance test of difference (Q test)
Word formation rules	Compounding	3	0.43	0.13,0.66	0.21	1.94 (p = 0.584)
	Derivation	8	0.30	0.17,0.43		
	Inflection	2	0.22	0.05,0.37		
	Mixed	7	0.26	0.17,0.34		
L1-L2 writing system distance	AP-AB	2	0.05	-0.17,0.26	0.38	5.28 (p = 0.071)
	AP-AP	12	0.28	0.21,0.35		
	M-AP	3	0.43	0.13,0.66		

Age	Kindergarten to grade two	10	0.28	0.16,0.39	0.06	0.44 (*p* = 0.804)
	Grades three to five	5	0.34	0.13,0.53		
	Mixed	5	0.33	0.21,0.43		
MA measurement type	Morphological structure awareness	2	0.55	0.38,0.68	0.33	11.54 (*p* = 0.021)*
	Mixed	4	0.32	0.21,0.41		
	Morphological relatedness	3	0.34	-0.09,0.66		
	Sentence analogy	2	0.22	0.05,0.37		
	Test of morphological structure	7	0.23	0.14,0.32		

Notes. AP, alphabet; AB, abjad; M, morphosyllabary. Three one-sample subgroups for the analysis of the moderator “L1-L2 writing system distance” were excluded; two one-sample subgroups were excluded from the analysis of the moderator "MA measurement type". The exclusion did not alter the *Q* test results. Details can be found in Appendix S4 in the Supporting Information online.

Table 4

Moderator Analysis Results for the Relationship Between L1 MA and L2 Word Decoding

	Moderator variable	Number of correlations (*k*)	Correlation (*r*)	95% CI	95%Difference in *r* (highest-lowest category) CI	Significance test of difference (*Q* test)
Word formation rules	Compounding	4	0.34	0.14,0.52	0.16	3.40 (p = 0.333)
	Derivation	8	0.35	0.16,0.52		
	Inflection	2	0.44	0.30,0.57		
	Mixed	5	0.28	0.19,0.37		
L1-L2 writing system distance	AB-AP	2	0.36	0.13,0.54	0.23	4.66 (p = 0.199)
	AP-AB	2	0.15	-0.02,0.32		
	AP-AP	11	0.38	0.25,0.51		

	M-AP	4	0.32	0.14,0.48		
Age	Kindergarten to grade two	8	0.29	0.19,0.39	0.10	1.93 (*p* = 0.381)
	Grades three to five	6	0.39	0.29, 0.49		
	Mixed	5	0.36	0.04, 0.62		
MA measurement type	Morphological structure awareness	3	0.41	0.30,0.50	0.28	9.27 (*p* = 0.055)
	Mixed	3	0.48	0.11, 0.74		
	Morphological relatedness	2	0.20	0.04, 0.36		
	Sentence analogy	2	0.44	0.30,0.57		
	Test of morphological	6	0.24	0.09,0.38		

	structure					
Word decoding task	Accuracy	15	0.35	0.24,0.45	0.01	0.01 (p = 0.910)
	Efficiency	4	0.34	0.15,0.50		

Notes. AP, alphabet; AB, abjad; M, morphosyllabary. Three one-sample subgroups were excluded from the analysis of the moderator "MA measurement type"; the exclusion altered the Q test result (Q = 24.84, p = 0.001 before exclusion). Details can be found in Appendix S4 in the Supporting Information online.

Table 5

Moderator Analysis Results for the Relationship Between L2 MA and L2 Word Decoding

	Moderator variable	Number of correlations (k)	Correlation (r)	95% CI	95%Difference in r (highest-lowest category) CI	Significance test of difference (Q test)
Word formation rules	Compounding	5	0.45	0.29,0.59	0.08	2.88 (p = 0.411)
	Derivation	16	0.48	0.41,0.54		
	Inflection	5	0.41	0.30,0.51		
	Mixed	5	0.40	0.32,0.47		
Age	Kindergarten to grade two	11	0.39	0.32,0.45	0.11	5.77 (p = 0.056)
	Grades three to five	8	0.48	0.36,0.58		

	Mixed	10	0.50	0.42,0.57		
MA measurement type	Morphological structure awareness	3	0.43	0.17,0.63	0.28	11.07 (p = 0.050)
	Mixed	7	0.46	0.37,0.53		
	Morphological relatedness	2	0.19	0.03,0.35		
	Sentence analogy	2	0.40	0.25,0.54		
	Test of morphological structure	11	0.47	0.40,0.54		
	Wug test	3	0.42	0.25,0.56		
Word decoding task	Accuracy	23	0.45	0.40,0.51	0.01	0.03 (p = 0.869)
	Efficiency	8	0.44	0.34,0.54		

Notes. Two one-sample subgroups for the analysis of the moderator "age" were excluded, which did not change the *Q* test result. Three one-sample subgroups were excluded from the analysis of the moderator "MA measurement type", and the exclusion altered the *Q* test result ($Q = 31.66$, $p < .001$ before exclusion). Details can be found in Appendix S4 in the Supporting Information online.

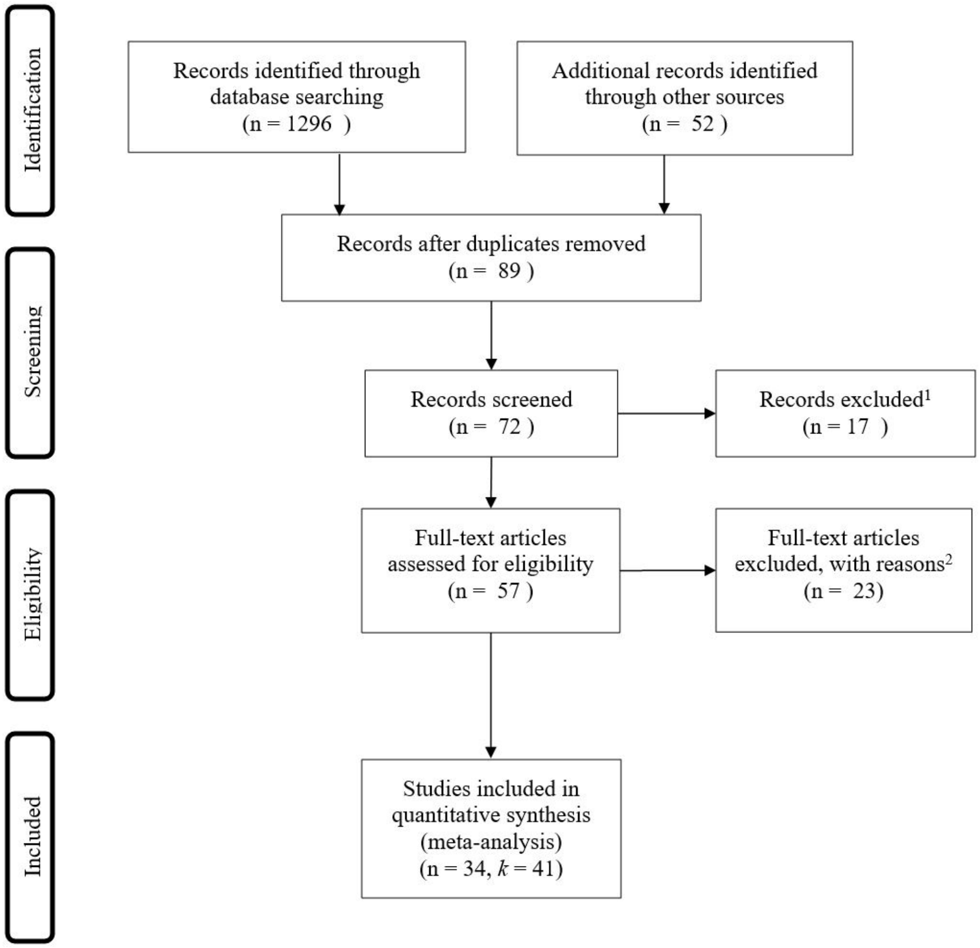
Table 6

Moderator Analysis Results for the Relationships Between L2 MA and L2 Reading Comprehension

	Moderator variable	Number of correlations (k)	Correlation (r)	95%CI	95%Difference in r (highest-lowest category) CI	Significance test of difference (Q test)
Word formation rules	Derivation	11	0.55	0.47,0.62	0.09	2.63 (p = 0.105)
	Mixed	4	0.46	0.38,0.54		
Age	Kindergarten to grade two	3	0.34	0.22, 0.45	0.25	12.31 (p = 0.002)**
	Grades three to five	6	0.59	0.50, 0.66		

	Mixed	6	0.52	0.46, 0.58		
MA measurement type	Mixed	4	0.49	0.42, 0.55	0.05	0.97 (*p* = 0.325)
	Test of morphological structure	11	0.54	0.46, 0.61		
Reading comprehension task	Cloze	3	0.56	0.49, 0.63	0.12	2.30 (*p* = 0.316)
	Multiple-choice	8	0.55	0.45, 0.63		
	Picture matching	2	0.44	0.29, 0.58		

Note. Two one-sample subgroups were excluded from the analysis of the moderator "word formation rules" (Q test results were altered); Two one-sample subgroups were excluded from the analysis of the moderator "age" (Q test results were not altered); two one-sample subgroups were excluded from the analysis of the moderator "MA measurement type" (Q test results were not altered);; and four one-sample subgroups were excluded from the analysis of the moderator "reading comprehension task"(Q test results were not altered). Details can be found in Appendix S4 in the Supporting Information online.



Notes. 1- studies that focused on adult L2 readers only or L2 readers with learning disabilities were excluded. 2-studies that focused on L2 vocabulary as the only outcome measure or did not provide correlational data (even after we contacted the article authors) were excluded.

Figure 1

110x121mm (240 x 240 DPI)

Study name	Correlation	Lower limit	Upper limit	Z-Value	p-Value
D'Angelo et al. (2017) ELs	0.460	0.150	0.688	2.813	0.005
D'Angelo et al. (2017) English only	0.140	-0.179	0.433	0.857	0.391
Deacon et al. (2007)	0.260	0.002	0.486	1.974	0.048
Deacon et al. (2009)	0.180	-0.047	0.390	1.555	0.120
Farran et al. (2012)	-0.040	-0.254	0.177	-0.358	0.720
Lam & Chen (2018) Older	0.130	-0.100	0.347	1.109	0.267
Lam & Chen (2018) Younger	0.320	0.109	0.503	2.929	0.003
Lok (2014) G1	0.310	0.098	0.495	2.831	0.005
Lok (2014) G2	0.140	-0.090	0.356	1.196	0.232
Luo et al. (2014)	0.620	0.475	0.732	6.801	0.000
Park (2004)	0.610	0.254	0.821	3.090	0.002
Ramirez et al. (2010)	0.340	0.151	0.505	3.433	0.001
Ramirez et al. (2013)	0.340	0.143	0.511	3.303	0.001
Saiegh-Haddad & Geva (2008)	0.190	-0.117	0.464	1.216	0.224
Schiff & Calif (2007)	0.440	0.203	0.629	3.470	0.001
Sun & Curdt-Christiansen (2016)	0.320	0.102	0.509	2.834	0.005
Tong et al. (2018)	0.470	0.299	0.612	4.945	0.000
Wang , Ko, & Choi (2009)	0.160	-0.087	0.389	1.271	0.204
Wang, Yang et al. (2009)	0.130	-0.095	0.343	1.132	0.258
Zhang et al. (2017)	0.430	0.279	0.560	5.203	0.000
	0.303	0.223	0.378	7.143	0.000

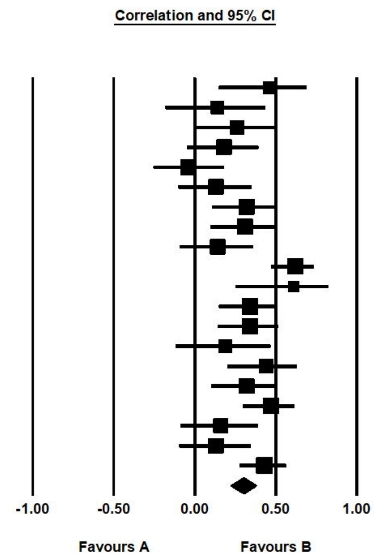


Figure 2

130x80mm (240 x 240 DPI)

Study name	Correlation	Lower limit	Upper limit	Z-Value	p-Value
Cho et al. (2011)	0.490	0.304	0.640	4.734	0.000
Deacon et al. (2007)	0.450	0.217	0.635	3.595	0.000
Deacon et al. (2009)	0.440	0.238	0.605	4.035	0.000
Farran et al. (2012)	0.190	-0.027	0.390	1.720	0.085
Hu (2013)	0.400	0.215	0.557	4.041	0.000
Kahn-Horwitz, Shimron, & Sparks (2005)	0.440	0.298	0.563	5.627	0.000
Lam & Chen (2018) Older	0.200	-0.028	0.408	1.720	0.085
Lam & Chen (2018) Younger	0.290	0.076	0.478	2.637	0.008
Lok (2014) G1	0.290	0.076	0.478	2.637	0.008
Lok (2014) G2	0.200	-0.028	0.408	1.720	0.085
Luo et al. (2014)	0.420	0.234	0.576	4.200	0.000
Ramirez et al. (2010)	0.710	0.595	0.797	8.602	0.000
Ramirez et al. (2013)	0.510	0.339	0.649	5.249	0.000
Saiegh-Haddad & Geva (2008)	0.080	-0.226	0.371	0.507	0.612
Schiff & Calif (2007)	0.220	-0.043	0.455	1.644	0.100
Tong et al. (2018)	0.400	0.218	0.555	4.107	0.000
Wang , Ko, & Choi (2009)	-0.100	-0.336	0.147	-0.790	0.430
Wang, Yang et al. (2009)	0.020	-0.203	0.241	0.173	0.862
Zhang et al. (2017)	0.500	0.359	0.618	6.215	0.000
	0.366	0.322	0.409	15.085	0.000

Statistics for each study

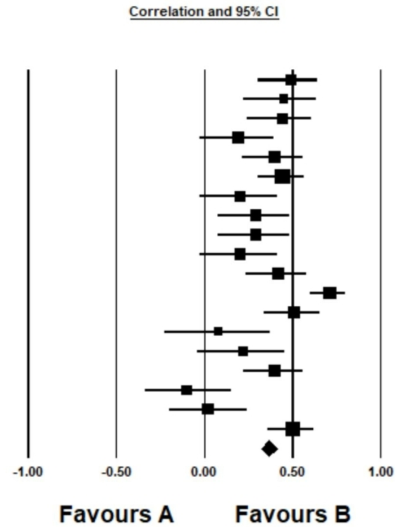


Figure 3

218x128mm (240 x 240 DPI)

Study name	Correlation	Lower limit	Upper limit	Z-Value	p-Value
	Statistics for each study				
Kahn-Horwitz, Shimron, & Sparks (2005)	0.470	0.332	0.588	6.078	0.000
Lok (2014) G1	0.440	0.245	0.601	4.171	0.000
Lok (2014) G2	0.410	0.202	0.583	3.696	0.000
Ramirez et al. (2013)	0.320	0.121	0.494	3.093	0.002
Sun & Curdt-Christiansen (2016)	0.480	0.285	0.637	4.468	0.000
Wang , Ko, & Choi (2009)	0.120	-0.128	0.354	0.949	0.342
	0.396	0.320	0.466	9.489	0.000

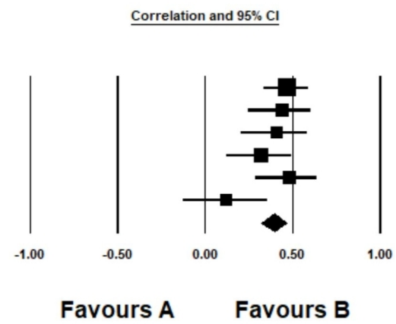


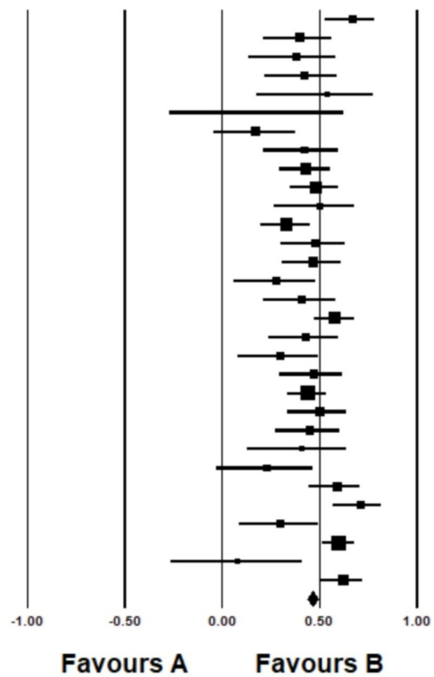
Figure 4

219x83mm (240 x 240 DPI)

Study name	Correlation	Lower limit	Upper limit	Z-Value	p-Value
Chen et al. (2012) (Chinese)	0.670	0.525	0.777	6.974	0.000
Chen et al. (2012) (Spanish)	0.400	0.209	0.561	3.929	0.000
Deacon et al. (2007)	0.380	0.135	0.581	2.967	0.003
Deacon et al. (2009)	0.420	0.215	0.590	3.825	0.000
Eviatar et al. (2018) L1Arabic	0.540	0.175	0.775	2.769	0.006
Eviatar et al. (2018) L1Hebrew	0.220	-0.275	0.623	0.866	0.386
Farran et al. (2012)	0.170	-0.047	0.372	1.535	0.125
Filippini (2007)	0.420	0.207	0.595	3.692	0.000
Goodwin (2010)	0.430	0.289	0.553	5.557	0.000
Goodwin et al. (2013)	0.480	0.350	0.592	6.490	0.000
Gottardo et al. (2018)	0.500	0.263	0.680	3.845	0.000
Jeon (2011)	0.330	0.196	0.452	4.663	0.000
Kieffer & Lesaux (2008) (G4)	0.480	0.300	0.627	4.793	0.000
Kieffer et al. (2013)	0.470	0.302	0.609	5.049	0.000
Lam & Chen (2018) Older	0.280	0.057	0.477	2.441	0.015
Lam & Chen (2018) Younger	0.410	0.210	0.577	3.847	0.000
Logan (2010)	0.580	0.468	0.674	8.380	0.000
Lok (2014) G1	0.430	0.234	0.593	4.062	0.000
Lok (2014) G2	0.300	0.078	0.493	2.626	0.009
Luo et al. (2014)	0.470	0.292	0.616	4.785	0.000
Neugebauer et al. (2015)	0.440	0.334	0.535	7.407	0.000
Ramirez et al. (2010)	0.500	0.334	0.636	5.326	0.000
Ramirez et al. (2013)	0.450	0.268	0.601	4.521	0.000
Saiegh-Haddad & Geva (2008)	0.410	0.125	0.632	2.755	0.006
Schiff & Calif (2007)	0.230	-0.033	0.463	1.721	0.085
Tong et al. (2018)	0.590	0.443	0.706	6.570	0.000
Wang , Ko, & Choi (2009)	0.710	0.564	0.813	6.986	0.000
Wang, Yang et al. (2009)	0.300	0.083	0.490	2.681	0.007
Zhang et al. (2019)	0.600	0.513	0.675	10.783	0.000
Zhou et al. (2018) (EM)	0.080	-0.265	0.407	0.446	0.655
Zhang et al. (2017)	0.620	0.502	0.715	8.202	0.000
	0.464	0.435	0.492	27.014	0.000

Statistics for each study

Correlation and 95% CI



197x157mm (240 x 240 DPI)

Study name	Lower limit	Upper limit	Z-Value	p-Value	Correlation
	Statistics for each study				
Filippini (2007)	0.050	0.482	2.372	0.018	0.280
Goodwin (2010)	0.404	0.637	7.131	0.000	0.530
Goodwin et al. (2013)	0.501	0.700	8.797	0.000	0.610
Gottardo et al. (2018)	0.391	0.750	4.852	0.000	0.600
Jeon (2011)	0.306	0.540	6.255	0.000	0.430
Kieffer & Lesaux (2008) (G4)	0.300	0.627	4.793	0.000	0.480
Kieffer & Lesaux (2012) (Spanish, LL journal)	0.652	0.742	19.316	0.000	0.700
Kieffer & Lesaux (2012) (Tagalog, LL journal)	0.403	0.700	5.755	0.000	0.570
Kieffer et al. (2013)	0.291	0.601	4.923	0.000	0.460
Logan (2010)	0.364	0.599	6.781	0.000	0.490
Lok (2014) G1	0.165	0.544	3.430	0.001	0.370
Lok (2014) G2	0.145	0.543	3.198	0.001	0.360
Neugebauer et al. (2015)	0.412	0.596	8.826	0.000	0.510
Ramirez et al. (2013)	0.339	0.649	5.249	0.000	0.510
Sun & Curdt-Christiansen (2016)	0.334	0.667	4.924	0.000	0.520
Wang , Ko, & Choi (2009)	0.417	0.736	5.458	0.000	0.600
Zhang et al. (2019)	0.445	0.623	9.398	0.000	0.540
	0.455	0.573	13.901	0.000	0.516

Correlation and 95% CI

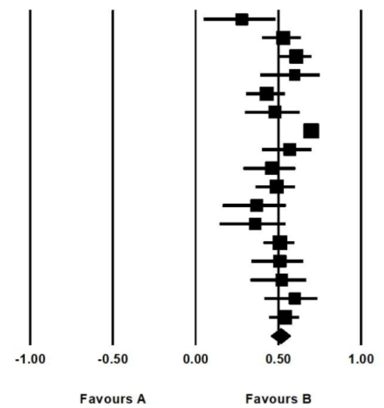


Figure 6

179x95mm (240 x 240 DPI)