Psychological Bulletin
2014, Vol. 140, No. 2, 409–433

0033-2909/14/$12.00 DOI: 10.1037/a0033890

Reading Comprehension and Its Underlying Components in Second-Language Learners: A Meta-Analysis of Studies Comparing First- and Second-Language Learners

Monica Melby-Lervåg and Arne Lervåg
University of Oslo

We report a systematic meta-analytic review of studies comparing reading comprehension and its underlying components (language comprehension, decoding, and phonological awareness) in first- and second-language learners. The review included 82 studies, and 576 effect sizes were calculated for reading comprehension and underlying components. Key findings were that, compared to first-language learners, second-language learners display a medium-sized deficit in reading comprehension (pooled effect size $d = -0.62$), a large deficit in language comprehension (pooled effect size $d = -1.12$), but only small differences in phonological awareness (pooled effect size $d = -0.08$) and decoding (pooled effect size $d = -0.12$). A moderator analysis showed that characteristics related to the type of reading comprehension test reliably explained the variation in the differences in reading comprehension between first- and second-language learners. For language comprehension, studies of samples from low socioeconomic backgrounds and samples where only the first language was used at home generated the largest group differences in favor of first-language learners. Test characteristics and study origin reliably contributed to the variations between the studies of language comprehension. For decoding, Canadian studies showed group differences in favor of second-language learners, whereas the opposite was the case for U.S. studies. Regarding implications, unless specific decoding problems are detected, interventions that aim to ameliorate reading comprehension problems among second-language learners should focus on language comprehension skills.

Keywords: reading comprehension, bilingual development, language comprehension, decoding

Supplemental materials: http://dx.doi.org/10.1037/a0033890.supp

In 2008, 21% (or 10.9 million) of children and youths between five and 17 years of age in the United States spoke a language other than English at home (National Center for Education Statistics, 2011). The number of second-language learners in school has steadily increased over the last decades both in the United States (National Center for Education Statistics, 2011) and in Europe (Organization for Economic Co-operation and Development [OECD] Reviews of Migrant Education, 2009). Given the large number of second-language learners, it is particularly concerning that these students have higher dropout rates and poorer educational outcomes than their monolingual first-language learner counterparts. In 2009, the dropout rate for foreign-born students in the United States was 21%, and the dropout rate for children born to foreign-born parents was 13%. However, the national average was only 8.4% (Child Trends Databank, 2011). The situation is similar in European countries (OECD Reviews of Migrant Education, 2009), where large-scale international comparative studies have shown that second-language learners demonstrate poorer learning outcomes in school than do first-language learners (e.g., Institute for Employment Studies, 2004; OECD, 2004).

A salient predictive factor for educational outcomes in most school subjects is reading comprehension (e.g., OECD, 2000). As the amount of text presented in all school subjects increases with each grade level, children who possess poor reading comprehension skills will struggle academically throughout their education. Poor reading comprehension skills can therefore be an important cause of lower academic success. In this study, we aim to increase our understanding of reading comprehension and its underlying skills in second-language learners when compared to first-language learners. Based on numerous prior studies, decoding (the process of accurately and fluently translating print into spoken words or units), phonological awareness (the ability to manipulate the sounds in spoken words), and language comprehension (the ability to understand the meaning of words and sentences in language) are crucial antecedents for reading comprehension (for a review, see National Institute for Literacy, 2008).

We present a meta-analysis of the differences and similarities between second-language learners and first-language learners in terms of reading comprehension and its underlying skills (i.e., language comprehension, decoding, and phonological awareness). As a background for the meta-analysis, we first present a narrative

This article was published Online First August 12, 2013.

Monica Melby-Lervåg, Department of Special Needs Education, University of Oslo, Oslo, Norway; Arne Lervåg, Department of Educational Research, University of Oslo.

The study was funded by the Norwegian Research Council, Grant Education2020.

Correspondence concerning this article should be addressed to Arne Lervåg, Department of Educational Research, University of Oslo, P.O. Box 1092 Blindern 0317, Oslo, Norway. E-mail: a.o.lervag@ped.uio.no

overview of prior studies and reviews on the typical development of reading comprehension as well as reading comprehension among second-language learners. In the meta-analysis, we first examine reading comprehension and then examine each of the underlying components (i.e., language comprehension, decoding, and phonological awareness). Our overall purpose in the review is twofold. First, we seek to identify important information regarding the areas of strengths and challenges for second-language learners compared with first-language learners. Second, we aim to uncover factors that may explain the performance differences between the two groups and to determine under the conditions in which the two groups perform at a similar level. Detecting factors that may explain this variation is crucial for understanding what affects the reading comprehension level and underlying skills of second-language learners. Taken together, this knowledge is of vital importance when examining theoretical claims in the area of literacy development among second-language learners and when providing effective instruction and targeted interventions for them. Providing effective instruction for the large number of second-language learners is a challenge not only for children learning English as a second language. Studies conducted by the European Union (EU) and the OECD (see EU European Commission's Directorate-General for Education and Culture, 2008; OECD Reviews of Migrant Education, 2009) reveal that challenges in countries and regions such as the Netherlands, France, or Scandinavia related to second-language learners in school are very similar to those demonstrated in Britain and also to some extent the United States. Thus, in our meta-analysis, we include studies in which second-language learners are learning either English or other European languages as their second language.

The Development of Reading Comprehension

Reading Comprehension Development in First-Language Learners

Individual differences in reading comprehension are often understood as the product of decoding and language comprehension skills (Gough & Tunmer, 1986). Decoding refers to the process of translating print into spoken words or units. It is often measured by tests in which the children are asked to decipher a printed word or a nonsense word into a pronounced unit (Hulme & Snowling, 2009). Fluency can be seen as a part of the decoding process and refers to the degree of automatization of the decoding. A fluent reader is able to read orally with speed, accuracy, and proper expression (National Institute of Child Health and Human Development [NICHD], 2000). Language comprehension is the ability to attribute semantic meaning to spoken words, often measured by tests of vocabulary (in this study, vocabulary refers to oral vocabulary and not reading vocabulary), word definitions, or listening comprehension (Gough & Tunmer, 1986). Numerous studies consistently reveal that these two skills explain the majority of individual differences in reading comprehension (for a review, see the National Institute for Literacy, 2008). As noted by Snow and Kim (2007), the area of language comprehension is a large problem space; compared to phonological awareness and decoding, vocabulary acquisition is the more difficult task.

Various studies of first-language learners have shown that the skills underlying the development of reading comprehension begin to evolve in early childhood, long before children receive formal reading instruction in school (for a review, see National Institute for Literacy, 2008). Furthermore, the relative importance of decoding and language comprehension in explaining differences in reading comprehension has been shown to change during the developmental course of schooling. In the early school years, much of the variation in reading comprehension is explained by individual differences in accuracy and the fluency of decoding words and texts. As children become older, their decoding skills are automatized, and more resources are allocated to comprehension (Lervåg & Aukrust, 2010; NICHD, 2000; NICHD Early Child Care Research Network, 2005; Roth, Speece, & Cooper, 2002; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004; Storch & Whitehurst, 2002). After the early primary school years, language comprehension gradually accounts for a larger proportion of individual differences in reading comprehension. For this reason, the sample age is a crucial factor in explaining the results of studies that compare reading comprehension between first- and second-language learners; it will subsequently serve as a moderator in our meta-analysis.

Reading Comprehension Development in Second-Language Learners

Throughout the years, various theoretical claims have been made concerning the nature of reading comprehension development among second-language learners. These theoretical perspectives underlie much extant research and are important for understanding what might moderate the differences between first- and second-language learners. One influential theoretical account is based on Cummins (1979). In this view, because of a common underlying language proficiency, second-language learner status can enhance second-language literacy skills because of the transference of skills from the first language. In addition, the development of second-language skills is moderated by socioeconomic status. Children from a higher socioeconomic background are more likely to use context-independent language at home that corresponds with the schooling language. This expansive use of language will presumably facilitate language transfer and lead to smaller group differences when second-language learners are compared with first-language learners. Thus, socioeconomic background will also serve as a moderator in our meta-analysis.

A second influential theory is based on the notion of contrastive analysis (Connor, 1996; Odlin, 1989). Within this perspective, the first and second languages are analyzed for the purpose of identifying structural (e.g., related to phonology, syntax, or semantics) similarities and differences (Odlin, 1986), which can either facilitate or impede the acquisition of the second language. The degree of structural similarity between the first and second languages may affect the size of the group differences in reading comprehension and underlying skills between first- and second-language learners. Hence, structural similarities between language 1 (L1) and language 2 (L2) constitute an important variable that will be examined as a moderator. This idea can also be applied to the similarities and differences between writing systems. In other words, learning to read a second orthography that is based on the same principles for converting print to sound (e.g., the alphabetic principle) as the first orthography should be easier than learning a second orthography that uses a different principle for the conver-

sion (e.g., idiographic). Therefore, orthography is another potentially important moderator factor.

The third theoretical perspective is the so-called time on task hypothesis (Porter, 1990). Here, the time spent learning the first language may have a negative impact on the learner's second-language skills. Consistent with this view, because learning a new language depends on exposure to that language, an emphasis on the first language at home and at school can negatively affect second-language learning. According to this perspective, first-language skills do not have a positive impact on second-language skills. Consequently, children who use both the first and the second language at home should have better second-language skills than children who use only their first language at home. Likewise, children who are instructed in the second language only at school should have better second-language skills than children who are instructed in both languages. Thus, language used at home and in instruction is a potentially important additional moderator variable.

Previous single studies. Various studies using different research designs have compared reading comprehension skills between first- and second-language learners. In concurrent studies, there is substantial variation in the size of the group differences in reading comprehension. Although some studies show that second-language learners perform at the same level as (or better than) first-language learners on reading comprehension tests (Chiappe, Glaeser, & Ferko, 2007; Lesaux, Rupp, & Siegel, 2007; Verhoeven & Vermeer, 2006), others indicate that second-language learners perform worse than their first-language-learner counterparts (Hannon & McNally, 1986; Kovelman, Baker, & Petitto, 2008; Lervåg & Aukrust, 2010).

Longitudinal studies of reading comprehension development are crucial for understanding the factors that cause this large variation in results. Unfortunately, only a few studies that examine reading comprehension in second-language learners have followed the same group of children across time. In one such study, Lervåg and Aukrust (2010) showed that for both first- and second-language learners (Urdu as L1 and Norwegian as L2), only language comprehension, and not decoding skills, explains the increase in reading comprehension skills from the middle of second to the end of third grade beyond mother's educational level and nonverbal abilities. They also found that language comprehension (vocabulary) was a particularly strong predictor among second-language learners compared to first-language learners. The limitations in language comprehension skills among the second-language learners were sufficient to explain the gap between the two groups in reading comprehension. The differences between the two groups with respect to reading comprehension were large at the onset of the study, and the gap between the two groups increased for both measures of reading comprehension (Woodcock Reading Mastery Tests—Revised and Neale Analysis of Reading Ability) during the study period.

Droop and Verhoeven (2003) demonstrated that for both first- and second-language learners (Moroccan and Turkish as L1, Dutch as L2), decoding and language comprehension explained variation in reading comprehension skills in third grade, but the influence of decoding skills ceased by the end of fourth grade. As for group differences, there was a large gap between the groups in favor of the first-language learners at the onset of the study, but the gap decreased, increased, or remained stable across measures of reading comprehension. Similarly, in a study by Verhoeven (2000) of children starting at age 6, language comprehension had a greater impact on reading comprehension among second-language learners (Mixed L1 languages, Dutch as L2) than among first-language learners. Also, whereas the reading comprehension level of second-language learners was similar to that of first-language learners at the study's onset, the level increased for one measure and decreased for another measure during the period of the study. On some tests, children approached the ceiling on reading comprehension; as a result, few differences could be found.

Moreover, Hutchinson, Whiteley, Smith, and Connors (2003) found that second-language learners (mixed Arabic languages as L1, English as L2) demonstrated poorer reading comprehension skills than first-language learners in second grade. This trend remained stable through the fourth grade. In addition, earlier language comprehension skills were more important for reading comprehension among second-language learners than among first-language learners. Nakamoto, Lindsey, and Manis (2007) showed that phonological processing and language comprehension explained the growth in reading comprehension from the first throughout the sixth grade in a manner consistent with the findings in studies of first-language learners. As for group differences, second-language learners (Spanish as L1, English as L2) started to lag behind beginning in third grade, and this gap increased through fifth grade. Hacquebord (1994) found a similar increase in group differences during secondary school (Turkish as L1; Dutch as L2). Further, Lesaux et al. (2007) found a similar predictive pattern for first- and second-language learners. With respect to group differences, Lesaux et al. found that in fourth grade, differences in reading comprehension between first- and second-language learners (mixed as L1; English as L2) were negligible.

Overall, longitudinal studies of first- and second-language learners confirm the pattern that phonological awareness, decoding, and language comprehension skills are crucial in predicting later reading comprehension, but the majority of studies indicate that language comprehension seems to be even more important for second-language learners than for first-language learners. However, when the size of the differences in reading comprehension between first- and second-language learners and how these differences change over time are considered, results from extant longitudinal studies are inconsistent.

Prior reviews and meta-analyses. Lesaux, Koda, Siegel, and Shanahan (2006) conducted a narrative review of reading comprehension in second-language learners. They concluded that the general tendency is that the second-language learners performed less well on reading comprehension when compared with first-language learners. Lesaux et al. further suggested that the factors that influence reading comprehension among second-language learners generally fall into two categories: contextual, such as the learners' socioeconomic background and the type of reading comprehension test; or individual, such as word-reading skills and background knowledge.

Additionally, a meta-analysis by Melby-Lervåg and Lervåg (2011) examined the cross-linguistic transfer of reading comprehension, language comprehension, decoding, and phonological awareness. For decoding and phonological awareness skills, second-language learners can benefit from the transference of skills used in their first language, thereby reducing the size of the

group difference for these skills between first-language learners and second-language learners. However, for language comprehension, the transfer in skills from the first to the second language is very small.

Development of Language Comprehension Skills

Development of Language Comprehension Skills in First-Language Learners

The acquisition of language comprehension skills is an essential aspect of child development, as these skills are crucial both for individual reasoning and for communicating with others. It has been estimated that high school graduates must know the meaning of approximately 75,000 English words; to accomplish this feat, they will have to learn an average of 10 to 12 words per day between the ages of 2 and 17 (Snow & Kim, 2007). Achieving this goal is clearly a complex and multiple-layered task. In addition, studies have shown that the rank-ordering of children according to their language comprehension skills remains nearly unchanged from the age of 4 to the fourth grade, suggesting that the skills that underlie language learning are formed at an early age (Gathercole, Willis, Emslie & Baddeley, 1992; Lervåg & Aukrust, 2010; Melby-Lervåg, Lervåg, et al., 2012; Storch & Whitehurst, 2002).

It is also well established that being raised in an impoverished environment leads to poorer overall outcomes with respect to language comprehension than being raised in a middle- or upper-class context (Hart & Risley, 1995; for a review, see Hoff, 2006). A significant number of studies have shown that joint book reading and exposure to literature enhance children's language skills (Mol & Bus, 2011) and that children from lower socioeconomic backgrounds are typically less exposed to such experiences (Hoff, 2006). In addition, parents of high socioeconomic status (SES) typically talk more often to their children, use a more elaborate vocabulary, and engage their children more often in context-independent conversation than do parents with lower socioeconomic backgrounds (Hart & Risley, 1995; Hoff, 2006; Pan, Rowe, Singer, & Snow, 2005). Thus, the weight of evidence suggests that SES affects the quality and the quantity of the language to which children are exposed. SES is thus an important variable that may affect the results of studies comparing reading and language comprehension among first- and second-language learners, and it will be used as a moderator variable in the subsequent meta-analysis.

Development of Language Comprehension Skills Among Second-Language Learners

Prior single studies. Although some cross-sectional studies show that second-language learners have language comprehension skills similar to those of first-language learners (Bialystok, Shenfield, & Codd, 2000; D'Angiulli, Siegel, & Sierra, 2001; Westman, Korkman, Mickos, & Byring, 2008), the majority of studies demonstrate that second-language learners have much poorer language comprehension skills than first-language learners (e.g., Droop & Verhoeven, 2003; Scruggs, Mastropieri, & Argulewicz, 1983). This finding is not surprising, given that second-language learners begin at a disadvantage in terms of their language comprehension skills. Even if they can derive benefits when their first language shares semantic features (cognates) with their second language (see Odlin, 1986), second-language learners must develop language comprehension skills at a faster pace if they are to achieve the same level as first-language learners. Therefore, compared to first-language learners, second- language learners often demonstrate a restricted second-language vocabulary.

Longitudinal studies that compare language comprehension between first-language learners and second-language learners show diverging results with respect to the gap between language comprehension skills and the stability or consistency of this difference across time. Verhoeven (2000) showed that the group differences in language comprehension were significant and that they favor first-language learners at the beginning of first grade; however, by the end of second grade, the gap between the two groups had decreased (mixed as L1, Dutch as L2). Notably, according to Verhoeven (2000), this gap may be due to ceiling effects on the measure for first-language learners. Droop and Verhoeven (2003) showed that second-language learners (Moroccan and Turkish as L1, Dutch as L2) start out with much poorer language comprehension skills than first-language learners and that from the beginning of the third grade to end of the fourth grade, the gap between the groups increased. Jean and Geva (2009) found that second-language learners began the fifth grade with less knowledge of word meanings than first-language learners and that this gap remained stable on a measure of receptive vocabulary but increased on a measure of root word meanings in the sixth grade (mixed as L1, English as L2).

Prior reviews and meta-analyses. No prior meta-analyses or narrative reviews have been conducted on language comprehension skills among second-language learners. However, Geva (2006) has conducted a narrative review of the relation between language comprehension, decoding, and reading comprehension among second-language learners. The conclusion was that although English language comprehension plays a significant role in the reading comprehension of second-language learners, multivariate studies suggest that this relation is moderated by contextual factors such as home language use, SES, and instructional experiences.

Development of Decoding Skills

Development of Decoding Skills in First-Language Learners

Developmental studies have shown that decoding is a code-related skill that is heavily influenced by instruction. Overall, decoding and fluency skills rapidly increase after the onset of reading instruction and then level off in the early and middle grades of primary school (Caravolas, Lervåg, Defior, Seidlová-Málková, & Hulme, 2013; Seymour, Aro, & Erskine, 2003).

Studies of first-language learners reveal that proficient decoding skills involve both visual and phonological processing (Seidenberg & McClelland, 1989), in which visual processes activate skills that allow the reader to link the visual symbol (letter) with the correct sound. As for phonological processing, both longitudinal and experimental training studies have shown that, in addition to letter knowledge, phonological skills are a critical precursor for developing efficient decoding skills (e.g., de Jong & van der Leij, 1999; Hulme et al., 2002; Lervåg, Bråten, & Hulme, 2009; Näslund &

Schneider, 1991; National Early Literacy Panel, 2008; Wagner, Torgesen, & Rashotte, 1994). In particular, the awareness of phonemes, rather than larger units such as rhymes or syllables, seems to play a pivotal role in the development of decoding skills (Castles & Coltheart, 2004; Hulme et al., 2002; Macmillan, 2002; Melby-Lervåg, Lyster, & Hulme, 2012). It has been argued that phonological awareness in children progresses from awareness of larger units (rhymes, syllable) to phoneme awareness and that phoneme awareness tasks are more difficult than corresponding tasks with larger units (Carroll, Snowling, Hulme, & Stevenson, 2003; McBride-Chang, 2004).

When considering the differences between alphabetical orthographies and the development of decoding skills, some have argued that phoneme awareness is most critical for learning to read in irregular orthographies, such as English, and that it is of less importance in more regular orthographies (Aro & Wimmer, 2003; Seymour et al., 2003; Share, 2008; Wimmer, 1993). In contrast, others have posited that the predictive relations between phonological skills and reading ability do not show any substantial differences between English and other, more consistent alphabetic orthographies (Caravolas et al., 2012; Caravolas, Volin, & Hulme, 2005; Vaessen et al., 2010; Ziegler et al., 2010). There is also a growing recognition that phonological skills may be involved in learning to read in nonalphabetic ideographic orthographies, such as Chinese (Hanley, 2005; Huang & Hanley, 1995; McBride-Chang et al., 2005). Given the inconclusive results, it is important to examine how the type of orthography (both in the first and second languages) affects the results of studies comparing decoding and phonological awareness skills among first- and second-language learners. Type of orthography will therefore be used as a moderator in the subsequent meta-analysis.

Development of Decoding Skills Among Second-Language Learners

Prior single studies. Numerous studies have compared decoding and phonological awareness skills between first-language learners and second-language learners. Theoretically, it has been hypothesized that because second-language learners have the option of comparing the structure between two languages, they have an advantage in developing metalinguistic awareness compared to first-language learners (see Bialystok, Majumder, & Martin, 2003). However, in studies comparing decoding and phonological awareness between first- and second-language learners, there are significant variations between the size of the group differences for both decoding (e.g., Chiappe et al., 2007; McBride-Chang, Bialystok, Chong, & Li, 2004) and phonological awareness (e.g., Bialystok, Luk, & Kwan, 2005; Kovelman, Baker, & Petitto, 2008).

As for the longitudinal studies that have examined growth in decoding skills among first-language learners compared to second-language learners, Verhoeven (2000) found that the two groups began at a similar level that remained stable throughout the study (mixed as L1, Dutch as L2). A study by Jongejan, Verhoeven, and Siegel (2007) showed similar findings; that is, first- and second-language learners (mixed as L1, English as L2) performed at the same level on both decoding and phonological awareness tasks, and this similarity remained stable from the first to the fourth grade. Jongejan et al. also determined that phonological awareness remains the most important predictor of decoding skills in the third and fourth grades. Similarly, the study by Hutchinson, Whiteley, Smith, and Connors (2004) found that the two groups start out at the same level with respect to phonological awareness and that this level remained stable from second to sixth grade (mixed Asian languages as L1, English as L2). In Droop and Verhoeven's (2003) study, the results were mixed and dependent on the type of decoding test. In general, second-language learners (Turkish or Moroccan as L1, Dutch as L2) were at the same level as first-language learners from low socioeconomic backgrounds at the onset of the study. However, the decoding skills of second-language learners were considerably lower than those of first-language learners from higher socioeconomic backgrounds. Thus, sample age can affect the size of the gap between first- and second-language learners when decoding and phonological awareness are considered, and it will therefore be used as a moderator in the subsequent meta-analysis.

Prior reviews and meta-analyses. Lesaux et al. (2006) conducted a meta-analysis of 10 studies that compared decoding skills among first- and second-language learners. The results showed a minute and insignificant difference in favor of second-language learners ($d = -0.09$). The results of the 10 studies were homogenous, and no analysis examined potential moderators to explain the differences between studies. However, Lesaux et al. concluded based on a narrative review that the process of learning to decode in a second language, as in a first language, is highly influenced by phonological processing. Because of large variations between the studies, they concluded that it was not possible to draw any conclusions regarding group differences in terms of phonological awareness.

A meta-analysis by Adesope, Lavin, Thompson, and Ungerleider (2010) found support for the hypothesis of a metalinguistic advantage for second-language learners. On the basis of 29 studies, they showed that second-language learners performed significantly better than first-language learner controls ($d = 0.33$) in terms of metalinguistic awareness (a broader construct than phonological awareness that included measures of language reasoning and grammatical judgments). There was a wide variation between studies regarding the size of the group differences, but no moderator analysis was conducted exclusively for metalinguistic awareness.

Measurement of Reading Comprehension and Underlying Skills

Measurement type can affect the results of the studies that compare reading comprehension and underlying skills between first- and second-language learners. For reading comprehension, different reading comprehension tests often demonstrate modest intercorrelations, suggesting that the tests are measuring different factors. Keenan, Betjemann, and Olson (2008) showed that tests that used a cloze procedure (i.e., the reader is asked to fill in a missing word in a sentence) relied heavily on decoding skills, whereas tests that use open-ended questions are more dependent on language comprehension skills. The role of decoding has also been related to text length, as tests that used single-sentence or two-sentence passages proved to be more sensitive to decoding skills than tests that used longer passages (Francis et al., 2006; Keenan et al., 2008). Keenan and Betjemann (2006) demonstrated

that a problem with tests using a multiple-choice format was that the child could answer test questions correctly independent of the passage. In an analysis of passage-independent items (in the Gray Oral Reading Test), these items were not sensitive to reading disability, and the learners' performance on such items did not correlate with performances on other reading comprehension tests. The type of test used to measure reading comprehension is thus an important factor that can affect the results of studies that compare reading comprehension between first- and second-language learners, and it will be used as a moderator in the subsequent meta-analysis.

With respect to assessing language comprehension, several measures—including receptive picture vocabulary tests (pointing to the correct picture after being presented a word), word definition tests (defining the meaning of a word), and oral cloze tests (orally filling in a missing word in a sentence)—are commonly used. It has been suggested that picture vocabulary and definition tests rely on different skills, as the picture vocabulary tests depend on the breadth of word knowledge, and word definition tasks depend on the depth of word knowledge (Ordoñez, Carlo, Snow, & Laughlin, 2002). Others, however, have found no conceptual distinction between the two types of tests (Vermeer, 2001). Another important factor is that the alpha reliability of these measure types can differ: Although picture vocabulary tests often demonstrate high reliability, word definition tests tend to show lower reliability (e.g., Lervåg & Aukrust, 2010). In a bivariate relationship, unreliability always attenuates a relationship (Shadish, Cook, & Campbell, 2002), which can mask true group differences between first- and second-language learners. The oral cloze assessment differs from the word definition and picture vocabulary tests, because it often uses a multiple-choice format and the words are presented in a sentence where the meaning of the word can be guessed from the context. Such a format may be easier than word definition and picture vocabulary tests (Pearson, Hiebert, & Kamil, 2007), thus perhaps reducing the differences between first- and second-language learners. Thus, task type will serve as a moderator in the meta-analysis.

Measures of decoding are generally assessed by using either untimed accuracy measures or timed fluency measures based on the reading of real words or nonwords. Fluency measures are often used to avoid ceiling effects in more transparent languages, in which children learn to read at a faster pace than English-speaking children do (Caravolas et al., 2013). When ceiling effects are avoided, fluency and accuracy measures using both words and nonwords tend to be highly reliable and highly correlated (see Lervåg et al., 2009). On this basis, there is little reason to believe that the decoding test type should affect the size of the gap between first- and second-language learners. Still, accuracy versus fluency tests will be used as a moderator for our later analyses. As for measures of phonological awareness, it has been suggested that tests of phoneme awareness (e.g., phoneme deletion) typically demonstrate higher alpha reliability than tests of rhyme awareness (e.g., rhyme detection; Muter, Hulme, Snowling, & Stevenson, 2004). This difference can deflate the size of the group variance between first- and second-language learners regarding rhyme awareness. Thus, test type will subsequently serve as a moderator in the meta-analysis.

The Current Study

Based on the narrative review of prior studies, there are several reasons for which a meta-analysis is necessary at this time. First, no prior meta-analysis has summarized group differences in language comprehension, reading comprehension, and phonological awareness between first- and second-language learners. As the findings from prior single studies are highly inconsistent, a meta-analysis that summarizes the differences in language comprehension skills and examines factors that affect those skills in second-language learners seems crucial. Second, no meta-analyses have systematically tested potential explanations for group differences in reading comprehension and underlying skills between first- and second-language learners. Such an analysis may shed light on factors that may explain the size of the group differences between first- and second-language learners. Also, because single studies examine children of different ages, a merging of the studies in a meta-analysis will provide data that can be used to generate hypotheses concerning developmental relations and the stability of group differences over time. Given the few longitudinal studies in this field and their inconsistent results regarding group differences, using age as a moderating variable can offer important directions for future research. Finally, the meta-analysis of decoding skills by Lesaux et al. (2006) must be updated, given that their analysis concluded in 2002.

Hypotheses

Our meta-analyses examine first- and second-language learners in relation to four different constructs: reading comprehension, language comprehension, decoding, and phonological awareness. First, in our meta-analysis of group differences in reading comprehension, we examined whether the moderators of age, socioeconomic status, home language, instructional language, differences between first and second language, consistency of first-language orthography, and test type could explain differences between the studies with respect to the size of group differences. Second, in our meta-analysis of studies comparing language comprehension between first- and second-language learners, we examined the moderators of age, socioeconomic status, home language, instructional language, language type, and test type. Finally, in our meta-analysis of both decoding and phonological awareness skills, we utilized the moderators of age, socioeconomic status, home language, instructional language, writing system in the first language, consistency of first-language orthography, and test type. For all outcomes, we used nonverbal IQ as a moderator to rule out the possibility that the group differences would be a function of this important factor. Further, we examined whether variables related to methodological quality (year of publication and distributional characteristics) could explain the differences between studies. Year of publication is important because it has been demonstrated that effect sizes in published studies tend to fade and decrease as a function of time (see Ioannidis, 1998; Jennions & Møller, 2001). Distributional characteristics are important, as floor effects among second-language learners could lead to small differences between the groups (related to overly complex tests), and ceiling effects could lead to small differences between the groups because of overly easy tests. Finally, we tested whether study origin (Asia, Australia, Europe, Canada, or the United States) could explain variations in the size of group differences between studies. Indeed,

Antecol, Cobb-Clark, and Trejo (2003) found that "Australian and Canadian immigrants have higher levels of English fluency, education, and income (relative to natives) than do U.S. immigrants" (p. 192). As our introduction suggests, these are issues that might affect the differences between the reading and language skills of first- and second-language learners.

On the basis of theory and prior studies, we identify the following main hypotheses to test in our meta-analyses. In the hypotheses below, group differences refer to differences between first- and second-language learners.

1. As reading comprehension is the product of decoding and language comprehension (Gough & Tunmer, 1986), we expect that the size of the mean effect size will fall between that of the studies pertaining to decoding and language comprehension. The relative importance of decoding and language comprehension skills for reading comprehension changes during the course of development (e.g., NICHD Early Child Care Research Network, 2005). Therefore, we expect that age will be an important moderating variable and that the group differences will be smaller for children in their first years of primary school than the group differences for older children. We also expect that characteristics related to reading-comprehension test type (Francis et al., 2006; Keenan & Betjemann, 2006; Keenan et al., 2008) will be important in explaining why the different studies yielded different results.

2. Because of the complexity of language comprehension (Snow & Kim, 2007) and because second-language learners begin at a disadvantage (with a limited degree of transference of first-language skills; Melby-Lervåg & Lervåg, 2011), we expect that studies examining language comprehension will demonstrate large group differences. We expect that the size of the group differences will be moderated by socioeconomic background (Hart & Risley, 1995; Hoff, 2006; Pan et al., 2005), the degree of exposure to the second language at home and in school (Odlin, 1989; Porter, 1990), and the extent to which their first and second languages share cognates. Because the child can receive contextual support from a sentence in oral cloze assessments, we expect this test to generate smaller group differences than studies that use tests with no contextual support (i.e., word definitions and picture vocabulary).

3. In the area of phonological awareness and decoding, because such skills are easily taught and are sensitive to transference from the first language (Melby-Lervåg & Lervåg, 2011; Snow & Kim, 2007), there will be small group differences that may possibly favor the second-language learners (Adesope et al., 2010). We further expect that group differences will be moderated by socioeconomic status (Hart & Risley, 1995; Hoff, 2006; Pan et al., 2005) and the degree of exposure to the second language at home and in school (Odlin, 1996; Porter, 1990). We also expect that second-language learners with an ideographic first language will have poorer phonological awareness and decoding skills than will those second-language learners who have an alphabetic first language (McBride-Chang et al., 2005). We also expect that group differences for phonological awareness will be moderated by the type of test used, as phoneme-awareness tasks are presumably more difficult than tasks using larger units (Castles & Coltheart, 2004; Hulme et al., 2002; Macmillan, 2002; Melby-Lervåg, Lyster, & Hulme, 2012).

Method

To ensure its methodological quality, our meta-analysis was designed and reported to be consistent with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations (www.prisma-statement.org).

Literature Search, Inclusion Criteria, and Study Coding

The literature search and inclusion criteria are shown in Figure 1. When we selected studies for the meta-analysis, "second-language learners" were operationally defined as children/youths who either use or study two languages. In addition, the child/youth must be exposed to each language either regularly at home with at least one parent or in school for at least 4 hours per day. Control groups with monolingual first-language learners were defined as samples consisting of children who spoke only one language at home, which had to be the same as the instructional language.

Multiple methods were used to obtain a sample of relevant studies. The electronic database search was conducted by investigators under the supervision of librarians. Searches were developed from the keywords *bilingual**, *L2 learners, second-language learners, English language learners (ELL), English second language (ESL), English additional language (EAL), language minority, limited English proficient (LEP), limited English speaking* and *multilingual** paired with *phon** *awareness, vocabulary, language comprehension and reading, decoding,* and *word attack.* Search limits included publications in English from 1965 to May 10, 2013. Abstracts for peer-reviewed studies, non-peer-reviewed studies, book chapters, dissertations, conference proceedings, and reports were also examined. All issues of *International Journal of Bilingual Education and Bilingualism, Bilingualism, TESOL Quarterly,* and *International Journal of Bilingualism* after 1980 were hand searched for relevant papers. Finally, authors who were represented by more than three independent studies in the meta-analysis were contacted by e-mail and asked for unpublished or in-press material.

The target constructs in this study were reading comprehension, language comprehension, phonological awareness, and decoding. For each of these constructs, criteria were established to determine the types of measures that represented each. The criteria established for the indicators of each construct were broad, and a broad range of tests were judged as valid indicators for the target constructs to increase the power of the overall analysis. Because the criteria for the indicators were broad, the differences between test types for each construct were also examined.

To be considered a measure of reading comprehension, studies in which a child read a passage or sentence and answered questions in relation to the text were included.

To be considered a measure of language comprehension, tests that aim to measure expressive or receptive vocabulary by means of pictures, oral cloze, or listening comprehension were included. The reason we used this broad language comprehension construct that also included listening comprehension was to increase the power of the meta-analysis. Treating vocabulary and listening comprehension as a single construct is also supported in a latent-variable study by Lervåg (2010), which showed that after measurement errors were taken into account by using latent variables,

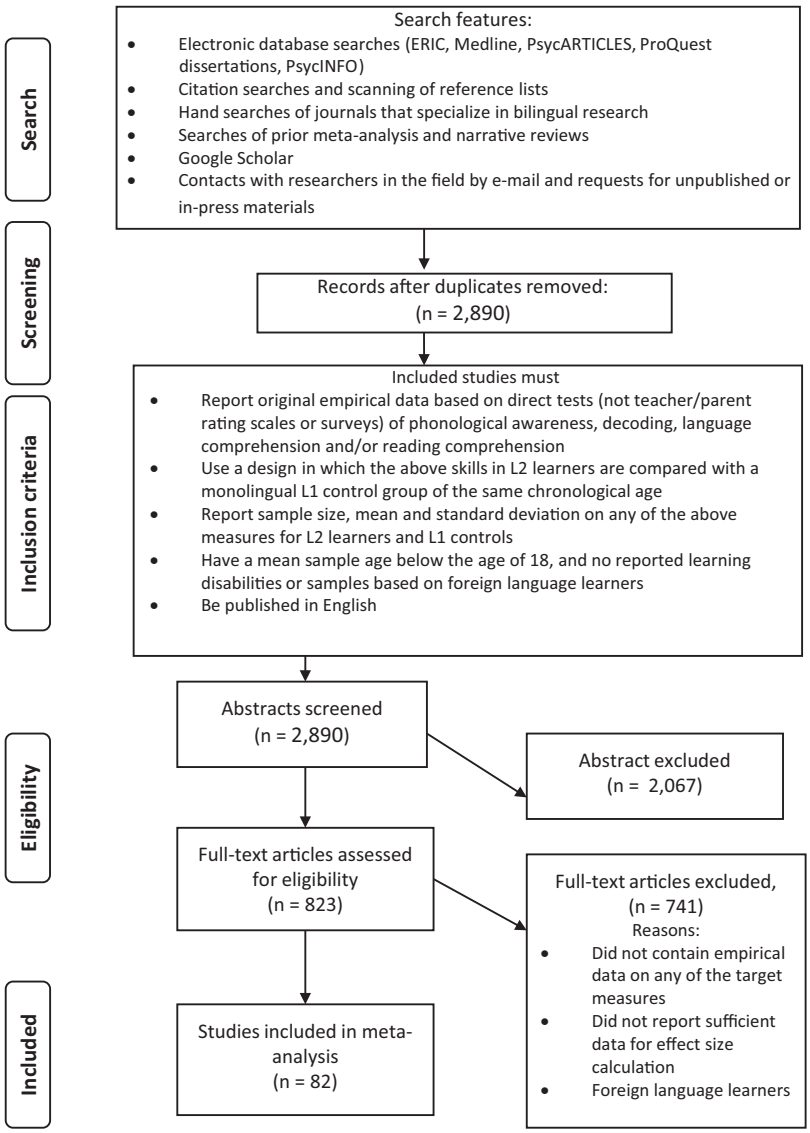


Figure 1. Flow diagram for the search and inclusion of studies. L1 = first language; L2 = second language.

listening comprehension was highly related to expressive and receptive vocabulary (in 7-year-olds). Furthermore, they all loaded on the same factor. To be considered as a measure of phonological awareness, the task must involve deletion, blending, counting, segmentation, generation, judgment, position analysis or replacement of phoneme, onset, rhymes, and/or syllables in words. To be considered as a decoding measure, the test should comprise reading fluency and/or reading accuracy of words, nonwords, sentence decoding, or passage decoding.

The abstracts from all search types were printed and judged according to relevance, and papers that seemingly met the criteria for inclusion based on the abstract were examined to determine whether sufficient statistics for an effect-size calculation were presented and to decide whether all inclusion criteria were met. This process resulted in the coding of a total of 82 studies with 160 independent group comparisons that included 15,137 second-language learners and 111,418 monolingual first-language learners.

Violating the assumption of independence by computing an overall effect size based on information from the same sample more than once can lead to incorrect estimates (Hunter & Schmidt, 1990). Thus, several considerations were made before the studies were coded. First, studies from the same author were examined to detect duplicate samples. When it was not possible to determine whether samples were dependent, independence was assumed. For longitudinal studies, information from only one time point was coded. Because of attrition, the first time point usually provides the largest sample and was therefore the preferred sample for coding. An exception to this practice was if the longitudinal study began before children had received formal reading instruction. In such cases, the time point at which the children were first measured on reading and reading-related measures was coded. For experimental

studies, only pretest data prior to any intervention were coded. In the analysis, each construct (i.e., reading comprehension, language comprehension, phonological awareness, and decoding) was analyzed separately, and the overall effect sizes were estimated across each type of construct.

With respect to independence, special considerations were made concerning the coding of the measures for each of the four constructs, given that some studies reported multiple measures for each of these constructs. Therefore, one indicator was coded for each construct based on an established set of guidelines. For reading comprehension, individual tests were coded before group tests, and open-ended tests were coded before multiple choice tests. For language comprehension, picture vocabulary tests were coded before other measurement types. For phonological awareness, phoneme-based measures were coded before other types of phonological awareness measures, such as awareness for larger units (i.e., rhymes or syllables) or composite scores. If a study reported several phoneme-based measures, phoneme deletion was chosen. For decoding, real-word reading was coded before nonword reading, and single word decoding was coded before passage reading.

The coding was conducted by the authors and one assistant. Independent double coding was used for a random sample of 30% of the studies. Before coding, the coder was trained in the procedures and the criteria. The coder was a full-time employed research assistant, with a master's degree in education, who was trained in meta-analyses. Intercoder correlation (Pearson's) for the main outcomes (i.e., reading comprehension, language comprehension, phonological awareness, and decoding) was $r = .99$ with an agreement rate of 89%. Also, intercoder correlation for continuous moderator variables was $r = .97$ with an agreement rate of 90%. Cohen's kappa was used for categorical moderator variables and was $K = .93$. Disagreements were resolved through discussion or by consulting the original paper.

Moderator Variables

We conducted a broad coding of a large number of moderators that could potentially be important for explaining variations between studies. In addition to the moderators used in the analysis (listed below), the age of second-language acquisition, length of residence in the host country for children and parents, parental second-language fluency, and motivational aspects were coded as moderator variables. However, the impact of these variables could not be analyzed, as too few studies (<five) reported data on any of these variables. As for methodological quality, publication status, sampling method, and alpha reliability were coded. However, none of these variables could be used as moderators, because only five studies (with different outcomes) reported information about alpha reliability. Also, despite special efforts to locate unpublished literature, only three studies that fulfilled the inclusion criteria could be classified as such. Furthermore, because the vast majority of studies used convenience sampling, this factor could not be used as a moderator. Finally, the mean and standard deviation of language comprehension skills in the first language and the mean and standard deviation for language comprehension skills in the second language were coded as moderators. Because of uncertainty and a lack of information in the original papers as to whether the first and second language tests used psychometrically comparable scales, it was not meaningful to calculate an effect size for differences in first- and second-language competence in the second-language learners.

Age. The mean ages of the second-language learners and the first-language control children were coded. Studies that reported information regarding only age range and in which the age range exceeded 2 years were excluded from the age moderator analysis. In cases where the study reported age within the range of 1 year, the median in years was coded. When studies reported age according to grade level, the median year that corresponded to the reported grade was coded.

Nonverbal IQ. The means and standard deviations for nonverbal IQ reported for each group in the original paper were coded.

Orthographic regularity. The degree of regularity between letter–sound relationships was used as a moderator, and languages were separated into two categories based on the degree: regular orthographies or irregular (English).

Language differences. The differences between the first and second language among second-language learners were coded into two categories: (a) Indo-European first language/Indo-European second language and (b) non-Indo-European first language/Indo-European second language. Indo-European languages are broadly a family of related languages that share cognates and a common origin. The Indo-European languages include most languages of Europe, the Middle East, and India and are distinct from a number of unrelated language families that predominate elsewhere in the world (Crystal, 1997).

Writing system. The writing system was coded into two categories: alphabetic writing system or ideographic writing system.

Instructional language. The language of instruction was coded into two categories: (a) instruction in the second language and (b) instruction in both first and second languages.

Home language. Home language was coded into two categories: (a) first language was the only language used with parents and (b) use of first language was with one parent only; use of second language was with the other parent.

Socioeconomic status. The information on the SES of second-language learners was separated into four categories: high, low, middle, or mixed. The coding of SES was based on information reported in the papers concerning family or neighborhood income and/or educational level. In the analysis, due to a small number of studies in some of the categories, three categories were used: (a) middle/high, (b) low, and (c) mixed.

Measure type. The test used to measure reading comprehension, language comprehension, phonological awareness, and decoding was coded.

Sample location. The location of the study—whether Asia, Australia, Europe, the United States, or Canada—was coded.

Methodological quality. The year of publication and the ratio between the standard deviation and the mean for each outcome measure were coded as possible indicators of methodological quality. The ratio between the standard deviation and the mean (coefficient of variability) was calculated for each study by dividing the standard deviation by the mean and multiplying by 100. This calculation expresses the standard deviation as a percentage of the mean. This moderator was used, as there were indications of non-normal distributions. If the standard deviation was lower than 15% or higher than 75% of the mean, this was coded as an indication of a non-normal distribution. We would have preferred

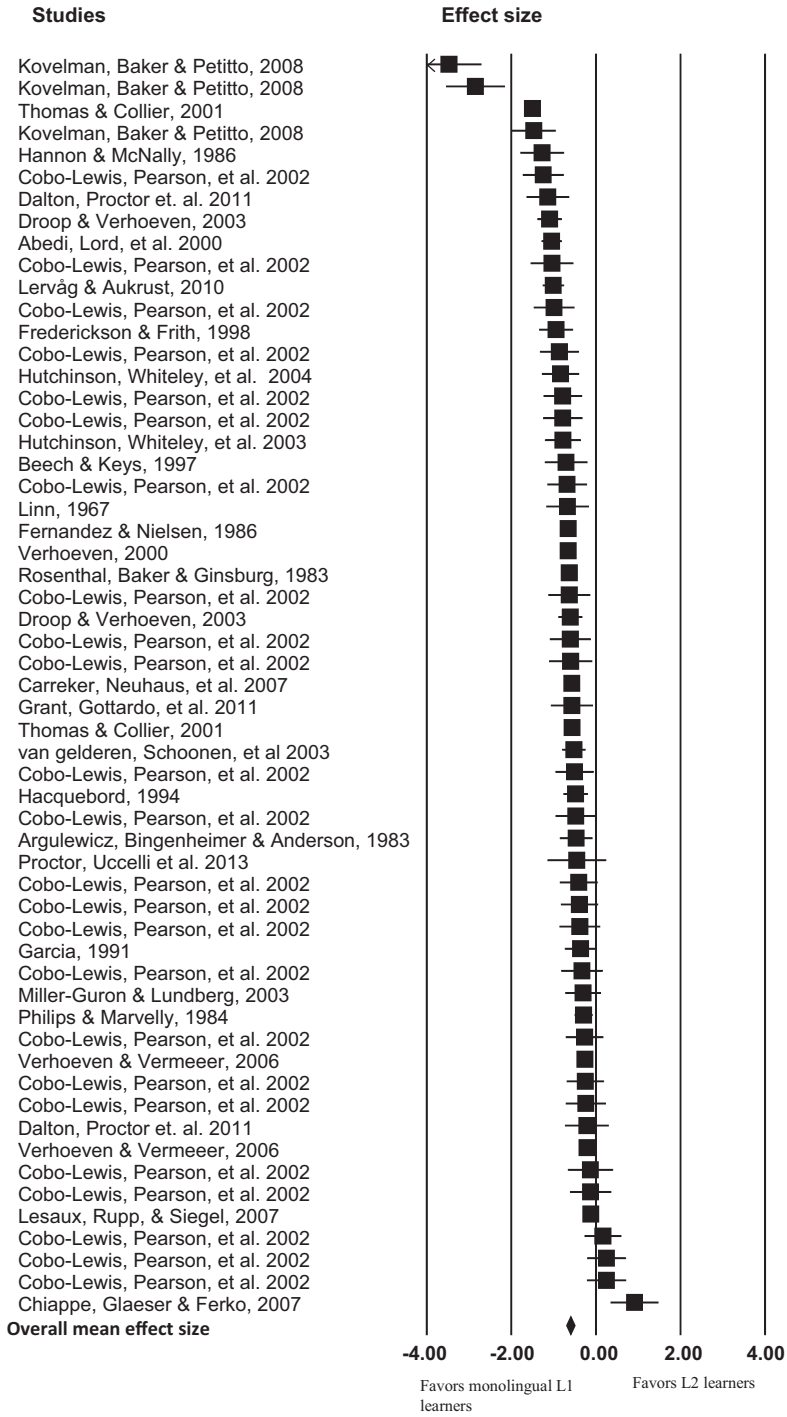


Figure 2. Forest plot of overall average effect size for group differences in reading comprehension between second-language learners and monolingual first-language learners (Cohen's *d*, displayed by ♦ with confidence intervals represented by horizontal lines) and effect sizes with confidence intervals for each study (Cohen's *d*, displayed by ■ with confidence intervals represented by horizontal lines). L1 = first language; L2 = second language.

to analyze the mean/standard deviation ratio as a continuous variable, but due to non-normality of the distribution this procedure was not possible. Although this was not ideal from a methodological point of view (see Preacher, Rucker, MacCallum, & Nicewander, 2005, for a discussion), the categorization used a cutoff based on the distribution of the mean/standard deviation ratio across the studies. Year of publication was also used as an indicator of methodological quality and a potential source of bias. It has been

Table 1
Number of Effect Sizes, Effect Size, 95% Confidence Interval (CI), Heterogeneity Statistics, Differences in d *Between Categories (With Significance Test), and* p *Values for Moderators of Reading Comprehension Differences Between First- and Second-Language Learners*

Moderator variable	Number of effect sizes (k)	Effect size (d)	95% CI	Heterogeneity (I^2)	Difference in d (highest − lowest category)	Significance test of differences between categories (Q test)
Socioeconomic status						
Low	25	−0.73**	[−0.88, −0.57]	83.98**		
High/middle	12	−0.53**	[−0.78, −0.27]	72.00**	0.20	.19
Instructional language						
Both L2 and L1	12	−0.47**	[−0.71, −0.23]	66.93**		
L2	34	−0.52**	[−0.62, −0.40]	77.79**	0.05	.74
Home language						
Both	15	−0.53**	[−0.70, −0.36]	47.23*		
L1	18	−0.76**	[−1.08, −0.44]	90.54**	0.23	.21
Orthography L2						
Irregular	48	−0.63**	[−0.74, −0.52]	87.69**		
Regular	9	−0.57**	[−0.77, −0.37]	85.25**	0.06	.60
Task type in test						
Cloze	34	−0.58**	[−0.78, −0.40]	85.06**		
Single questions to text	15	−0.78**	[−0.93, −0.61]	87.10**		
Multiple choice questions	5	−0.38**	[−0.60, −0.17]	67.23**	0.40	.01**
Text type in test						
Sentence	28	−0.43**	[−0.60, −0.27]	70.19**		
Passage	26	−0.78**	[−0.94, −0.62]	91.47**	0.35	.005**
Study origin						
Europe	15	−0.63**	[−0.79, −0.47]	80.86**		
United States	40	−0.63**	[−0.75, −0.50]	87.26**	0	.95
Distribution						
Studies with floor or ceiling effects	30	−0.47**	[−0.59, −0.35]	90.07**		
Other studies	27	−0.76**	[−0.90, −0.62]	61.98**	0.31	.003**

Note. d = the effect size for subsets of studies belonging to different categories of the moderator variable; k = number of studies; I^2 = the proportion of total variation between the effect sizes that are caused by real heterogeneity rather than by chance; L1 = first language; L2 = second language.
* $p < .05$. ** $p < .01$.

demonstrated that the effect sizes in published studies tend to fade and decrease as a function of time (Ioannidis, 1998; Jennions & Møller, 2001). This trend is mainly due to publication bias: Studies with large effect sizes tend to be published more easily and will be published first, whereas studies with smaller or zero effect sizes take longer to be published and will be published later.

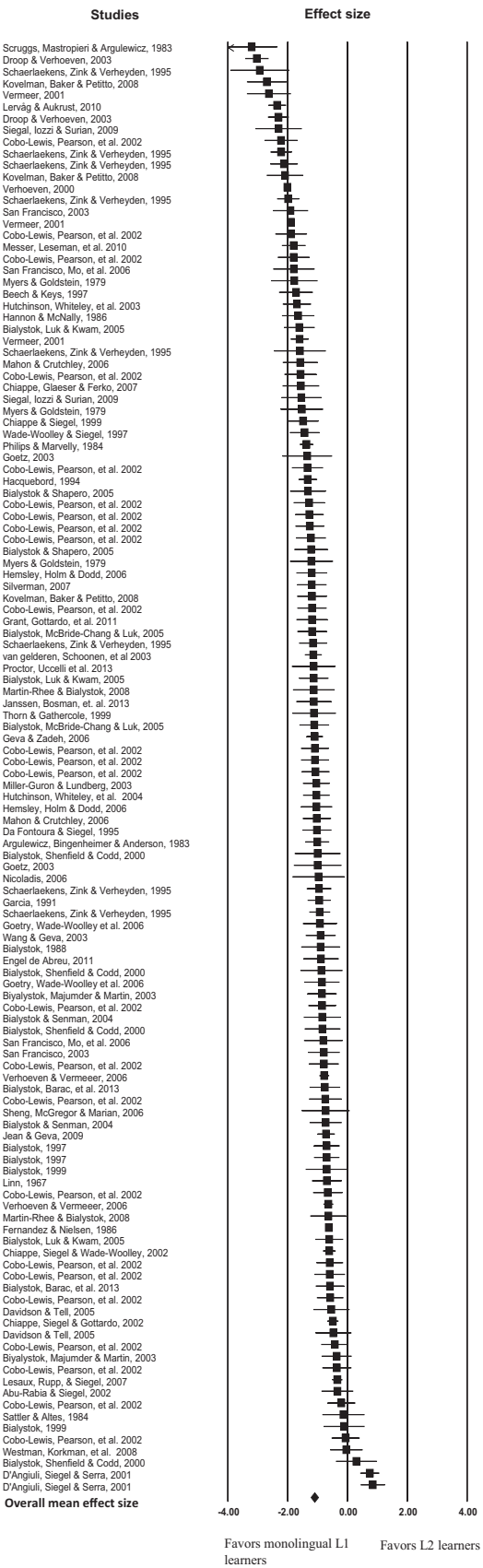
Meta-Analytic Procedures

Effect size and heterogeneity. The majority of analyses were conducted with Comprehensive Meta-Analysis software (Borenstein, Hedges, Higgins, & Rothstein, 2005). The analytic procedures included the following steps. First, the effect sizes for the studies entailing group comparisons were computed separately by means of Cohen's d based on Hedges' formula (Hedges, 1981). We used this calculation because it is corrected for sample size, and, therefore, unlike other effect size estimates, it does not tend to be upwardly biased for small samples. When Cohen's d is expressed in positive terms, the second-language learners have better performance on the test (i.e., a higher group mean) than the monolingual children. A 95% confidence interval was calculated for each effect size to examine whether it was larger than zero. If the confidence interval does not cross zero, the effect is statistically significant.

The overall effect size was estimated by calculating a weighted average of the effect sizes for each outcome construct. The computation of an overall effect size was based on a random effects model, which rests on the assumption that variations between studies can be systematic and are therefore not only due to random error as in the fixed-effect model. Whether the overall effect size differed from zero was tested with a z test, and a sensitivity analysis was used to determine the impact from outliers. A sensitivity analysis allows for an adjusted overall effect size to be estimated after removing studies, one by one, when extreme effect sizes are detected.

The Q test of homogeneity was used to examine the variation in effect sizes between studies (Hedges & Olkin, 1985); I squared (I^2) was used to determine the magnitude of heterogeneity. I^2 is the proportion of the total variation between the effect sizes that are caused by real heterogeneity rather than by chance (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Moderator variables. In all analyses of moderator variables, as when estimating an overall effect size, random effects models were used. For the continuous moderator variables, meta-regression based on a method of moments regression analyses for random effects models is used to predict study outcomes from the moderator variables. In a random effects regression analysis based on the method of moments (also known as the DerSimonian and



Laird method) weights are assigned to each study by estimating variance between studies on the basis of the sum of within-study variance and between-studies variance (Borenstein et al., 2009). Thus, this procedure will lead to each study being weighted more evenly in the regression analysis, and the apparent difference in weighting between small studies versus large studies that is present in a fixed effect model will be less tangible. Because we expect true differences between studies, this method of regression will be a more plausible model than just assigning weights to studies on the basis of their sample size. The meta-regression was conducted with macros developed for SPSS (Lipsey & Wilson, 2001; Wilson, 2006). To determine the strength of the predictors on the study's outcome, a percentage of between-study variance explained (R^2) was used as an effect size.

For the categorical moderator variables, a Q test was used to test the effect size differences between subgroups of studies belonging to different categories in a moderator variable. Unfortunately, the number of studies was not sufficient to undertake multivariate analysis of the categorical moderator variables. We therefore report results based on multiple one-way significance tests (Q tests). Because these significance tests are based on parts of the same data set, this increases the possibility for making Type I errors (i.e., conclude that there is a significant difference between subsets of studies when there are none; see Pigott, 2012, for discussion). In order to deal with the elevated level of Type I errors due to multiple significance tests, we therefore emphasize the degree of overlap in confidence intervals between the categories of each moderator variable rather than significance tests. The degree of overlap between the confidence intervals for the mean effect size of each category will yield the same information as a significance test but is not affected with problems related to multiple comparisons (Pigott, 2012; Valentine, Pigott, & Rothstein, 2010).

For the categorical moderator analysis, if the omnibus test was significant and there were three or more categories, we conducted post hoc pairwise comparisons. Studies were separated in subsets based on the categories of the moderator variable. The degree of differences between the subsets of studies were determined by comparing the overlap between confidence intervals for effect sizes generated for each subset of studies and by comparing the size of Cohen's d between the study subsets.

Publication bias. A funnel plot was used to determine the presence of publication bias. In the funnel plot, sample size is plotted on the y-axis, and effect size is plotted on the x-axis. In the absence of retrieval bias, this plot is expected to form an inverted funnel. In the presence of bias, the funnel will be asymmetric. To detect publication bias, one examines funnel plots for all analyses presented. The trim and fill (Duval & Tweedie, 2000) method was used to examine the impact from

Figure 3. Forest plot of overall average effect size for group differences in language comprehension between second-language learners and monolingual first-language learners (Cohen's *d,* displayed by ♦ with confidence intervals represented by horizontal lines) and effect sizes with confidence intervals for each study (Cohen's *d,* displayed by ■ with confidence intervals represented by horizontal lines). L1 = first language; L2 = second language.

Table 2
Number of Effect Sizes, Effect Size, 95% Confidence Interval (CI), Heterogeneity Statistics, Differences in d *Between Categories (With Significance Test), and* p *Values for Moderators of Language Comprehension Differences Between First- and Second-Language Learners*

Moderator variable	Number of effect sizes (k)	Effect size (d)	95% CI	Heterogeneity (I^2)	Difference in d (highest − lowest category)	Significance test of differences between categories (Q test)
Socioeconomic status						
Low	42	−1.31**	[−1.55, −1.07]	95.03**		
High/middle	31	−0.82**	[−1.08, −0.57]	86.81**	0.49	.007**
Instructional language						
Both L2 and L1	30	−0.95**	[−.1.26, −0.65]	89.09**		
L2	78	−1.20**	[−1.34, −1.04]	91.61**	0.25	.15
Home language						
Both	36	−0.76**	[−0.97, −0.56]	84.41**		
L1	45	−1.47**	[−1.69, −1.25]	91.39**	0.71**	.0001**
Language comprehension test type						
Expressive vocabulary test	11	−1.31**	[−1.63, −0.99]	87.55**		
Oral cloze	10	−0.55*	[−0.96, −0.14]	93.56**		
Receptive picture vocabulary	89	−1.14**	[−1.28, −1.00]	87.07**	0.76**	.01**
Language type						
Indo-European L1 and L2	68	−1.09**	[−1.24, −0.93]	88.78**		
Non-Indo-European L1/ Indo-European L2	43	−1.24**	[−1.44, −1.02]	91.90**	0.15	.24
Study origin						
Canada	33	−0.73**	[−0.91, −0.55]	83.27**		
United Kingdom	13	−1.33**	[−1.51, −1.15]	39.16		
Other European countries (non-English L2)	25	−1.59**	[−1.86, −1.32]	94.71**		
United States	48	−1.10**	[−1.26 −0.95]	82.75**	0.86	.0001**
Distribution						
Studies with floor or ceiling effects	35	−0.95**	[−1.23 −0.64]	92.68**		
Other studies	89	−1.19**	[−1.32, −1.05]	91.30**	0.24	.13

Note. d = the effect size for subsets of studies belonging to different categories of the moderator variable; k = number of studies; I^2 = the proportion of total variation between the effect sizes that are caused by real heterogeneity rather than by chance; L1 = first language; L2 = second language.
* $p < .05$. ** $p < .01$.

possible missing studies. The trim and fill method imputes values in the funnel plot to render it symmetrical and then calculates an estimated overall effect size.

Missing data. During coding, numerous instances of missing data became apparent. If the data that were missing were critical for calculating the effect size for the main outcomes, we contacted the author. However, this approach was usually unsuccessful. If data were missing for moderator variables, the study was excluded from the moderator analysis for which data were missing but was included in all moderator analyses for which data were provided.

Results

Characteristics for each study included in the meta-analysis are presented in Table S1 in the online supplemental materials. A correlation matrix of all outcomes and continuous moderator variables is shown in Table S2 (see supplemental materials). The group differences between the first- and second-language samples in reading comprehension, language comprehension, phonological awareness, and decoding with confidence intervals (CI), overall and for each study, are shown in Figures 2 to 5. The results for the categorical moderator variables are presented in Tables 1 to 4. The tables show the number of studies in each category of the moderator variable, the effect sizes with 95% CI for the subsets of studies in each category, and the differences between the highest and lowest categories with significance test.

Reading Comprehension

A total of 57 independent effect sizes, including 6,464 second-language learners (mean sample size = 113.40, SD = 268.64, range = 17–1,876) and 33,534 monolingual first-language learners (mean sample size = 588.32, SD = 2,278.98, range = 11–13,436), examined the differences in reading comprehension between the two groups. As demonstrated in Figure 2, the overall mean effect size was moderate in favor of the monolingual first-language learners, $d = -0.62$, 95% CI [−0.71, −0.52], and significant, $z(56) = -12.43$, $p < .01$. The effect sizes varied between $d = -3.47$ and 0.91, and this variation was significant and large, $Q(56) = 416.45$, $p < .01$, $I^2 = 86.55$. A sensitivity analysis showed that after removal of outliers, the overall effect size was within the range of $d = -0.64$, 95% CI [−0.73, −0.54] to $d = -0.59$, 95% CI [−0.68, −0.50]. As for publication bias, the funnel plot indicated that studies to the left of the mean were missing (i.e.,

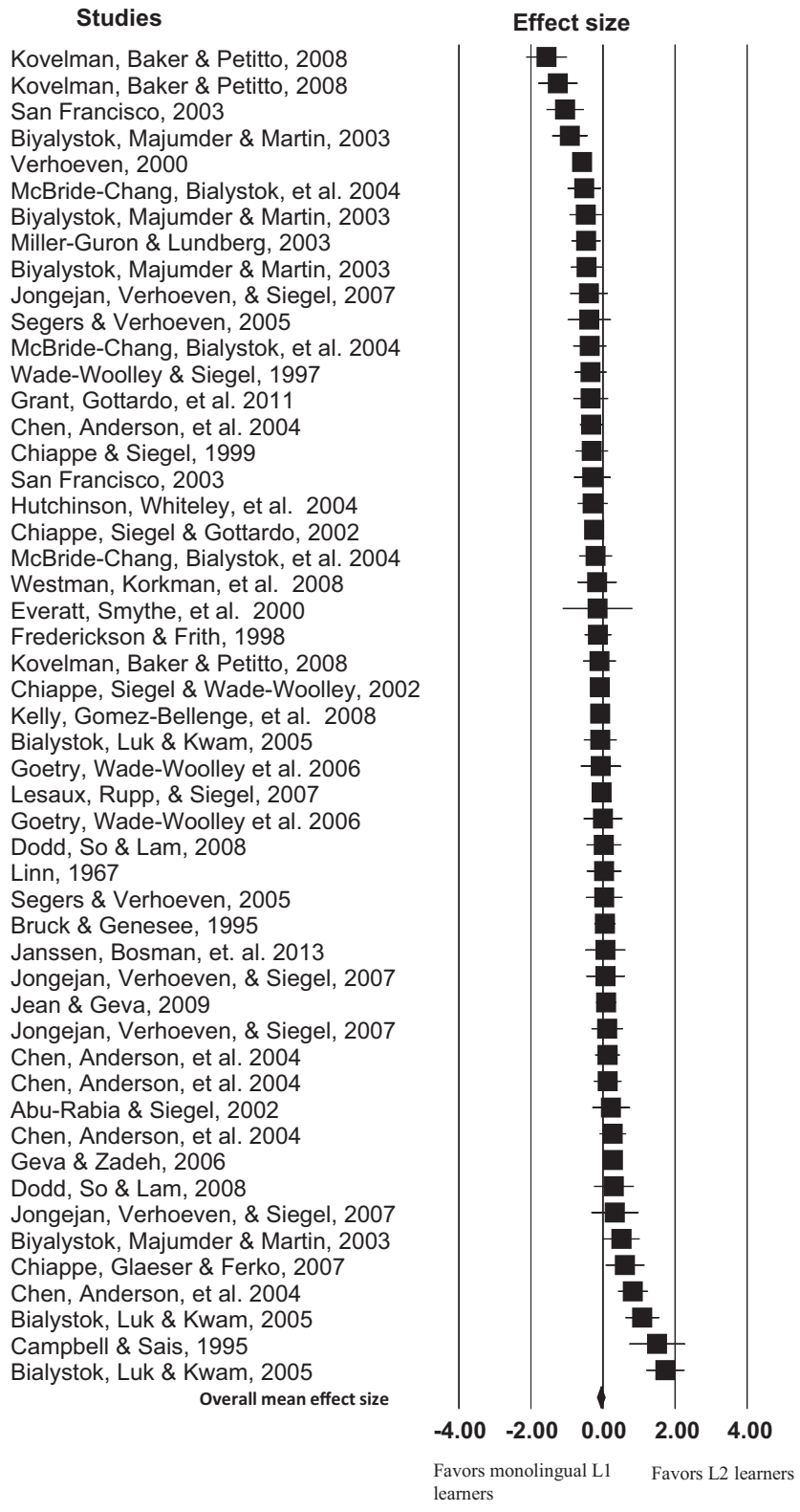


Figure 4. Forest plot of overall average effect size for group differences in phonological awareness between second-language learners and monolingual first-language learners (Cohen's *d*, displayed by ♦ with confidence intervals represented by horizontal lines) and effect sizes with confidence intervals for each study (Cohen's *d*, displayed by ■ with confidence intervals represented by horizontal lines). L1 = first language; L2 = second language.

Table 3
Number of Effect Sizes, Effect Size, 95% Confidence Interval (CI), Heterogeneity Statistics, Differences in d *Between Categories (With Significance Test), and* p *Values for Moderators of Phonological Awareness Differences Between L1 and L2 Learners*

Moderator variable	Number of effect sizes (k)	Effect size (d)	95% CI	Heterogeneity (I^2)	Difference in d (highest − lowest category)	Significance test of differences between categories (Q test)
Socioeconomic status						
Low	11	−0.26*	[−0.51, −0.02]	91.41**		
High/middle	9	−0.23*	[−0.46, −0.02]	62.88**	0.03	.84
Instructional language						
Both L2 and L1	5	0.24	[0.28, −0.80]	94.55**		
L2	34	−0.01	[−0.15, 0.13]	78.71**	0.25	.64
Home language						
Both	8	−0.26	[−0.50, 0.01]	61.72**		
L1	16	−0.01	[−0.31, 0.28]	86.80**	0.25	.22
Phonological awareness test						
Rhyme/syllable test	8	0.29	[−0.05, 0.63]	76.00**		
Phoneme test	43	−0.14*	[−0.25, −0.02]	82.52**	0.15	.02*
Writing system L1						
Alphabetic	25	−0.07	[−0.26, 0.12]	89.13**		
Ideographic	12	−0.05	[−0.30, 0.19]	73.52**	0.02	.97
Orthography L2						
Irregular (English L2)	36	−0.09	[−0.22, 0.04]	81.08**		
Regular (non-English L2)	8	−0.29*	[−0.51, −0.07]	51.50	0.20	.12
Study origin						
Asia	8	0.10	[−0.18, 0.38]	72.11**		
Canada	23	0.00	[−0.17, 0.19]	78.84**		
Europe	12	−0.13	[−0.40, 0.14]	74.78**		
United States	8	−0.44*	[−0.85, −0.03]	88.52**	0.44	.15
Distribution						
Studies with floor or ceiling effects	5	0.31	[−0.22, 0.84]	80.97**		
Other studies	46	−0.09	[−0.22, 0.04]	90.73**	0.40	.12

Note. d = the effect size for subsets of studies belonging to different categories of the moderator variable; k = number of studies; I^2 = the proportion of total variation between the effect sizes that are caused by real heterogeneity rather than by chance; L1 = first language; L2 = second language.
* $p < .05$. ** $p < .01$.

studies in which second-language learners compared with first-language learners had poorer reading comprehension skills than the overall average effect size). In a trim and fill analysis, four studies were imputed to the left of the mean, and the adjusted overall effect size was $d = -0.67$, 95% CI [−0.77, −0.58].

Table 1 shows the results of the moderator analysis. As shown in Table 1, variables related to text type in tests reliably explained the variations between the studies, and tests using passage reading yielded significantly higher effect sizes than tests using sentence reading. The confidence intervals for the two text types are nonoverlapping. Similarly, task type reliably explained the variations between the studies, as there were no overlapping confidence intervals for multiple choice tests and for single open-ended question. In short, open-ended question tests yielded a higher mean effect size than multiple choice tests. Also, as shown in Table 1, the difference between studies with floor or ceiling effects and studies with more normal distributions was significant, with nonoverlapping confidence intervals. Those studies with floor or ceiling effects in their reading comprehension measure had attenuated sizes of the difference between groups.

Language comprehension reliably explained variance in reading comprehension, $\beta = 0.34$, $p < .01$, $k = 46$, $R^2 = .12$. However, there were four outliers that demonstrated a large deficit in language comprehension for second-language learners, but these outliers showed a moderate to large advantage in favor of the second-language learners on reading comprehension. Presumably, these studies are outliers because they examine children who are in the very beginning stages of developing reading skills. After removal of these outliers, language comprehension explained 30% of the variation in reading comprehension, $\beta = 0.55$, $p < .01$, $k = 42$, $R^2 = .30$. When age and language comprehension were analyzed together as predictors of reading comprehension, age was significant, $\beta = -0.23$, $p = .03$, $k = 45$. The total R^2 for the two variables was 0.17, which means that the group differences in reading comprehension between first- and second-language learners (in favor of the first-language learners) decreases as a function of sample age. Notably, the influence of language comprehension on reading comprehension increases as children get older. Decoding significantly explains variations in reading comprehension skills, $\beta = 0.49$, $p < .01$, $k = 41$, $R^2 = .25$. When age was combined with decoding skills, age was not a significant predictor, $\beta = -0.15$, $p = .21$, $k = 41$. When language comprehension and decoding skills were entered together (after removal of the four outliers), the total R^2 for explaining variations in reading comprehension skills was 0.46; for decoding, $\beta = 0.36$, $p < .01$; and for language comprehension, $\beta = 0.52$, $p < .01$. The moderator variable of age entered alone did not reliably explain variances in effect sizes across studies, $\beta = -0.16$, $p = .10$, $k = 55$, $R^2 = .03$.

As for methodological quality, no significant impact was detected related to publication year, $\beta = -0.08$, $p = .55$, $k = 57$, $R^2 = .01$.

Language Comprehension

A total of 124 independent effect sizes consisting of 7,973 second-language learners (mean sample size = 64.29, *SD* = 181.19, range = 8–1,876) and 23,345 monolingual first-language learners (mean sample size = 188.27, *SD* = 1,210.29, range = 10–13,436) examined the differences in language comprehension between the two groups. As shown in Figure 3, the overall mean effect size was large in favor of first-language learners, $d = -1.12$, 95% CI [−1.24, −1.00], and significant, $z(123) = -18.12$, $p < .01$. The effect sizes varied between $d = -3.20$ and 0.85, and this variation was significant and large, $Q(123) = 1{,}609.93$, $p < .01$, $I^2 = 92.36$. A sensitivity analysis shows that after removing outliers, the overall effect size was within the range of $d = -1.14$, 95% CI [−1.26, −1.02] to $d = -1.10$, 95% CI [−1.22, −0.99]. As for publication bias, the funnel plot indicated that studies on the left side of the mean were missing (i.e., studies in which second-language learners, compared with first-language learners, had language comprehension skills poorer than the overall average effect size). In a trim and fill analysis, 21 studies were imputed on the left side of the mean, and the adjusted overall effect size was $d = -1.31$, 95% CI [−1.45, −1.19].

Table 2 shows the results for the categorical moderator variables. As shown in Table 2, socioeconomic status was a significant moderator variable and children from low socioeconomic backgrounds tend to have poorer language comprehension skills than do their monolingual first-language peers and children from middle and high socioeconomic backgrounds. The home language was also found to be a significant moderator variable, and second-language learners who used the first language exclusively at home tend to have poorer language comprehension skills than second-language learners who used both first language and second language at home. Also, the differences between language comprehension test type was significant, but here all the confidence intervals were overlapping. Finally, study origin was a significant variable, and pairwise comparisons showed that studies conducted in Canada tended to both demonstrate smaller group differences than did the U.S. studies, $Q(1) = 9.40$, $p < .01$; European studies from non-English countries, $Q(1) = 26.33$, $p < .0001$; and studies from the United Kingdom, $Q(1) = 21.02$, $p < .0001$. European studies from non-English countries also showed significantly higher mean effect sizes than the U.S. studies, $Q(1) = 9.14$, $p < .002$. Studies from the United States showed smaller group differences than studies from the United Kingdom, but this difference did not reach significance, $Q(1) = 9.14$, $p = .06$.

In addition, we conducted a meta-regression to examine whether group differences are affected by the age of the sample. It was determined that age had no significant impact on effect size, $\beta = 0.14$, $p = .15$, $k = 124$, $R^2 = .02$. Thus, the group differences in language comprehension between monolingual first-language learners and second-language learners were stable across the age range (3.5 to 15.5 years). Similarly, nonverbal IQ did not reliably explain variations between studies, $\beta = 0.41$, $p = .12$, $k = 13$, $R^2 = .17$. As for methodological quality, no significant impact was detected according to publication year, $\beta = 0.06$, $p = .53$, $k = 127$, $R^2 = .00$.

Phonological Awareness

A total of 51 independent effect sizes comprising 7,053 second-language learners (mean sample size = 138.29, *SD* = 618.24, range = 7–4,494) and 77,987 monolingual first-language learners (mean sample size = 1,529.16, *SD* = 10,071.15, range = 7–72,716) examined differences in phonological awareness between the two groups. As displayed in Figure 4, the overall mean effect size was small, $d = -0.08$, 95% CI [−0.18, 0.03], and was not significant, $z(50) = -1.32$, $p = .19$. The effect sizes varied between $d = -1.57$ and 1.73, and this variation is significant and large, $Q(50) = 283.75$, $p < .01$, $I^2 = 82.37$. A sensitivity analysis shows that after removal of outliers, the overall effect size was within the range of $d = -0.11$, 95% CI [−0.21, −0.01] to $d = -0.05$, 95% CI [−0.16, 0.05]. As for publication bias, the funnel plot indicated that studies on the right side of the mean were missing (i.e., studies in which second-language learners compared with first-language learners had better phonological awareness skills than the overall average effect size). In a trim and fill analysis, 11 studies were imputed on the right side of the mean, and the adjusted overall effect size was $d = 0.11$, 95% CI [−0.02, 0.21].

Table 3 shows the analysis of the moderator variables. As seen in Table 3, the only moderator variable that showed a reliable difference between categories on the *Q* test was phonological awareness test type, but here the confidence intervals were overlapping. The impact of age on effect size was analyzed (age range = 5.0 to 15.0), and the results showed that this impact could not reliably explain variations in study outcomes, $\beta = -0.13$, $p = .22$, $k = 50$, $R^2 = .02$. As for methodological quality, no significant impact was detected due to publication year, $\beta = -0.32$, $p = .12$, $k = 51$, $R^2 = .10$.

Decoding

A total of 79 independent effect sizes comprised of 8,452 second-language learners (mean sample size = 106.99, *SD* = 501.28, range = 4–4,494) and 79,860 monolingual first-language learners (mean sample size = 1,010.89, *SD* = 8,122.2, range = 7–72,716) examined differences in decoding between the two groups. As demonstrated in Figure 5, the overall mean effect size was small in favor of first-language learners, $d = -0.12$, 95% CI [−0.22, −0.02], but significant, $z(78) = -2.40$, $p = .02$. The effect sizes varied between $d = -2.85$ and 3.38, and this variation was significant and large, $Q(78) = 587.63$, $p < .01$, $I^2 = 86.73$. A sensitivity analysis shows that after removing outliers, the overall effect size was within the range of $d = -0.15$, 95% CI [−0.24, −0.06] and $d = -0.09$, 95% CI [−0.19, 0.004]. As for publication bias, the funnel plot indicated that studies on the right side of the mean were missing (i.e., studies in which second-language learners compared with first-language learners had better decoding skills than the overall average effect size). In a trim and fill analysis, 27 studies were imputed to the right of the mean, and the adjusted overall effect size was $d = 0.14$, 95% CI [0.04, 0.25].

Table 4 shows the results for the analysis of moderator variables. As apparent from Table 4, the origin of the study was a

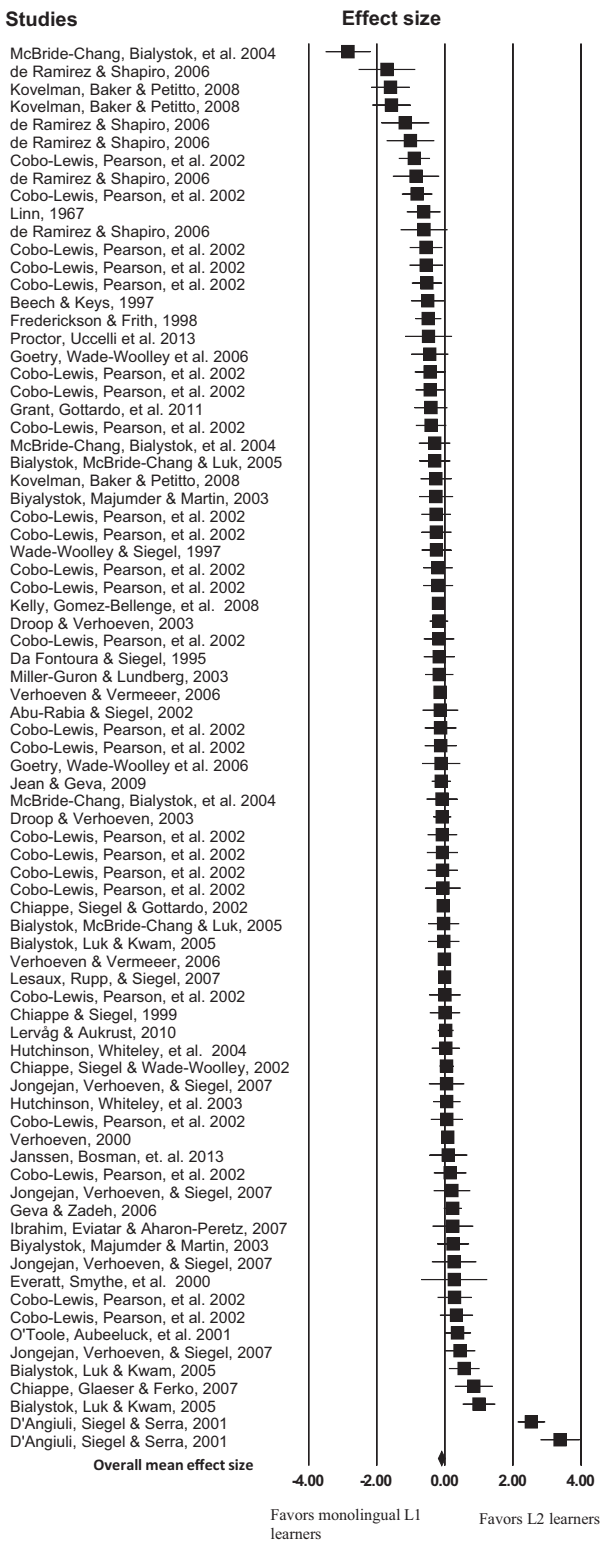


Figure 5. Forest plot of overall average effect size for group differences in decoding between second-language learners and monolingual first-language learners (Cohen's *d,* displayed by ♦ with confidence intervals represented by horizontal lines) and effect sizes with confidence intervals for each study (Cohen's *d,* displayed by ■ with confidence intervals represented by horizontal lines). L1 = first language; L2 = second language.

significant moderator variable: Studies conducted in Canada reliably showed group differences in favor of second-language learners, whereas the same was not true for either the U.S. or the European studies. Pairwise comparisons showed that there was a significant difference between Canadian studies and U.S. studies, $Q(1) = 17.01$, $p < .001$, and between Canadian and European studies, $Q(1) = 4.76$, $p = .03$. The European studies also demonstrated significantly smaller group differences in favor of first-language learners than did the U.S. studies, $Q(1) = 14.44$, $p = .001$. There was no overlap between the confidence intervals. The difference between low and middle/high socioeconomic background was also significant, but here the confidence intervals were overlapping. The impact of age on effect size was analyzed (age range = 5.0 to 15.0 years), and the results showed that this could not explain variations in study outcomes, $\beta = 0.003$, $p = .97$, $k = 75$, $R^2 = .00$. As for methodological quality, no significant impact was detected based on the publication year, $\beta = -0.01$, $p = .90$, $k = 79$, $R^2 = .00$.

Discussion

Our empirical review of reading comprehension skills and their underlying components revealed a number of critical findings concerning the differences and factors that moderate the differences between first- and second-language learners. First, first-language learners demonstrated moderately better reading comprehension skills than did second-language learners. Both language comprehension and decoding skills moderated this relationship. Good language and decoding skills were associated with good reading comprehension skills, and the impact of language comprehension on reading comprehension increased with age. Furthermore, the differences between first- and second-language learners in terms of reading comprehension were also moderated by test-specific characteristics: For second-language learners, answering single open-ended questions is more difficult than answering multiple choice or cloze questions. Moreover, answering questions from passages was more difficult than answering questions from single sentences for second-language learners. Finally, floor and ceiling effects on the reading comprehension tests were associated with attenuated effect sizes of the differences between first- and the second-language learners.

Regarding the components of reading comprehension skills, first-language learners had better oral language comprehension skills than second-language learners. This difference was greater for children from low than from middle or high SES families. The difference was also greater for children who spoke only their first language at home. Finally, there were less-pronounced differences between first- and second-language learners in Canada than in other Western countries.

Overall, second-language learners were slightly worse on decoding skills than were second-language learners. However, second-language learners had better decoding skills than first-language learners in Canada, although they had poorer decoding skills than first-language learners in the USA. There were no reliable differences between the first- and second-language learners on phonological awareness tasks.

Table 4
Number of Effect Sizes, Effect Size, 95% Confidence Interval (CI), Heterogeneity Statistics, Differences in d *Between Categories (With Significance Test), and* p *Values for Moderators of Decoding Differences Between First- and Second-Language Learners*

Moderator variable	Number of effect sizes (k)	Effect size (d)	95% CI	Heterogeneity (I^2)	Difference in d (highest − lowest category)	Significance test of differences between categories (Q test)
Socioeconomic status						
Low	32	−0.30**	[−0.43, −0.19]	78.84**		
High/middle	19	0.14	[−0.26, 0.55]	93.69**	0.44	.04*
Instructional language						
Both L2 and L1	22	−0.15	[−0.65, 0.35]	94.99**		
L2	47	−0.06	[−0.13, 0.02]	40.37**	.09	.72
Home language						
Both	21	0.18	[−0.19, 0.54]	92.87**		
L1	32	−0.16**	[−0.31, −0.01]	74.04**	0.34	.09
Writing system L1						
Alphabetic	58	−0.13	[−0.28, 0.03]	88.06**		
Ideographic	7	−0.45	[−1.06, 0.16]	90.40**	0.34	.31
Orthography L2						
Irregular (English L2)	70	−0.13*	[−0.26, −0.01]	87.78*		
Regular (non-English L2)	9	−0.03	[−0.12, 0.05]	17.63	0.10	.19
Test type						
Reading accuracy test	58	−0.23*	[−0.31, −0.14]	69.79**		
Reading fluency test	11	−0.27*	[−0.48, −0.06]	78.52**	0.04	.69
Study origin						
Canada	25	0.26*	[0.00, 0.57]	91.82**		
Europe	17	−0.05	[−0.14, 0.04]	33.95		
United States	36	−0.36**	[−0.49, −0.23]	72.66**	0.62	.0001**
Distribution						
Studies with floor or ceiling effects	22	−0.14*	[−0.27, −0.02]	90.42**	0.01	.79
Other studies	57	−0.13	[−0.22, 0.04]	34.16		

Note. d = the effect size for subsets of studies belonging to different categories of the moderator variable; k = number of studies; I^2 = the proportion of total variation between the effect sizes that are caused by real heterogeneity rather than by chance; L1 = first language; L2 = second language.
*$p < .05$. **$p < .01$.

Reading Comprehension

In accordance with our hypothesis, both language comprehension and decoding skills moderated the degree to which the reading comprehension skills of the second-language learner samples lagged behind those of first-language-learner samples. Samples with small differences between first- and second-language learners in language comprehension and decoding skills showed comparable small differences between reading comprehension skills. This finding is in accordance both with single studies and with the simple view of reading that sees reading comprehension as the product of language comprehension and decoding skills. Unexpectedly, SES did not moderate the differences between first- and second-language learners' reading comprehension skills. At first glance, this finding is in disagreement with Cummins' notion that children of high SES are more likely to be exposed to a rich language that is less dependent on the here-and-now context, which in turn should facilitate the transference of skills to the second language and lead to better second-language skill acquisition (Cummins, 1979). Such facilitation could then be expected to decrease the gap between first- and second-language learners. Still, SES does moderate one of the underlying components of reading comprehension skills (language comprehension): The gap between first- and second-language learners was smaller for children of middle and high SES. Given this finding, the inability of SES to moderate reading comprehension differences between first- and second-language learners becomes even more surprising. However, a closer look at the 38 independent group comparisons of reading comprehension and SES skills reveal that 24 of the comparisons emanate from the study by Cobo-Lewis, Pearson, Eilers, and Umbel (2002). In this study, all of the children came from the Spanish-speaking population in Florida, and one may question whether there are features specific to this study that led to this unexpected result. It is not entirely clear how the children in Cobo-Lewis et al. (2002) are separated into SES categories, but when the mother's language skills are examined, it appears that there were small differences between children of low and middle/high SES. Such small differences may explain why SES does not seem to affect reading comprehension. Notably, the reading comprehension test used in Cobo-Lewis et al. is an oral cloze test based on sentence reading. As noted, this type of reading comprehension test tends to be more sensitive to decoding skills than to language comprehension (Keenan et al., 2008). The test type may also explain why the differences between the effect sizes of SES on reading comprehension were small for the two groups. Without the samples from the study by Cobo-Lewis et al., the number of samples that reported SES is too few to conduct a moderator analysis. Thus, more research on how SES relates to second-language learner reading comprehension is warranted. In future studies, one should also measure exposure to decontextualized language or, preferably, decontextualized language skills to deter-

mine whether a decontextualized first language is a moderator between SES and second-language skills in reading comprehension and its underlying components. No such study seems to exist.

The fact that the gap between first- and second-language learners was smaller for tests based on cloze and sentences than for tests using passages and open-ended questions may relate to the components of reading comprehension skills tapped by these different kinds of tests. As has been found in several studies, short-passage and sentence-based cloze tests are more dependent on decoding skills than are tests with longer open-ended questions (e.g., Keenan et al., 2008; Lervåg & Aukrust, 2010). This difference is likely to arise because failure to decode one word will have more severe consequences in shorter than in longer open-ended tests, in which there are more words to "lean on" (Keenan et al., 2008). Similarly, longer tests may have additional and more complex plots that interact, and such tests will favor children with good oral language skills, given that larger amount of language might be needed to understand these plots. Further, open-ended tests might require more detailed and lengthy answers than the cloze tests. Thus, the findings that short-passage and sentence-based cloze tests are easier for second-language learners may reflect the fact that second-language learners are as good as first-language learners at decoding yet have poorer language comprehension skills. The reason that the difference between first- and second-language learners was smaller with multiple choice tests (compared to single questions) may relate to the validity of multiple choice tests. Keenan and Betjemann (2006) found that students were able to correctly answer multiple-choice questions without reading the corresponding passages of the Grey Oral Reading Test. They argued that such passage-independent items resulted in a lack of both content and concurrent validity. Such lack of validity would lead to an incorrect estimation of reading comprehension skills and thus erroneously indicate a smaller difference between first- and second-language learners.

Language Comprehension

The difference between first- and second-language learners was clearly moderated by SES. There was a smaller difference in language comprehension between first- and second-language learners coming from a middle or high SES than from a low SES home. This finding accords with theories that consider SES to be a proxy for exposure to greater amounts of decontextualized language (e.g., Cummins, 1979) and with research finding language enhancement in children of high SES parents who use more elaborate vocabulary and engage in more non-context-dependent conversations with their children (Hart & Risley, 1995; Hoff, 2006; Pan et al., 2005). Furthermore, it seems likely that parents with high educational levels are more able to support their children's learning of a second language at home and at school. Parents with higher educational levels are likely more experienced in learning and using a second language in an academic context, better able to transfer their knowledge to a second language and more likely to influence their children with positive attitudes towards education and learning in general. All of these considerations could help to decrease the gap between first- and second-language learners.

That second-language learners who regularly use both the first and the second language at home had language comprehension skills at a level more similar to first- than second-language learners (who used only their first language at home) is consistent with the "time on task" hypothesis (Porter, 1990). That is, children who spend more time using the second language at home will gain better second-language comprehension skills. However, even if this finding seems to support the time on task hypothesis, this finding may be partly confounded with parental education and/or skills in the second language. If the parents have an in-depth knowledge of the second language, they are most likely able to use both languages in a way that facilitates the development of second-language comprehension skills in their children. Furthermore, and somewhat consistent with this supposition, using both languages at home may reflect how long the child and his or her family have lived in their country of residence and been exposed to the second language. The longer they have stayed in their resident country, the more they have been exposed to the second language; the more they have been exposed to the language, the better they speak it. Thus, it is more likely that the family speaks the second language at home. This hypothesis could not be analyzed because too few studies reported the parental length of residence in the country.

Another interesting finding is that the gap between first- and second-language learners was smaller in Canada than in Europe and the United States. The reasons for this difference may be linked to differences in the countries' immigration patterns. Whereas Europe and the United States have many immigrants from poorly educated developing countries, Canada has many well-educated immigrants from various Asian countries (Antecol et al., 2003). In addition, Canadian studies focusing on English–French bilingualism are not affected by additional problems associated with immigration and the cultural diversity found in most of the European and U.S. studies (e.g., lack of schooling, attitudes toward education). If the reasoning regarding aberrant educational levels in the Canadian studies is correct, it should partially explain the relatively small gap between first- and second-language learners in Canada. Well-educated immigrants from developing countries often have relatively smaller incomes (Bonikowska, Hou, & Picot, 2011) and are therefore likely to settle in neighborhoods of a lower SES. When the children of these well-educated immigrants are compared to the children of less educated nonimmigrants based on their academic skills, they do not lag as far behind on second-language comprehension skills as do less educated immigrants in Europe and the United States.

Finally, neither instructional language nor language type moderated the differences between first- and second-language learner language comprehension skills. It may be that cognates do not play a strong role in overall second language comprehension acquisition. Still, using language type as a proxy of cognate transfer may be considered rather crude.

Decoding and Phonological Awareness

There was a reliable difference between the decoding skills of first- and second-language learners, in favor of first-language learners. Notably, when publication bias is corrected for in a trim and fill analysis, the difference in decoding skills is significant in favor of second-language learners. Also, there was a reliable difference between second-language learners in Canada and the United States. Although the Canadian second-language learners

were better at decoding than then their first-language-learner peers, the opposite was true for the second-language learners in the United States. As mentioned within the discussion of language comprehension skills, this finding is likely to reflect differences in educational levels between the immigrant populations in Canada and the United States. None of the other moderators of the differences in decoding skills between first- and second-language learners were reliable. Thus, having an ideographic first language or learning to read a regular second-language orthography did not affect the size of the differences in decoding skills between the two groups. Furthermore, the test format was inconsequential, as no differences were found between accuracy and fluency tests.

There were no reliable differences between the two groups on phonological awareness, but the overall effect showed that first-language learners had slightly better phonological awareness skills than did second-language learners. This is to some extent in opposition to the hypothesis that phonological awareness will be better among second-language learners. Thus, second-language learners' position to compare the two different languages does not seem to benefit their development of phonological awareness in the second language. It may also be that the additional complications of having to learn two languages override the presumable advantage of being able to compare two languages. Still, because the group difference in favor of first-language learners is very small and not significant, this should be interpreted with caution.

The meta-analysis by Adesope et al. (2010) found a moderate metalinguistic advantage among second-language learners (d = 0.33). There are two reasons that they reach a different conclusion than does our meta-analysis. First, their construct, metalinguistic awareness, was defined more broadly than our phonological awareness construct, as they included measures of language reasoning and grammatical judgment. Second, they coded group differences in metalinguistic awareness based on the size of profile effects among bilingual children compared with monolingual children, as reported in a publication by Oller, Pearson, and Cobo-Lewis (2007). The coding of the profile effect as an effect size of group differences between bilingual and monolingual children is misleading, as the profile effect refers to how even the performance of bilingual children is across different test types compared to that of monolinguals, rather than to the differences between the groups per se (D. K. Oller, personal communication, September, 13, 2010).

Limitations and Future Directions

Regarding the limitations with this meta-analysis, an important caveat is that in spite of efforts to locate unpublished literature, only three of the studies we located could be included in the meta-analysis. It is therefore possible that group differences can be overestimated due to publication bias (see Scherer, Langenberg, & von Elm, 2007). In addition, moderators such as age of L2 acquisition, length of residence in the host country for children and parents, parental L2 fluency, and alpha reliability could not be examined, as these factors were reported in too few studies. For future research, this potentially important information should be examined and reported.

Furthermore, although a number of cross-sectional studies have compared first- and second-language learners, results of longitudinal studies are mixed (see the introduction). Additional longitudinal studies are required to clarify the developmental trajectories of second-language learners. Such studies can create a foundation for more targeted interventions for this group.

An important implication from our review is that to ameliorate reading comprehension problems among second-language learners, language comprehension skills are important. Several studies of vocabulary interventions, in which specific words are trained, have shown positive effects (Elleman, Lindo, Morphy, & Compton, 2009). However, the effects of this intervention tend to be specific to the trained words; transfer effects to other words and to reading comprehension in general are small (Elleman et al., 2009). Still, a recent study by Clarke, Snowling, Truelove, and Hulme (2010) has identified promising developments for more general effects on both language comprehension and reading comprehension skills by employing broader language interventions. Randomized controlled studies should examine the effects of language comprehension interventions to develop effective programs for second-language learners.

References

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

*Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice, 19*(3), 16–26. doi:10.1111/j.1745-3992.2000.tb00034.x

*Abu-Rabia, S., & Siegel, L. S. (2002). Reading, syntactic, orthographic, and working memory skills of bilingual Arab Canadian children. *Journal of Psycholinguistic Research, 31,* 661–678. doi:10.1023/A:1021221206119

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80,* 207–245. doi:10.3102/0034654310368803

Antecol, H., Cobb-Clark, D. A., & Trejo, S. J. (2003). Immigration policy and the skills of immigrants to Australia, Canada, and the United States. *Journal of Human Resources, 38,* 192–218. doi:10.2307/1558761

*Argulewicz, E. N., Bingenheimer, L. T., & Anderson, C. C. (1983). Concurrent validity of the PPVT-R for Anglo-American and Mexican-American students. *Journal of Psychoeducational Assessment, 1,* 163–167. doi:10.1177/073428298300100207

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics, 24,* 621–635. doi:10.1017/S0142716403000316

*Beech, J. R., & Keys, A. (1997). Reading, vocabulary and language preference in 7- to 8-year-old bilingual Asian children. *British Journal of Educational Psychology, 67,* 405–414. doi:10.1111/j.2044-8279.1997.tb01254.x

*Bialystok, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology, 24,* 560–567. doi:10.1037/0012-1649.24.4.560

*Bialystok, E. (1997). Effects of bilingualism and biliteracy on children's emerging concepts of print. *Developmental Psychology, 33,* 429–440. doi:10.1037/0012-1649.33.3.429

*Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development, 70,* 636–644. doi:10.1111/1467-8624.00046

*Bialystok, E., Barac, R., & Blaye, A. (2010). Word mapping and executive functioning in young monolingual and bilingual children. *Journal of Cognition and Development, 11,* 485–508. doi:10.1080/15248372.2010.516420

*Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading, 9,* 43–61. doi:10.1207/s1532799xssr0901_4

*Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics, 24,* 27–44. doi:10.1017/S014271640300002X

*Bialystok, E., McBride-Chang, C., & Luk, G. (2005). Bilingualism, language proficiency, and learning to read in two writing systems. *Journal of Educational Psychology, 97,* 580–590. doi:10.1037/0022-0663.97.4.580

*Bialystok, E., & Senman, L. (2004). Executive processes in appearance–reality tasks: The role of inhibition of attention and symbolic representation. *Child Development, 75,* 562–579. doi:10.1111/j.1467-8624.2004.00693.x

*Bialystok, E., & Shapero, D. (2005). Ambiguous benefits: The effect of bilingualism on reversing ambiguous figures. *Developmental Science, 8,* 595–604. doi:10.1111/j.1467-7687.2005.00451.x

*Bialystok, E., Shenfield, T., & Codd, J. (2000). Languages, scripts, and the environment: Factors in developing concepts of print. *Developmental Psychology, 36,* 66–76. doi:10.1037/0012-1649.36.1.66

Bonikowska, A., Hou, F., & Picot, G. (2011). *Do highly educated immigrants perform differently in the Canadian and U.S. labour markets?* (Analytical Studies Branch Research Paper Series 2011329e). Retrieved from http://www.statcan.gc.ca/pub/11f0019m/11f0019m2011329-eng.pdf

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-Analysis Version 2 [Computer software]*. Englewood, NJ: Biostat.

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

*Bruck, M., & Genesee, F. (1995). Phonological awareness in young second language learners. *Journal of Child Language, 22,* 307–324. doi:10.1017/S0305000900009806

*Campbell, R., & Sais, E. (1995). Accelerated metalinguistic (phonological) awareness in bilingual children. *British Journal of Developmental Psychology, 13,* 61–68. doi:10.1111/j.2044-835X.1995.tb00664.x

Caravolas, M., Lervåg, A., Defior, S., Seidlová-Málková, G., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science*. Advance online publication. doi:10.1177/0956797612473122

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., . . . Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science, 23,* 678–686. doi:10.1177/0956797611434536

Caravolas, M., Volin, J., & Hulme, C. (2005). Phoneme awareness is a key component of alphabetic literacy skills in consistent and inconsistent orthographies: Evidence from Czech and English children. *Journal of Experimental Child Psychology, 92,* 107–139. doi:10.1016/j.jecp.2005.04.003

*Carreker, S., Neuhaus, G. F., Swank, P. R., Johnson, P., Monfils, M. J., & Montemayor, M. L. (2007). Teachers with linguistically informed knowledge of reading subskills are associated with a Matthew effect in reading comprehension for monolingual and bilingual students. *Reading Psychology, 28,* 187–212. doi:10.1080/02702710601186456

Carroll, J. M., Snowling, M. J., Hulme, C., & Stevenson, J. (2003). The development of phonological awareness in preschool children. *Developmental Psychology, 39,* 913–923. doi:10.1037/0012-1649.39.5.913

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91,* 77–111. doi:10.1016/S0010-0277(03)00164-1

*Chen, X., Anderson, R. C., Li, W., Hao, M., Wu, X., & Shu, H. (2004). Phonological awareness in bilingual and monolingual Chinese children. *Journal of Educational Psychology, 96,* 142–151. doi:10.1037/0022-0663.96.1.142

*Chiappe, P., Glaeser, B., & Ferko, D. (2007). Speech perception, vocabulary, and the development of reading skills in English among Korean- and English-speaking children. *Journal of Educational Psychology, 99,* 154–166. doi:10.1037/0022-0663.99.1.154

*Chiappe, P., & Siegel, L. S. (1999). Phonological awareness and reading acquisition in English- and Punjabi-speaking Canadian children. *Journal of Educational Psychology, 91,* 20–28. doi:10.1037/0022-0663.91.1.20

*Chiappe, P., Siegel, L. S., & Gottardo, A. (2002). Reading-related skills of kindergartners from diverse linguistic backgrounds. *Applied Psycholinguistics, 23,* 95–116. doi:10.1017/S014271640200005X

*Chiappe, P., Siegel, L. S., & Wade-Woolley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading, 6,* 369–400. doi:10.1207/S1532799XSSR0604_04

Child Trends Databank. (2011). Child Trends' calculations of U. S. Census Bureau, School Enrollment—Social and Economic Characteristics of Students: October 2005 (Detailed Tables: Table 1). Retrieved from http://www.census.gov/population/www/socdemo/school/cps2005.html

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading comprehension difficulties: A randomised controlled trial. *Psychological Science, 21,* 1106–1116. doi:10.1177/0956797610375449

*Cobo-Lewis, A., Pearson, B. Z., Eilers, R. E., & Umbel, V. C. (2002). Effects of bilingualism and bilingual education on oral and written English skills: A multifactor study of standardised test outcomes. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 64–99). Clevedon, United Kingdom: Multilingual Matters.

Connor, U. (1996). *Contrastive rhetoric: Cross-cultural aspects of second-language writing*. Cambridge, England: Cambridge University Press. doi:10.1017/CBO9781139524599

Crystal, D. (1997). *The Cambridge encyclopedia of language*. Cambridge, England: Cambridge University Press.

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

*da Fontoura, H. A., & Siegel, L. S. (1995). Reading, syntactic, and working memory skills of bilingual Portuguese–English Canadian children. *Reading and Writing, 7,* 139–153. doi:10.1007/BF01026951

*Dalton, B., Proctor, C. P., Uccelli, P., Mo, E., & Snow, C. E. (2011). Designing for diversity: The role of reading strategies and interactive vocabulary in a digital reading environment for fifth-grade monolingual English and bilingual students. *Journal of Literacy Research, 43,* 68–100. doi:10.1177/1086296X10397872

*D'Angiulli, A., Siegel, L. S., & Sierra, E. (2001). The development of reading in English and Italian in bilingual children. *Applied Psycholinguistics, 22,* 479–507. doi:10.1017/S0142716401004015

*Davidson, D., & Tell, D. (2005). Monolingual and bilingual children's use of mutual exclusivity in the naming of whole objects. *Journal of Experimental Child Psychology, 92,* 25–45. doi:10.1016/j.jecp.2005.03.007

de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology, 91,* 450–476. doi:10.1037/0022-0663.91.3.450

*de Ramírez, R. D., & Shapiro, E. S. (2006). Curriculum-based measurement and the evaluation of reading skills of Spanish-speaking English language learners in bilingual education classrooms. *School Psychology Review, 35,* 356–369.

*Dodd, B., So, L. K. H., & Lam, K. K. C. (2008). Bilingualism and literacy: The effect of language pair on phonological awareness abilities. *Australian Journal of Learning Difficulties, 13,* 99–113. doi:10.1080/19404150802380514

*Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly, 38,* 78–103. doi:10.1598/RRQ.38.1.4

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463. doi:10.1111/j.0006-341X.2000.00455.x

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2,* 1–44. doi:10.1080/19345740802539200

*Engel de Abreu, P. M. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory, 19,* 529–537. doi:10.1080/09658211.2011.590504

EU European Commission's Directorate-General for Education and Culture. (2008). *Education and migration: Strategies for integrating migrant children in European schools and societies*. Retrieved from ec.europa.eu/culture/documents/education_migration_nesse.pdf

*Everatt, J., Smythe, I., Adams, E., & Ocampo, D. (2000). Dyslexia screening measures and bilingualism. *Dyslexia, 6,* 42–56. doi:10.1002/(SICI)1099-0909(200001/03)6:1<42::AID-DYS157>3.0.CO;2-0

*Fernandez, R. M., & Nielsen, F. (1986). Bilingualism and Hispanic scholastic achievement: Some baseline results. *Social Science Research, 15,* 43–70. doi:10.1016/0049-089X(86)90003-7

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

*Frederickson, N., & Frith, U. (1998). Identifying dyslexia in bilingual children: A phonological approach with inner London Sylheti speakers. *Dyslexia, 4,* 119–131. doi:10.1002/(SICI)1099-0909(199809)4:3<119::AID-DYS112>3.0.CO;2-8

*García, G. E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. *Reading Research Quarterly, 26,* 371–392. doi:10.2307/747894

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28,* 887–898. doi:10.1044/1092-4388(2005/028)

Geva, E. (2006). Second-language oral proficiency and second-language literacy. 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. 123–139). Mahwah, NJ: Erlbaum.

*Geva, E., & Zadeh, Z. Y. (2006). Reading efficiency in native-English-speaking and English-as-a-second language children: The role of oral proficiency and underlying cognitive-linguistic processes. *Scientific Studies of Reading, 10,* 31–57. doi:10.1207/s1532799xssr1001_3

*Goetry, V., Wade-Woolley, L., Kolinsky, R., & Mousty, P. (2006). The role of stress processing abilities in the development of bilingual reading. *Journal of Research in Reading, 29,* 349–362. doi:10.1111/j.1467-9817.2006.00313.x

*Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition, 6,* 1–15. doi:10.1017/S1366728903001007

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading and reading disability. *Remedial and Special Education, 7,* 6–10. doi:10.1177/074193258600700104

*Grant, A., Gottardo, A., & Geva, E. (2011). Reading in English as a first or second language: The case of grade 3 Spanish, Portuguese, and English speakers. *Learning Disabilities: Research & Practice, 26*(2), 67–83. doi:10.1111/j.1540-5826.2011.00327.x

*Hacquebord, H. (1994). L2-reading in the content areas: Text comprehension in secondary education in the Netherlands. *Journal of Research in Reading, 17,* 83–98. doi:10.1111/j.1467-9817.1994.tb00056.x

Hanley, J. R. (2005). Learning to read in Chinese. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 316–335). Oxford, England: Blackwell. doi:10.1002/9780470757642.ch17

*Hannon, P., & McNally, J. (1986). Children's understanding and cultural factors in reading test performance. *Educational Review, 38,* 237–246. doi:10.1080/0013191860380304

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

Hedges, L. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics, 6,* 107–128. doi:10.3102/10769986006002107

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

*Hemsley, G., Holm, A., & Dodd, B. (2006). Diverse but not different: The lexical skills of two primary age bilingual groups in comparison to monolingual peers. *International Journal of Bilingualism, 10,* 453–476. doi:10.1177/13670069060100040401

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26,* 55–88. doi:10.1016/j.dr.2005.11.002

Huang, H. S., & Hanley, J. R. (1995). Phonological awareness and visual skills in learning to read Chinese and English. *Cognition, 54,* 73–98. doi:10.1016/0010-0277(94)00641-W

Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002). Phoneme awareness is a better predictor of early reading skills than onset-rime awareness. *Journal of Experimental Child Psychology, 82,* 2–28. doi:10.1006/jecp.2002.2670

Hulme, C., & Snowling, M. J. (2009). *Developmental disorders of language learning and cognition*. Chichester, England: Wiley-Blackwell.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

*Hutchinson, J. M., Whiteley, H. E., Smith, C. D., & Connors, L. (2003). The developmental progression of comprehension-related skills in children learning EAL. *Journal of Research in Reading, 26,* 19–32. doi:10.1111/1467-9817.261003

*Hutchinson, J. M., Whiteley, H. E., Smith, C., & Connors, L. (2004). The early identification of dyslexia: Children with English as an additional language. *Dyslexia, 10,* 179–195. doi:10.1002/dys.275

*Ibrahim, R., Eviatar, Z., & Aharon-Peretz, J. (2007). Metalinguistic awareness and reading performance: A cross language comparison. *Journal of Psycholinguistic Research, 36,* 297–317. doi:10.1007/s10936-006-9046-3

Institute for Employment Studies. (2004). *Why the difference? A closer look at higher education minority ethnic students and graduates*. Retrieved from http://www.employment-studies.co.uk/pubs/report.php?id=rr552

Ioannidis, J. P. A. (1998). Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA: Journal of the American Medical Association, 279,* 281–286. doi:10.1001/jama.279.4.281

*Janssen, M., Bosman, A. M. T., & Leseman, P. P. M. (2013). Phoneme awareness, vocabulary and word decoding in monolingual and bilingual Dutch children. *Journal of Research in Reading, 36,* 1–13. doi:10.1111/j.1467-9817.2011.01480.x

*Jean, M., & Geva, E. (2009). The development of vocabulary in English as a second language children and its role in predicting word recognition ability. *Applied Psycholinguistics, 30,* 153–185. doi:10.1017/S0142716408090073

Jennions, M. D., & Møller, A. P (2002). Relationships fade with time: A meta-analysis of temporal trends in publication in ecology and evolution. *Proceedings of the Royal Society, B: Biological Sciences, 269,* 43–48. doi:10.1098/rspb.2001.1832

*Jongejan, W., Verhoeven, L., & Siegel, L. S. (2007). Predictors of reading and spelling abilities in first- and second-language learners. *Journal of Educational Psychology, 99,* 835–851. doi:10.1037/0022-0663.99.4.835

Keenan, J., & Betjemann, R. (2006). Comprehending the Gray Oral Reading Test without reading it: Why comprehension tests should not include passage-independent items. *Scientific Studies of Reading, 10,* 363–380. doi:10.1207/s1532799xssr1004_2

Keenan, J., Betjemann, R., & Olson, R. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12,* 281–300. doi:10.1080/10888430802132279

*Kelly, P. R., Gómez-Bellengé, F. X., Chen, J., & Schulz, M. (2008). Learner outcomes for English language learner low readers in an early intervention. *TESOL Quarterly, 42,* 235–260.

*Kovelman, I., Baker, S. A., & Petitto, L. A. (2008). Age of first bilingual language exposure as a new window into bilingual reading development. *Bilingualism: Language and Cognition, 11,* 203–223. doi:10.1017/S1366728908003386

Lervåg, A. (2010, July). *The structure and development of beginning reading comprehension skills*. Paper presented at the conference of the Society for the Scientific Study of Reading, Berlin, Germany.

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

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A Norwegian latent variable longitudinal study. *Developmental Psychology, 45,* 764–781. doi:10.1037/a0014132

Lesaux, N. K., Koda, K., Siegel, L. S., & Shanahan, T. (2006). Development of literacy. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners* (pp. 75–122). Mahway, NJ: Erlbaum.

*Lesaux, N. K., Rupp, A. A., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds: Findings from a 5-year longitudinal study. *Journal of Educational Psychology, 99,* 821–834. doi:10.1037/0022-0663.99.4.821

*Linn, G. B. (1967). Linguistic functions of bilingual Mexican-American children. *Journal of Genetic Psychology, 111,* 183–193. doi:10.1080/00221325.1967.10533761

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Macmillan, B. M. (2002). Rhyme and reading: A critical review of the research methodology. *Journal of Research in Reading, 25,* 4–42. doi:10.1111/1467-9817.00156

*Mahon, M., & Crutchley, A. (2006). Performance of typically-developing school-age children with English as an additional language (EAL) on the British Picture Vocabulary Scales II (BPVS II). *Child Language Teaching and Therapy, 22,* 333–351. doi:10.1191/0265659006ct311xx

*Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition, 11,* 81–93. doi:10.1017/S1366728907003227

McBride-Chang, C. (2004). *Children's literacy development*. New York, NY: Oxford University Press.

*McBride-Chang, C., Bialystok, E., Chong, K. K. Y., & Li, Y. (2004). Levels of phonological awareness in three cultures. *Journal of Experimental Child Psychology, 89,* 93–111. doi:10.1016/j.jecp.2004.05.001

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 and morphological awareness to reading in Beijing, Hong Kong, Korea, and America. *Journal of Experimental Child Psychology, 92,* 140–160. doi:10.1016/j.jecp.2005.03.009

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

Melby-Lervåg, M., Lervåg, A., Lyster, S. A. H., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological Science, 23,* 1092–1098. doi:10.1177/0956797612443833

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138,* 322–352. doi:10.1037/a0026744

*Messer, M. H., Leseman, P., Boom, J., & Mayo, A. Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology, 105,* 306–323. doi:10.1016/j.jecp.2009.12.006

*Miller Guron, L., & Lundberg, I. (2003). Identifying dyslexia in multilingual students: Can phonological awareness be assessed in the majority language? *Journal of Research in Reading, 26,* 69–82. doi:10.1111/1467-9817.261006

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137,* 267–296. doi:10.1037/a0021890

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40,* 665–681. doi:10.1037/0012-1649.40.5.665

*Myers, B. J., & Goldstein, D. (1979). Cognitive development in bilingual and monolingual lower-class children. *Psychology in the Schools, 16,* 137–142. doi:10.1002/1520-6807(197901)16:1<137::AID-PITS2310160124>3.0.CO;2-Z

Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2007). A longitudinal analysis of English language learners' word decoding and reading comprehension. *Reading and Writing, 20,* 691–719. doi:10.1007/s11145-006-9045-7

Näslund, J. C., & Schneider, W. (1991). Longitudinal effects of verbal ability, memory capacity, and phonological awareness on reading performance. *European Journal of Psychology of Education, 6,* 375–392. doi:10.1007/BF03172772

National Center for Education Statistics. (2011). [Information]. Retrieved from http://nces.ed.gov/programs/coe/indicator_lsm.asp

National Institute for Literacy. (2008). *Developing early literacy: A scientific analysis of early literacy development and implications for intervention*. Retrieved from http://www.nifl.gov/publications/pdf/NELPReport09.pdf

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00–4754). Washington, DC: U.S. Government Printing Office.

National Institute of Child Health and Human Development Early Child Care Research Network. (2005). Pathways to reading: The role of language in the transition to reading. *Developmental Psychology, 41,* 428–442. doi:10.1037/0012-1649.41.2.428

*Nicoladis, E. (2006). Cross-linguistic transfer in adjective–noun strings by preschool bilingual children. *Bilingualism: Language and Cognition, 9,* 15–32. doi:10.1017/S136672890500235X

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Oxford, England: Cambridge University Press. doi:10.1017/CBO9781139524537

OECD. (2000). *Knowledge and skills for life. First results from the OECD Programme for International Student assessment (PISA) 2000*. Retrieved from http://www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/33691596.pdf

OECD. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris, France: Author.

OECD Reviews of Migrant Education. (2009). *Closing the gap for immigrant students: Policies, practice and performance*. Retrieved from http://www.oecd.org/document/53/0,3343,en_2649_39263231_44870901_1_1_1_1,00.html

Oller, D. K., Pearson, B., & Cobo-Lewis, A. (2007). Profile effects in early bilingual language and literacy acquisition. *Applied Psycholinguistics, 28,* 198–229. doi:10.1017/S0142716407070117

Ordoñez, C. L., Carlo, M. S., Snow, C. E., & McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology, 94,* 719–728. doi:10.1037/0022-0663.94.4.719

*O'Toole, S., Aubeeluck, A., Cozens, B., & Cline, T. (2001). Development of reading proficiency in English by bilingual children and their monolingual peers. *Psychological Reports, 89,* 279–282. doi:10.2466/PR0.89.6.279-282

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development, 76,* 763–782. doi:10.1111/1467-8624.00498-i1

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to know. *Reading Research Quarterly, 42,* 282–296. doi:10.1598/RRQ.42.2.4

*Phillips, C. J., & Marvelly, A. (1984). Basic educational skills of Asian children at the end of infant schooling. *Educational Research, 26,* 36–45. doi:10.1080/0013188840260107

Pigott, T. D. (2012). *Advances in meta-analysis*. New York, NY: Springer. doi:10.1007/978-1-4614-2278-5

Porter, R. (1990). *Forked tongue*. New York, NY: Basic Books.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods, 10,* 178–192. doi:10.1037/1082-989X.10.2.178

*Proctor, C. P., Uccelli, P., Dalton, B., & Snow, C. E. (2009). Understanding depth of vocabulary and improving comprehension online with bilingual and monolingual children. *Reading and Writing Quarterly, 25,* 311–333. doi:10.1080/10573560903123502

*Rosenthal, A., Baker, K., & Ginsburg, A. (1983). The effect on language background on achievement level and learning outcome among elementary school students. *Sociology of Education, 56,* 157–169. doi:10.2307/2112545

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and reading. *Journal of Educational Research, 95,* 259–272. doi:10.1080/00220670209596600

*San Francisco, A. R. (2003). *Cross-language interference in the phonological awareness of Spanish–English bilingual children* (ERIC Report ED482581). Retrieved from http://www.eric.ed.gov/pdf_availability.html

*San Francisco, A. R., Mo, A., Carlo, M., August, D., & Snow, C. (2006). The influence of language of literacy instruction and vocabulary on the spelling of Spanish–English bilinguals. *Reading and Writing, 19,* 627–642. doi:10.1007/s11145-006-9012-3

*Sattler, J. M., & Altes, L. M. (1984). Performance of bilingual and monolingual Hispanic children on the Peabody Picture Vocabulary Test–Revised and the McCarthy Perceptual Performance Scale. *Psychology in the Schools, 21,* 313–316. doi:10.1002/1520-6807(198407)21:3<313::AID-PITS2310210307>3.0.CO;2-B

*Schaerlaekens, A., Zink, I., & Verheyden, L. (1995). Comparative vocabulary development in kindergarten classes with a mixed population of monolinguals, simultaneous and successive bilinguals. *Journal of Multilingual and Multicultural Development, 16,* 477–495. doi:10.1080/01434632.1995.9994619

Schatschneider, D., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96,* 265–282. doi:10.1037/0022-0663.96.2.265

Scherer, R., Langenberg, P., & von Elm, E. (2007). Full publication of results initially presented in abstracts. *Cochrane Database of Systematic Reviews, 2007*(2), Article MR000005.pub3. doi:10.1002/14651858

*Scruggs, T., Mastropieri, M., & Argulewicz, W. (1983). Stability of performance on the PPVT-R for three ethnic groups attending a bilingual kindergarten. *Psychology in the Schools, 20,* 433–435. doi:10.1002/1520-6807(198310)20:4<433::AID-PITS2310200407>3.0.CO;2-B

*Segers, E., & Verhoeven, L. (2005). Long-term effects of computer training of phonological awareness in kindergarten. *Journal of Computer Assisted Learning, 21,* 17–27. doi:10.1111/j.1365-2729.2005.00107.x

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568. doi:10.1037/0033-295X.96.4.523

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation of literacy acquisition in European orthographies. *British Journal of Psychology, 94,* 143–174. doi:10.1348/000712603321661859

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Share, D. L. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin, 134,* 584–615. doi:10.1037/0033-2909.134.4.584

*Sheng, L., McGregor, K. K., & Marian, V. (2006). Lexical-semantic organization in bilingual children: Evidence from a repeated word association task. *Journal of Speech, Language, and Hearing Research, 49,* 572–587. doi:10.1044/1092-4388(2006/041)

*Siegal, M., Iozzi, L., & Surian, L. (2009). Bilingualism and conversational understanding in young children, *Cognition, 110,* 115–122. doi:10.1016/j.cognition.2008.11.002

*Silverman, R. (2007). Vocabulary development of English-language and English-only learners in kindergarten. *Elementary School Journal, 107,* 365–383. doi:10.1086/516669

Snow, C. E., & Kim, Y.-S. (2007). Large problem spaces: The challenge of vocabulary for English language learners. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 123–139). New York, NY: Guilford Press.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38,* 934–947. doi:10.1037/0012-1649.38.6.934

*Thomas, W. P., & Collier, V. P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement*. Santa Cruz, CA: Center for Research on Education, Diversity and Excellence, University of California, Santa Cruz.

*Thorn, A. S. C., & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 52*(A), 303–324. doi:10.1080/713755823

Vaessen, A., Bertrand, D., Tóth, D., Csépe, V., Faísca, L., Reis, A., & Blomert, L. (2010). Cognitive development of fluent word reading does not qualitatively differ between transparent and opaque orthographies. *Journal of Educational Psychology, 102,* 827–842. doi:10.1037/a0019465

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power in meta-analysis. *Journal of Educational and Behavioral Statistics, 35,* 215–247. doi:10.3102/1076998609346961

*van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Snellings, P., Simis, A., & Stevenson, M. (2003). Roles of linguistic knowledge,

metacognitive knowledge and processing speed in L3, L2 and L1 reading comprehension: A structural equation modeling approach. *International Journal of Bilingualism, 7,* 7–25. doi:10.1177/13670069030070010201

*Verhoeven, L. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading, 4,* 313–330. doi:10.1207/S1532799XSSR0404_4

*Verhoeven, L., & Vermeer, A. (2006). Literacy achievement of children with intellectual disabilities from diverse linguistic backgrounds. *Journal of Intellectual Disability Research, 50,* 725–738. doi:10.1111/j.1365-2788.2006.00838.x

*Vermeer, A. (2001). Breadth and depth of vocabulary in relation to acquisition and frequency of input. *Applied Psycholinguistics 22,* 217–234. doi:10.1017/S0142716401002041

*Wade-Woolley, L., & Siegel, L. S. (1997). The spelling performance of ESL and native speakers of English as a function of reading skill. *Reading and Writing, 9,* 73–92. doi:10.1023/A:1007994928645

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). The development of reading-related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study. *Developmental Psychology, 30,* 73–87. doi:10.1037/0012-1649.30.1.73

*Wang, M., & Geva, E. (2003). Spelling performance of Chinese ESL children: Lexical and visual-orthographic processes. *Applied Psycholinguistics, 24,* 1–25. doi:10.1017/S0142716403000018

*Westman, M., Korkman, M., Mickos, A., & Byring, R. (2008). Language profiles of monolingual and bilingual Finnish preschool children at risk for language impairment. *International Journal of Language & Communication Disorders, 43,* 699–711. doi:10.1080/13682820701839200

Wilson, D. B. (2006). *Meta-analysis macros for SAS, SPSS, and Stata*. Retrieved from http://mason.gmu.edu/~dwilsonb/ma.html

Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics, 14,* 1–33. doi:10.1017/S0142716400010122

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., . . . Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science, 21,* 551–559. doi:10.1177/0956797610363406

Received April 3, 2012
Revision received April 13, 2013
Accepted April 29, 2013 ■