



Foreign Language Attainment of Children/Adolescents with Poor Literacy Skills: a Systematic Review and Meta-analysis

Alexa von Hagen¹  · Saskia Kohnen²  · Nicole Stadie³ 

Published online: 11 September 2020

© The Author(s) 2020

Abstract

This systematic review investigated how successful children/adolescents with poor literacy skills learn a foreign language compared with their peers with typical literacy skills. Moreover, we explored whether specific characteristics related to participants, foreign language instruction, and assessment moderated scores on foreign language tests in this population. Overall, 16 studies with a total of 968 participants (poor reader/spellers: $n = 404$; control participants: $n = 564$) met eligibility criteria. Only studies focusing on English as a foreign language were available. Available data allowed for meta-analyses on 10 different measures of foreign language attainment. In addition to standard mean differences (SMDs), we computed natural logarithms of the ratio of coefficients of variation (CVRs) to capture individual variability between participant groups. Significant between-study heterogeneity, which could not be explained by moderator analyses, limited the interpretation of results. Although children/adolescents with poor literacy skills on average showed lower scores on foreign language phonological awareness, letter knowledge, and reading comprehension measures, their performance varied significantly more than that of control participants. Thus, it remains unclear to what extent group differences between the foreign language scores of children/adolescents with poor and typical literacy skills are representative of individual poor readers/spellers. Taken together, our results indicate that foreign language skills in children/adolescents with poor literacy skills are highly variable. We discuss the limitations of past research that can guide future steps toward a better understanding of individual differences in foreign language attainment of children/adolescents with poor literacy skills.

Keywords Poor literacy · Dyslexia · Foreign language · Bilingualism · Meta-analysis

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10648-020-09566-6>) contains supplementary material, which is available to authorized users.

✉ Alexa von Hagen
vonHagen@em.uni-frankfurt.de

Extended author information available on the last page of the article

Introduction

While research has investigated native language abilities of children/adolescents with poor literacy skills quite extensively, less attention has been paid to their success in learning a foreign language in formal educational settings. Relatives, teachers, and allied health professionals often assume that the difficulties children/adolescents with poor literacy skills experience in their native language will transfer to the new language being learned (Sparks 2016). As a consequence, children/adolescents with poor literacy skills receive less support and are even exempted from foreign language instruction in many countries. One example is an Italian law, allowing children/adolescents with poor literacy skills to be completely excused from foreign language learning (see Palladino et al. 2013). Researchers in the field of foreign language learning express their concern about these policies, as to date the evidence regarding foreign language difficulties in poor readers/spellers is scarce. Wight (2015) suggests that the policies and practices of exempting students from foreign language study demonstrate that they are often discharged “(1) based on personal beliefs and preferences rather than on the basis of a carefully considered consensus of inclusion, and (2) in the absence of actual data about the potential successes of students with special needs” (pp. 41–42, Wight 2015). Although these policies aim to protect children/adolescents with poor literacy skills from experiencing failure, they also impede those students to gain cognitive and professional advantages associated with foreign language learning (e.g., on inhibitory control—Bialystok and Majumder 1998; in theory of mind development—Kovács 2009; on metalinguistic knowledge—Bialystok 2012). Moreover, access to cultural diversity remains limited without being able to speak an additional language. Despite having a profound impact on students’ future opportunities, these decisions are—to date—not based on a systematic evaluation of the existing evidence.

Thus, this study aimed to address this gap by conducting a systematic review of the available evidence on foreign language attainment in children/adolescents with poor literacy skills. More specifically, we identified, critically appraised and synthesized existing evidence reported by past research studies to address the following two research questions:

1. How successful are children/adolescents with poor literacy skills in learning a foreign language, as compared with children/adolescents with typical literacy skills?
2. Is successful foreign language attainment in children/adolescents with poor literacy skills influenced by moderators such as participant characteristics, foreign language instruction, and foreign language assessment?

In this way, we intended to provide a systematic overview of the current state of research and provide useful input for future research in this field by highlighting limitations of past research, as well as research gaps that need to be addressed.

In coherence with the preferred reporting items for systematic reviews and meta-analyses from the PRISMA statement (Moher et al. 2009—see S1 in the Supplemental Materials for a completed checklist of the PRISMA items: <https://osf.io/be9x4/>), we define the main elements of our research questions according to the PICO acronym (population, intervention, comparison, outcome—Moher et al. 2009) in the following sections.

Population: Children/Adolescents with Poor Literacy Skills

Worldwide a significant proportion of children/adolescents present with poor literacy skills that cannot be explained by medical, emotional, or neurological difficulties or insufficient literacy instruction. Prevalence rates of children/adolescents with poor literacy skills range from 3.1 to 17.5% across languages (e.g., 3.1–3.2% in Italian, Barbiero et al. 2012; 5% in German, Müller et al. 2014; 17.5% in American English, Shaywitz et al. 2008). The severity and type of difficulties children/adolescents with poor literacy skills experience varies across languages as a function of the complexity of the target writing system they are learning (Goulandris 2003; Landerl et al. 1996; Sprenger-Charolles et al. 2011). For instance, within alphabetic writing systems, “transparent” scripts with simple phoneme–grapheme rules (e.g., German, Italian, Spanish) are easier to learn than “opaque” scripts with complex phoneme–grapheme rules (e.g., English and French—Katz and Frost 1992; Ziegler et al. 2010).

Many different terms have been used to describe poor literacy skills: e.g., developmental dyslexia and/or dysgraphia, specific reading and/or spelling difficulty, and reading and/or spelling impairment, deficit, or disability (Elliott and Grigorenko 2014; Siegel 1988, 2007). In the present review, we use the term “children/adolescents with poor literacy skills” or “poor readers/spellers” to refer to students with low scores on reading and/or spelling tests.

The term poor literacy encompasses both reading and spelling difficulties in the present review. With respect to reading difficulties, we included inaccurate or slow reading of words, nonwords, sentences, or texts (International Dyslexia Association 2012). While some poor readers/spellers may additionally struggle with reading comprehension tasks, students solely facing reading comprehension difficulties and no other reading deficits (e.g., inaccurate or slow word or nonword reading) have been excluded from the group of children/adolescents with poor literacy skills in this study. This is mainly due to the fact that reading comprehension difficulties have been shown to be associated with oral language deficits (e.g., poor vocabulary knowledge, poor sentence comprehension), especially when they occur without additional difficulties in reading accuracy and/or fluency (e.g., Oakhill et al. 2003). Concerning spelling difficulties, children/adolescents with poor literacy skills may struggle in spelling words or nonwords either in dictation tasks and/or in spontaneous text production (Kohnen et al. 2015).

Moreover, we defined poor literacy skills as a below average performance (i.e., either one standard deviation, 1 year or grade below the expected level) in either a reading or spelling task or in both (Elliott and Grigorenko 2014). Studies in which participants were included based on self- or teacher reports were excluded (Snowling et al. 2011). We only incorporated studies assessing participants from their first year of formal schooling up to the last year of secondary education, whereas studies investigating students in post-secondary education were excluded.

Comparison: Children/Adolescents with Typical Literacy Skills

To be included in this review, studies had to compare foreign language performance of poor readers and spellers with control participants demonstrating typical literacy skills. Once again, reading or spelling tests had to be used to confirm typical literacy skills. Control participants had to score within the expected age group, grade, or less than one standard deviation below the expected level (Elliott and Grigorenko 2014).

Intervention: Foreign Language Instruction

Although the manner in which a foreign language is being instructed also influences its attainment (Saito and Hanzawa 2016), in this review, we considered all types of classroom-based foreign language instruction (e.g., language vs. content-based approaches; instruction vs. immersion approaches). However, children could only have little or no access to the foreign language being instructed outside of the classroom (see p. 9 Dixon et al. 2012 for a similar delimitation of foreign language context). Thus, all the studies stating that participants had additional access to the foreign language, other than limited access for example to music, films, or travel experiences, were excluded. Examples of excluded reports are studies focusing on the foreign language learning abilities of heritage speakers (i.e., children/adolescents that are exposed to a minority language at home).

Outcome: Foreign Language Attainment

Foreign language attainment involves mastering distinct subskills, such as for example discriminating foreign language speech sounds, comprehending spoken words, or reading and spelling words. It could be possible that children/adolescents with poor literacy skills show a lower performance in only some (e.g., reading and spelling), but not all foreign language subskills. A detailed investigation of existing research on different foreign language subskills of poor readers/spellers can shed light on this issue. However, as different research traditions in the foreign language learning literature have used different labels to describe these subskills, it is sometimes difficult to reconcile classification systems. For example, some authors distinguish between oral and written language, while others contrast receptive and expressive modalities (Nation 2013). Again, others segregate tasks focusing on prelexical, lexical, and nonlexical processing mechanisms (de Bot 1992; de Bot et al. 1997). For the purpose of this review, we consider a classification based on the tasks used to measure foreign language attainment (e.g., picture naming, rhyme detection, lexical decision, etc.) as the most appropriate one to give the reader a concrete idea of the measures that were used in past research. Table 1 shows this classification.

As we were mainly interested in exploring the linguistic aspects involved in foreign language attainment, we decided to exclude studies solely focusing on foreign language learning motivation. We included both self-developed and standardized language tests.

Potential Moderators of Foreign Language Attainment in Children/Adolescents with Poor Literacy Skills

Our secondary research question aimed to investigate if successful foreign language attainment in children/adolescents with poor literacy skills might be influenced by moderators such as participant characteristics, foreign language instruction, and foreign language assessment. We anticipated 11 potential moderators and predefined subgroups for each of these variables. This information is summarized in Table 2 and described in the following sections.

Participant Characteristics A first potential moderator refers to children/adolescents' native language profile. Past research has shown that children/adolescents with poor literacy skills have very heterogeneous performance patterns in their native language (Bishop and Snowling 2004; Friedmann and Coltheart 2016; McArthur et al. 2013; Moll and Landerl 2009). This

Table 1 Foreign language outcome measures

FL outcome measure	Examples of tasks
Discrimination of speech sounds	Auditory discrimination of phonemes, syllables, words, or nonwords
Production of speech sounds	Repetition of phonemes, syllables, words, or nonwords
Receptive vocabulary knowledge	Spoken word-picture matching
Spoken word production	Picture naming Serial rapid naming Semantic fluency
Sentence comprehension	Spoken sentence-picture matching Grammaticality judgments
Sentence production	Sentence elicitation Picture description Conversation
Short-term memory	Digit repetition
Phonological awareness	Rhyme detection Spoonerisms
Letter knowledge	Initial and final phoneme deletion
Word reading	Speeded and unspeeded letter naming Speeded and unspeeded word reading Speeded sentence and text reading
Nonword reading	Speeded and unspeeded nonword reading
Orthographic knowledge	Visual lexical decision
Reading comprehension	Text reading and multiple choice questions Open questions Cloze
Spelling	Word, nonword, and sentence spelling to dictation Word or nonword copy Written story telling
Translation	Translation of words, sentences, or texts

FL = foreign language

heterogeneity may extend to foreign language skills such that only some, but not all poor readers/spellers show a lower attainment than their peers with typical literacy skills. To investigate this issue, we planned to explore the impact of participants' native language profiles on their foreign language attainment.

Based on past research, we first distinguished between poor readers/spellers that only show difficulties in written native language skills and others that also show impaired oral language abilities (Bishop and Snowling 2004). Within the written language domain, we aimed to explore evidence on three subgroups of children/adolescents with poor literacy skill: poor readers/good spellers, good readers/poor spellers, and poor readers/poor spellers (Moll and Landerl 2009).

Finally, we were interested in the impact of different types of reading deficits on foreign language attainment. More specifically, we differentiated poor readers/spellers with sublexical deficits (i.e., difficulties converting letters to sounds), lexical deficits (i.e., difficulties in recognizing written words leading to inaccurate or slow word reading), mixed deficits (i.e., sublexical and lexical impairments), and other reading deficits (e.g., processing letter order, recognizing letters, ordering letters, and moving letters between words—Friedmann and Coltheart 2016; Friedmann and Lukov 2008; Friedmann and Rahamim 2007; Kohnen et al. 2012, 2018; McArthur et al. 2013; Sotiropoulos and Hanley 2017).

Another moderator that could contribute to variable foreign language performance is the diversity of linguistic backgrounds. Several studies have reported higher foreign language

Table 2 Moderators and subgroups

Moderators	Subgroups
Participant characteristics	
NL profile	
Oral/written language deficits ^a	Poor oral NL and poor written NL Average oral NL and poor written NL
Reading/spelling deficits ^b	Poor readers/good spellers Good readers/poor spellers Poor readers/poor spellers
Reading deficit subtype ^c	Sublexical reading deficit Lexical reading deficit Mixed reading deficit Other reading deficits
Linguistic background	Monolingual Bilingual Monolingual and bilingual
FL instruction	
Frequency of FL classes	Less than 2 classes per week Between 2 and 4 classes per week More than 4 classes per week
Duration of FL classes	Less than 30 min Between 30 and 60 min More than 60 min
Language pairing between NL and FL	
Structural differences ^d	NL Indo-European/FL Indo-European NL Indo-European/FL non-Indo-European NL non-Indo-European/FL Indo-European NL non-Indo-European/FL non-Indo-European
Writing system differences ^d	Alphabetic NL/alphabetic FL Alphabetic NL/ideographic FL Ideographic NL/alphabetic FL Ideographic NL/ideographic FL
Orthographic regularity ^e	Regular NL/regular FL Regular NL/irregular FL Irregular NL/regular FL Irregular NL/irregular FL
Onset age of FL instruction ^f	Early childhood: onset age before 6 years Late childhood: onset age from 6 to 11 years Adolescence: onset age from 12 to 17 years Transition from adolescence to adulthood: from 18 years onwards
FL assessment	
Age of FL assessment ^f	Early childhood: before 6 years of age Late childhood: from 6 to 11 years of age Adolescence: from 12 to 17 years of age Transition from adolescence to adulthood: from 18 years onwards

NL = native language; FL = foreign language

^aBishop and Snowling (2004)

^bMoll and Landerl (2009)

^cFriedmann and Coltheart (2016)

^dMelby-Lervåg and Lervåg (2014)

^eOnly categorized within alphabetic writing systems: Seymour et al. (2003) and Ziegler et al. (2010)

^fAbrahamsson and Hyltenstam (2009)

attainment in bilingual as compared with monolingual students with typical literacy skills. For example, Nair et al. (2016) found a significantly better performance in early and late bilinguals

as compared with monolinguals in a novel-word-learning task. Similarly, Tremblay and Sabourin (2012) reported significantly higher speech perception abilities in multi- and bilinguals as compared with monolinguals. Therefore, we aimed to distinguish between studies including only monolingual, bilingual, or both monolingual and bilingual participants.¹

Foreign Language Instruction The quantity of foreign language input has often been highlighted as an important variable to explain foreign language attainment (Wright 2013). Thus, we decided to explore the impact of frequency and duration of foreign language classes on foreign language attainment in children/adolescents with poor literacy skills. In addition, language pairing between native and foreign language can moderate foreign language attainment (Connor 1996; Odlin 1989). Structural similarities or differences between the native and foreign language, for example between Indo-European and non-Indo-European languages, have been shown to either facilitate or impede the acquisition of a foreign language (Connor 1996; Melby-Lervåg and Lervåg 2014; Odlin 1989).

Moreover, Bialystok et al. (2005) pointed out that the orthographic similarity between two writing systems (e.g., alphabetic vs. ideographic writing systems) could explain the extent to which bilingual children were able to positively transfer literacy skills across languages. Within alphabetic writing systems, it seems that the consistency with which a grapheme is mapped onto a phoneme is an important moderator of literacy performance (e.g., Seymour et al. 2003; Ziegler et al. 2010). This is especially relevant in children/adolescents with poor literacy skills. Although students with poor literacy skills have been identified in many different languages (e.g., Frost 2012; Ziegler et al. 2003), some performance patterns seem to differ across languages (Goulandris 2003; Landerl et al. 1996; Moll et al. 2014). Lastly, onset age of foreign language instruction has been shown to influence foreign language skills in children/adolescents with typical literacy skills and was therefore also included as a potential moderator in this review (e.g., Bialystok 1997; DeKeyser et al. 2010; Friederici et al. 2002; Johnson and Newport 1989).

Foreign Language Assessment Similar to onset age, the age at which foreign language abilities are assessed could play a role in explaining foreign language attainment in poor readers/spellers. Indeed, Bialystok (1997) highlighted that older learners rely on wider previous knowledge than younger ones and are therefore able to include new information into already existing conceptual categories. In contrast, younger learners tend to create new categories for the input they receive, which sometimes involves a longer learning process. Similarly, DeKeyser (2000) reported that younger learners rely to a greater extent on implicit mechanisms that may no longer be available to older learners. Older learners depend much more on explicit learning mechanisms. Both types of learning mechanisms have been shown to be beneficial in developing different foreign language subskills. For instance, it appears that speech production relies to a greater extent on implicit learning, while grammatical knowledge is acquired faster through explicit teaching (DeKeyser 2000). Thus, the age at which a foreign language assessment is conducted might impact the magnitude of group differences between children/adolescents with poor and typical literacy skills.

¹ We planned to allocate studies including pluri-/multilingual participants to the bilingual participant group.

Method

The procedures used in this review were predefined in a protocol registered in the *International Prospective Register of Systematic Reviews* PROSPERO (see https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=69980). Differences between the protocol and this final report are detailed in the Supplemental Materials (see S2).

Eligibility Criteria

We summarized the selection criteria in nine “signaling questions” shown in Table 3.

These questions were used during title and abstract as well as full text screening. Furthermore, we only included group comparison studies (as opposed to case studies or case series) that reported on a quantitative comparison of both participant groups on a foreign language outcome measure. We only excluded data reported within doctoral dissertations if the same information was published in a peer-reviewed article.

Information Sources

We searched the following databases: Ovid databases (PsycINFO, PsycARTICLES, Medline, Embase), Wiley Database, PubMed, ProQuest (ERIC, ProQuest dissertations and Linguistics and Language Behavior Abstracts (LLBA)), and Web of Science. Furthermore, we used PsycExtra and Google Scholar to identify gray literature and hand searched the following journals: *International Journal of Bilingual Education and Bilingualism*, *Bilingualism: Language and Cognition*, *Second Language Research*, *TESOL Quarterly*, and *International Journal of Bilingualism*. We also planned to contact authors of more than three independent studies that met inclusion criteria per e-mail and ask for unpublished material, although this did not occur.

Table 3 Eligibility criteria

Signaling questions	Inclusion	Exclusion
1. Do the participants in the study attend primary schools from year 1 onwards or secondary schools?	Yes	No
2. Does the study include participants with poor literacy skills?	Yes	No
3. Do the students with poor literacy skills show a performance of at least 1 SD, 1 year or 1 grade below the expected level on any of the following measures: word/nonword reading accuracy, reading fluency, spelling?	Yes	No
4. Is the condition of average or below average literacy performance of all participants measured by a test as opposed to self- or teacher reports?	Yes	No
5. Are the poor readers/spellers compared to control participants on any foreign language measure?	Yes	No
6. Are the participants allocated to the control or experimental group based on their performance on foreign language tasks?	No	Yes
7. Do the participants have access to the target foreign language outside of the foreign language instruction context (except limited access to music, films, or travel experience)?	No	Yes
8. Does the study measure the foreign language attainment of students with poor literacy skills (oral/ written language)?	Yes	No
9. Is the foreign language attainment being measured thought to be a result of a foreign language instruction received in a classroom context?	Yes	No

Search

Figure 1 details the search strategy used in this review.

This strategy was entered into the databases detailed above between February 10th and 26th of 2017 with no language or date limits. We adapted this search strategy to the requirements of the different databases (see Supplemental Materials S3).

Study Selection

First, we imported the titles and abstracts of the identified studies to Covidence (www.covidence.org). Duplicates were automatically deleted. Further studies that we clearly identified as duplicates were also dismissed. Using the signaling questions (see Table 3), two reviewers independently determined eligibility of studies based on titles and abstracts. Disagreements between two reviewers were resolved through an additional judgment by a third reviewer. In a second step, we repeated the same procedure for the full texts of the studies included from the first stage. If the study did not meet the eligibility criteria, we registered the reason for exclusion in Covidence.

Data Collection Process

In order to extract relevant data from included studies, we customized a data extraction form in Covidence (see Supplemental Materials S4). The first author completed the data extraction form for each of the included studies. The other two authors double-checked data entry against the original studies. Any incongruities were resolved by returning to the original data in the study. In cases of missing data, the corresponding authors of studies were contacted, and if this information was not available, the study (or measure) was excluded from the review.

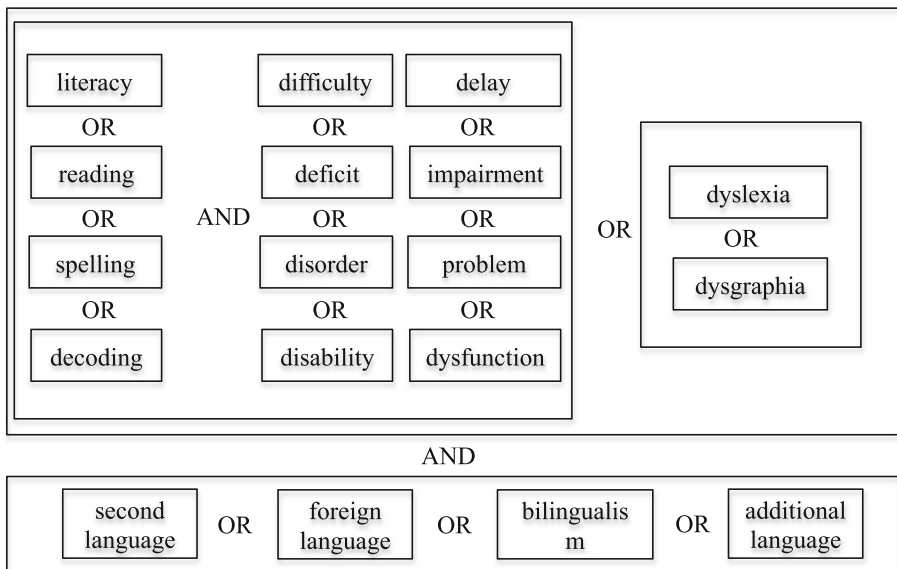


Fig. 1 Search strategy

Data Items

Overall, we collected data on 15 oral and written foreign language tasks (see Table 1). Similar tasks were grouped together (e.g., auditory discrimination of phonemes and syllables as discrimination of speech sounds). Included studies presented data on foreign language outcome measures as continuous data. Therefore, we extracted the mean, standard deviation, and number of participants of the group of children/adolescents with poor literacy skills and the control group for each outcome measure reported in each study. In relation to our second research question, we also gathered information on 11 moderators (see above Table 2).

Risk of Bias in Individual Studies

Two reviewers independently assessed the risk of bias of each included study by applying an adaptation of the ROBINS-I rating scale (Sterne et al. 2016—see Supplemental Materials S5). We judged the quality of each study within the following domains: (a) confounding, (b) selection of participants into the study, (c) classification of participant groups, (d) deviations from intended interventions, (e) missing data, (f) measurement of outcomes, and (g) selection of the reported result. For the last three domains, we did not only assess each study but also each outcome measure. Each domain was judged as having a low, moderate, serious, or critical risk of bias or providing no information in this respect. Again, judgments from two reviewers were juxtaposed, and in case of disagreements, decisions were discussed between all three reviewers, leading to a re-evaluation of the risk of bias in some cases. The entire process was completed in Covidence. As Sterne et al. (2016) suggest that most nonrandomized interventions studies will at least present an overall moderate risk of bias, we decided to only exclude studies with serious or critical risk of bias from further analyses.

Summary Measures

We planned separate meta-analyses for the 15 foreign language outcome measures described in Table 1. However, we only completed analyses if information from at least two studies was available (Borenstein et al. 2009). Following common meta-analytic procedures, we used standard mean differences (*SMDs*) with Hedges correction *g* for small sample sizes (Borenstein et al. 2009) as the unit of analysis. This allowed us to compare the average foreign language performance between children/adolescents with poor and typical literacy skills.

However, we were concerned that this information might not be representative of the performance of individual children/adolescents with poor literacy skills. Based on the heterogeneous performance of poor readers/spellers documented in past research, the extent to which group averages capture individual performances of children/adolescents with poor literacy skills would be expected to vary. Results from meta-analyses solely based on *SMDs* might therefore have limited potential to guide practical implications for individual children/adolescents with poor literacy skills.

To address this limitation, we also computed a second overall effect focusing on the variability across participant groups (the natural logarithm of the ratio between the coefficients of variation of both participant groups—*CVR*; Nakagawa et al. 2015). This allowed us to compare the magnitude of performance variability between poor and typical readers/spellers' foreign performance. Such meta-analytic procedures have recently proven useful to determine the magnitude of intersubject variability in the field of biological evolution and nutrition

(Nakagawa et al. 2015; Senior et al. 2016). To our knowledge, our study is the first to apply this procedure to synthesize evidence on children/adolescents with poor literacy skills so far. Given the well-documented heterogeneity of the study population (Bishop and Snowling 2004; Friedmann and Coltheart 2016; McArthur et al. 2013; Moll and Landerl 2009), adopting this procedure seems justified.

Synthesis of Results

Both types of effect sizes (*SMDs* and *CVRs*) were derived from the mean (*M*), standard deviation (*SD*), and number (*n*) of participants for each foreign language task. Some studies reported information on more than one group of children/adolescents with poor literacy skills or more than one control group. In those cases, we combined the *M*, *SD*, and *n* of both groups or excluded one of the groups (Borenstein et al. 2009).

Furthermore, many studies used more than one task for the same outcome measure (e.g., phoneme deletion and substitution tasks to assess phonological awareness). In these cases, *SMDs* and *CVRs* were computed separately for each task, and subsequently, values were aggregated for each outcome measure (Borenstein et al. 2009). Such aggregation methods take into account the correlation between aggregated tasks. However, as this information was not available for many studies, we assumed a large correlation of $r = 0.50$ (Cohen 1988) based on the similarity of the tasks being aggregated (Borenstein et al. 2009). The same procedure was followed for longitudinal studies reporting more than one data-point per outcome measure (Borenstein et al. 2009).

Before aggregating *SMDs*, we ensured that a negative difference indicated that the control group performed better than the group of children/adolescents with poor literacy skills for all comparisons. If measures were based on the occurrence of errors (instead of accuracy rates), the sign of the *SMDs* was reversed (Borenstein et al. 2009).

Based on all potential moderators that could influence foreign language attainment, we expected to find significant heterogeneity between studies. Therefore, we decided a priori to use random effects modeling to consider the study inverse variance and the between-study variance. We used Cochran's *Q* statistic with a significance level of $p < 0.05$ to determine the presence of heterogeneity among effect sizes. Furthermore, to quantify heterogeneity, we calculated τ^2 and I^2 , as an index of the variation between study effect sizes. We followed the guidelines of Higgins et al. (2003) and considered I^2 values around 25%, 50%, and 75% as low, moderate, and high heterogeneity, respectively. To measure the overall effect, we used *Z* statistics with a Bonferroni-corrected significance level according to the number of comparisons that we were performing (Borenstein et al. 2009). Overall effects were only interpreted in the absence of significant heterogeneity between study effect sizes (*Q* statistic $p > 0.05$).

Risk of Bias Across Studies

In order to assess reporting bias, we completed a funnel plot analysis and applied the trim and fill method by Duval and Tweedie (2000a, b) and Duval (2005), as implemented in the *metafor* package in *R* (Viechtbauer 2010). Following the recommendations of Viechtbauer (2010), we selected the estimator “R0,” as it provides a test of the null hypothesis that the number of missing studies on the chosen side of the funnel plot is zero. We tested this for both sides of the funnel plot.

Moderator Analyses

To explore the impact of specific moderators on the foreign language attainment of children/adolescents with poor literacy skills, we planned to compute separate analyses for 11 moderators (see Table 2). However, in line with Littell et al. (2008), these analyses were only computed if data was available from at least 10 studies. We completed separate random mixed modeling meta-analyses for each moderator subgroup using the *metafor* package in *R* (Viechtbauer 2010). Finally, we used a *Z* test with an adjusted significance level according to the number of comparisons computed to detect significant differences between the overall effects of each moderator subgroup (Borenstein et al. 2009).

Results

Study Selection

The study selection process is illustrated in Fig. 2 and involved two rounds.

Round 1 started with the identification of a total of 1913 study reports, of which 543 duplicates were removed. Titles and abstracts of the remaining 1370 references were screened, and in total, we excluded 1312 studies that did not meet inclusion criteria. We reviewed full texts of the remaining 58 studies and excluded 42 of them on the basis of our inclusion criteria. Therefore, round 1 of the search procedure ended with 16 studies meeting inclusion criteria. In *round 2*, we additionally checked the reference lists of these 16 included studies and found 44 additional references that were imported into Covidence. We also screened the titles and abstracts of studies that cited the 16 included studies in Google Scholar and found 53 new relevant references that were also added to Covidence. Of these 97 new references, 70 duplicates were removed. We screened titles, abstracts, and full texts of the remaining 27 studies, but none of these studies met inclusion criteria and, thus, did not enter the current review.

Study Characteristics

Within the 16 studies included in this review, a total of 968 participants (404 poor readers/spellers and 564 control participants) were assessed. All studies focused on the comparison of children/adolescents with poor and typical literacy skills on at least one oral or written foreign language outcome measure. Studies were completed in the Netherlands (six studies), Italy (3 studies), Hong Kong (3 studies), Norway (2 studies), Poland (1 study), and China (1 study).

Participants In the Supplemental Materials (see S6), we provide details on the information extracted from each study. With respect to participants' native language profile, only three studies reported this information. Bekebrede et al. (2009) and Morfidi et al. (2007) included only children/adolescents with average oral native language skills, but poor written native language skills. None of the studies distinguished between selective reading and spelling deficits, and only two studies listed information on reading difficulty subtypes. While Bekebrede et al. (2009) included children/adolescents with lexical and other reading deficits, Haisma (2009) assessed participants with sublexical and lexical reading difficulties. Information was completely missing for 13 studies.

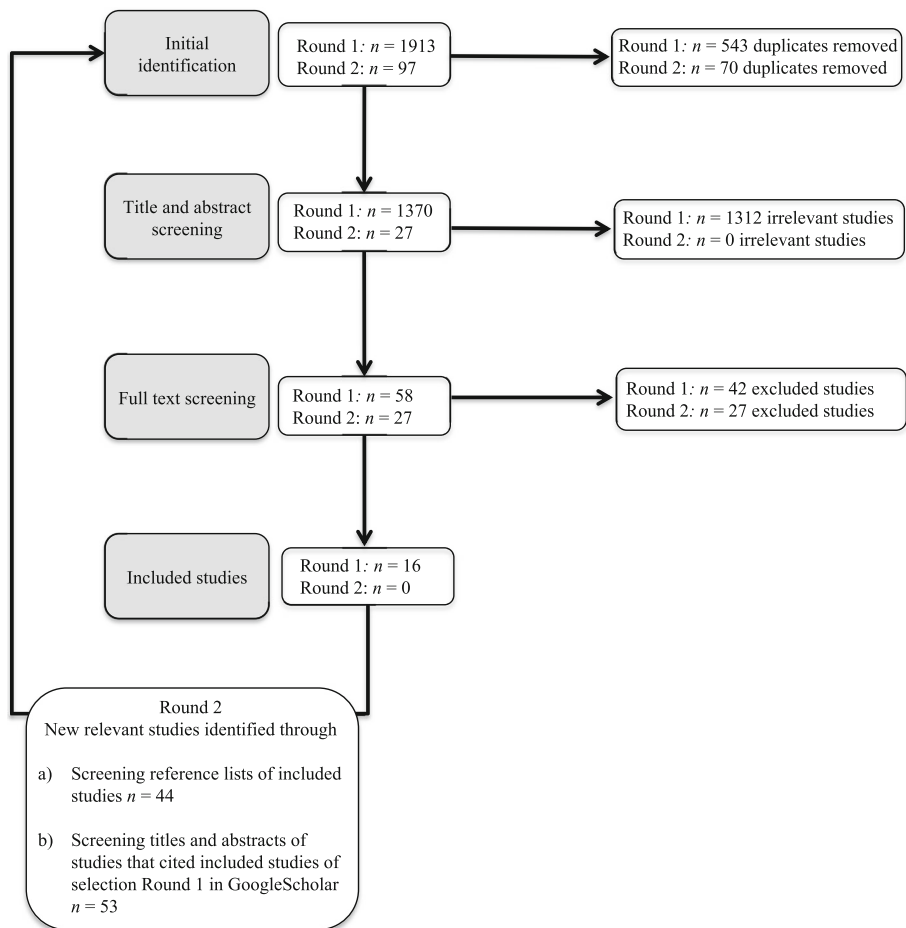


Fig. 2 Flowchart of the search procedure

Concerning linguistic background, four studies reported that their participants had a purely monolingual background. Only Morfidi et al. (2007) stated that their sample also included bilingual children. Information was missing for the remaining 11 studies.

Foreign Language Instruction In three studies, frequency of foreign language instruction consisted of two to four classes per week, and in two studies, participants received between 30 and 60 min of instruction per class. Information was missing for 13 studies.

No language limits were introduced in the search criteria, yet, the foreign language assessed in all studies was English. The native languages of participants were Dutch (six studies), Italian (3 studies), Cantonese (3 studies), Norwegian (2 studies), Polish (1 study), and Mandarin (1 study). Therefore, in terms of language pairings, 12 studies focused on the combination of two Indo-European languages, with alphabetic writing systems. In all cases, the native language was a predominantly regular orthography paired with a predominantly irregular foreign language orthography. In the remaining four studies, the language combination was a non-

Indo-European native language with an ideographic writing system combined with a foreign Indo-European language with an alphabetic writing system.

Onset age of foreign language learning was early childhood (onset age before 6 years) for three studies, late childhood (onset age from 6 to 11 years) for five studies, and adolescence (onset age from 12 to 17 years) for three studies. None of the studies reported on participants in the transition from adolescence to adulthood (onset age from 18 years onwards). Lastly, information was missing for five studies.

Foreign Language Assessment The age at foreign language assessment was late childhood (6–11 years of age) for seven studies and adolescence (12–17 years of age) for nine studies. None of the studies assessed participants in early childhood (before 6 years of age) or the transition from adolescence to adulthood (from 18 years of age onwards).

An overview of the foreign language outcome measures collected by each study can be found in the Supplemental Materials (see S7). No information was available on foreign language speech discrimination and production. Receptive vocabulary knowledge and word production skills were investigated by six and five studies, respectively. One study tested sentence comprehension and production and another one measured short-term memory. Four studies explored phonological awareness, while two studies measured letter knowledge. Word and nonword reading skills were assessed by 14 and seven studies, respectively. Seven studies gathered information on orthographic knowledge and four assessed reading comprehension. Spelling skills were measured by eight studies and two studies tested translation skills.

Excluded Studies

Reasons for excluding studies at the full text screening phase can be found in the Supplemental Materials (see S8).

Risk of Bias within Studies

Figure 3 provides an overview of the risk of bias of the included studies following the ROBINS-I rating scale (Sterne et al. 2016—see S9 for details).

As the information contained in the study reports was not sufficient to assess all of the domains of risk of bias, in many cases, we contacted the authors to obtain additional information.

Confounding All studies had at least a moderate risk of bias in this domain, as they were all nonrandomized control. This means that by default confounding is expected, as different baseline characteristics of participants could have influenced the results (Sterne et al. 2016). However, all studies reported information on the measurement and control of other important confounding domains such as age, socioeconomic status (SES), nonverbal reasoning, or oral native language skills in a reliable and valid manner. Studies also employed pairwise matching between experimental and control participants or statistical adjustment with Bayesian analysis or ANCOVAs.

Selection of Participants into the Study Two conditions had to be fulfilled to reflect a low risk of bias in this domain. First, the start of foreign language instruction and the timing of

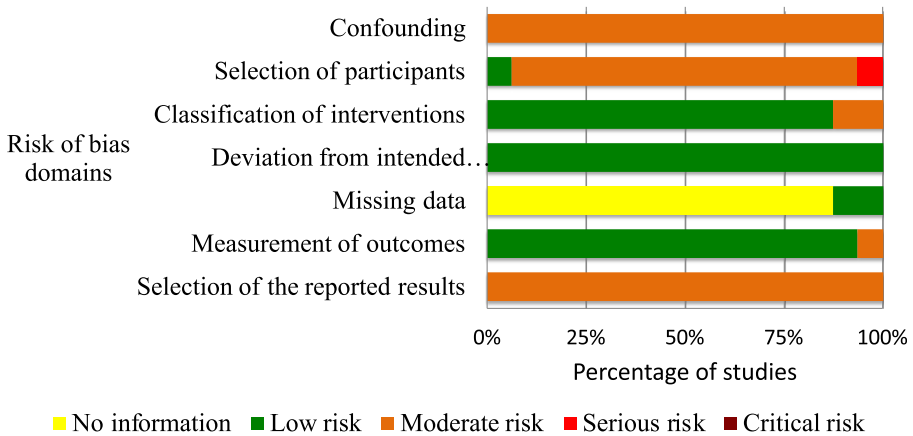


Fig. 3 Overview of risk of bias of included studies ($n = 16$)

foreign language assessment had to be the same for most participants (Sterne et al. 2016). All included studies fulfilled this condition. Second, participant selection had to be unrelated to foreign language attainment (Sterne et al. 2016). This condition was only validated for the Helland and Morken (2016) study, as they explicitly stated that parents of all children in the participating schools were contacted. Therefore, this study received a low risk of bias judgment. In contrast, the remaining studies either did not report how they selected participants into the study or mentioned that school staff (e.g., counselors, teachers, etc.) had selected the participants. This indicates a moderate risk of bias, because school staff could have selected participants based on their knowledge of the participants' foreign language performance. However, all studies, except de Bree and Unsworth (2014), confirmed the literacy status of each of the participants either through an external diagnosis of poor literacy skills or by administering a literacy test. In contrast, de Bree and Unsworth (2014) solely relied on the information of school staff for the selection of participants into the study. Although the authors acknowledged the presence of selection bias in their study in a footnote, this still represents a serious risk of bias. Therefore, we excluded this study from further analysis.

Classification of Participant Groups Two aspects were crucial in this risk of bias domain. First, participant group status had to be well defined (Sterne et al. 2016). Children/adolescents with poor literacy skills had to perform at least 1 *SD*, year, or grade below the expected level on one or more of the following measures: word/nonword reading accuracy, reading fluency, and/or spelling. While all studies fulfilled this condition, Ding et al. (2013) and Helland and Morken (2016) still showed a moderate risk of bias. The group of children/adolescents with poor literacy skills in the study of Ding et al. (2013) was selected because they scored 1 *SD* below the mean of the study sample itself ($n = 102$). This represents a moderate risk, because the identification of children/adolescents with poor literacy skills could have been biased by the characteristics of the sample of total participants from which they were selected. Helland and Morken (2016) identified poor literacy skills based on a below average performance (< 25th percentile) on at least two out of four literacy measures. While for three of these measures independent standardized test norms were available, one (i.e., text reading) was developed by the authors (see Helland et al. 2011, for additional information to Helland and Morken 2016). For this measure, the cutoff criterion (< 25th percentile) was based on the sample of the study

($n = 42$). Therefore, similar to Ding et al. (2013), the identification of children/adolescents with poor literacy skills could have been biased by the characteristics of the specific study sample. Based on this information, we assigned a moderate risk of bias to Helland and Morken (2016).

A second crucial aspect pertaining to classification was that the assignment of participant status should not have been determined retrospectively (Sterne et al. 2016). Although two studies were especially vulnerable with respect to this condition, after a careful analysis, we decided that no risk of bias was present. First, Zhou et al. (2014) completed a longitudinal study in which poor literacy status was defined at the last testing point and retrospectively assigned to two previous testing points at earlier developmental stages. With this risk of bias domain in mind, we had already excluded the first two previous testing points during data extraction. Therefore, no risk of bias was present for the third data-point which was included in this review. Similarly, in their longitudinal study, Helland and Morken (2016) determined poor literacy status at the last testing point and retrospectively assigned to two previous testing points at earlier developmental stages. However, the authors only reported on foreign language outcome measures for the third testing point, when poor literacy status was defined. Therefore, no risk of bias was identified.

Deviation from Intended Interventions All studies showed a low risk of bias in this domain. Łockiewicz and Jaskulska (2016) were the only authors who reported information on a potential risk of bias of deviation from intended interventions in the form of extracurricular private foreign language tutoring. Furthermore, foreign language teachers could have provided educational accommodations to children/adolescents with poor literacy skills if they had knowledge on their students' native language performance (e.g., dyslexia diagnosis). However, both situations of potential risk of bias reflect usual practice and are therefore assigned a low risk of bias (Sterne et al. 2016).

Missing Data We were not able to assess this risk of bias domain for most studies, because no explicit information was given in the studies, with the exception of Helland and Morken (2016) and van Viersen et al. (2017). A low risk was assigned to studies where the number of participants in the results matched the number of participants in the methods section. However, other studies did not provide the number of participants when presenting results. These cases were marked as providing no information to assess this risk of bias domain.

Measurement of Outcomes All studies showed a low risk of bias with the exception of Palladino et al. (2016). In this study, spelling errors were scored according to a predefined grid. Since no information was available regarding the reliability of this measure, we assigned a moderate risk of bias.

Selection of Reported Results We identified a moderate risk of bias for all studies because no preregistered protocols or statistical analysis plans were available for any of the studies (Sterne et al. 2016). However, the information on outcome measurements in the methods and results section of each study report was consistent.

Overall Risk of Bias We determined the overall risk of bias as the highest risk of bias judgment received by a study in any of the domains (Sterne et al. 2016). This was a moderate risk of bias for all studies, with the exception of de Bree and Unsworth (2014). In this study,

we found a serious risk of bias in the selection of participants and, therefore, assigned an overall serious risk of bias. This led to the exclusion of this study.

Meta-analyses of Foreign Language Outcome Measures

Before conducting the analyses, we completed data transformations (e.g., merging data from two data points, two control or experimental groups, etc.). Details are specified in the Supplemental Materials (see S10) and the datasets and code used to compute the analyses can be accessed on the Open Science Framework at <https://osf.io/np97f/>. Overall, we were able to compute separate random effects modeling meta-analyses for 10 out of 15 foreign language outcome measures. Analyses for the remaining five foreign language outcome measures were not possible due to limited data (less than 2 study reports). Results are presented in Table 4.

Meta-analytic results revealed overall *SMDs* and *CVRs* for 10 foreign language outcome measures. However, only four out of 10 overall *SMDs* could be interpreted, due to significant heterogeneity between study effects (Q statistic $p < 0.05$). The interpretable effects concerned foreign language receptive vocabulary knowledge, phonological awareness, letter knowledge, and reading comprehension. For foreign language receptive vocabulary knowledge, children/adolescents with poor literacy skills on average showed a similar performance to their peers with typical literacy skills. In contrast, poor readers/spellers scored poorer on foreign language phonological awareness, letter knowledge, and reading comprehension.

Results based on *SMDs* are derived from group averages, and they may not take into consideration individual differences within each participant group. Therefore, in addition to overall *SMDs*, we computed overall *CVRs*. This complementary analysis allowed us to estimate to what extent results from overall *SMDs* were likely to vary across individual poor readers/spellers. Due to significant heterogeneity between study effects (Q statistic $p < 0.05$), *CVRs* could only be interpreted with respect to two foreign language outcome measures, namely for phonological awareness and reading comprehension. According to these results, poor readers/spellers varied significantly more than control participants in their foreign language phonological awareness performance. In contrast, performance on foreign language reading comprehension varied to a similar extent across participant groups. Figures 4 and 5 depict results for each foreign language outcome measure in the form of forest plots.

Moderator Analyses

Due to serious data limitations (less than 10 studies), only three out of 11 planned moderator analyses (see Table 2) could be conducted on only one of the 15 foreign language outcome measures planned for this review (Littell et al. 2008). Thus, we focused on investigating whether (a) language pairing between native and foreign language, (b) onset age of foreign language instruction, and (c) age at foreign language assessment provided a better understanding of the heterogeneity observed between study effects in foreign language word reading. An adjusted significance level of $p < 0.003$ was applied to correct for the five additional comparisons (on top of the 10 previous comparisons) that were computed as part of the moderator analyses (Borenstein et al. 2009). Results revealed no significant impact of any of the three moderators mentioned above, neither for the *SMD*, nor for the *CVR* of the foreign language word reading performance between children/adolescents with poor and typical literacy skills.

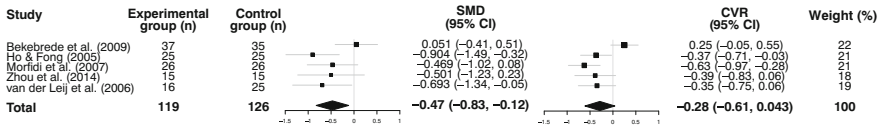
Table 4 Meta-analytic results

FL outcome measure	Number of studies	Participant groups		Mean difference between groups				Variance difference between groups					
		Poor readers/ spellers (<i>n</i>)	Control participants (<i>n</i>)	<i>Q</i>	<i>I</i> ² (%)	SMD	95% CI	<i>Z</i>	<i>Q</i>	<i>I</i> ² (%)	CVR	95% CI	<i>Z</i>
Receptive vocabulary knowledge	5	119	126	7.40 ^{ns}	45.98	-0.47	[-0.82, -0.11]	-2.59 ^{ns}	16.43*	75.66	-0.28	[-0.61, 0.04]	-1.70 ^{ns}
Spoken word production	5	113	188	38.52*	89.62	-1.01	[-1.79, -0.22]	-2.51 ^{ns}	23.80*	83.19	-0.30	[-0.65, -0.05]	-1.68 ^{ns}
Phonological awareness	4	95	104	3.90 ^{ns}	23.17	-1.08	[-1.40, -0.76]	-6.66*	3.03 ^{ns}	1.15	-0.37	[-0.56, -0.19]	-3.98*
Letter knowledge	2	54	54	2.64 ^{ns}	62.16	-1.23	[-1.90, -0.56]	-3.59*	43.96*	97.72	-0.81	[-2.44, 0.83]	-0.97 ^{ns}
Word reading	13	319	455	155.30*	92.27	-1.59	[-2.16, -1.02]	-5.48*	76.93*	84.40	-0.54	[-0.77, -0.31]	-4.63*
Nonword reading	6	169	235	43.91*	88.61	-0.90	[-1.50, -0.30]	-2.93*	9.85 ^{ns}	49.22	-0.18	[-0.36, -0.00]	-2.00 ^{ns}
Orthographic knowledge	6	160	155	31.17*	83.96	-1.42	[-1.60, -1.24]	-15.66*	81.67*	93.87	-0.70	[-1.25, -0.15]	-2.49 ^{ns}
Reading comprehension	4	79	211	3.00 ^{ns}	0.16	-1.02	[-1.29, -0.75]	-7.34*	6.87 ^{ns}	56.37	-0.24	[-0.49, 0.01]	-1.85 ^{ns}
Spelling	8	211	308	46.68*	85.00	-1.33	[-1.82, -0.84]	-5.35*	44.42*	84.24	-0.24	[-0.54, -0.06]	-1.50 ^{ns}
Translation	2	33	48	4.33*	76.92	-1.24	[-1.54, -0.94]	-2.34 ^{ns}	2.89 ^{ns}	65.73	-0.79	[-1.29, -0.29]	-3.09*

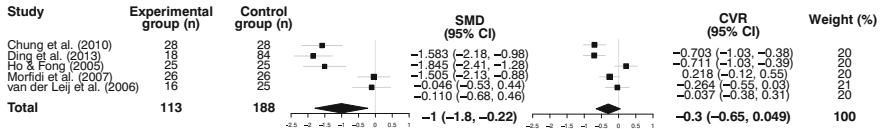
The significance level was **p* < 0.05 for the *Q* statistic and **p* < 0.005 for the *Z* statistic (Bonferroni correction for 10 comparisons). Data in italics show overall effects that could not be interpreted due to the presence of significant between-study heterogeneity

FL = foreign language; SMD = standardized mean difference; CVR = natural logarithm of the ratio of variation coefficients (Nakagawa et al. 2015), ns = non significant

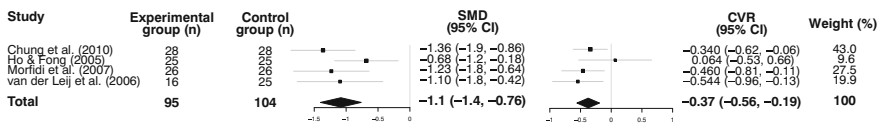
a.Foreign language receptive vocabulary



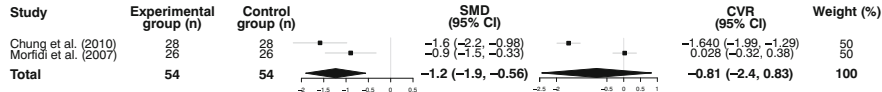
b.Foreign language spoken word production



c.Foreign language phonological awareness



d.Foreign language letter knowledge



e.Foreign word reading

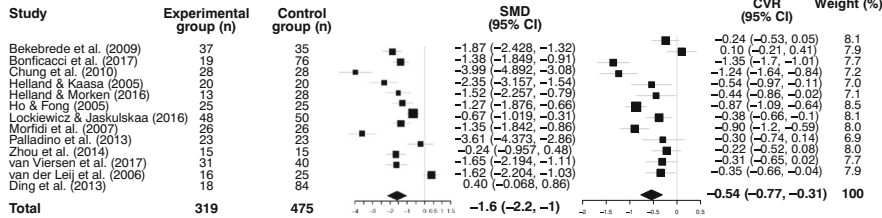


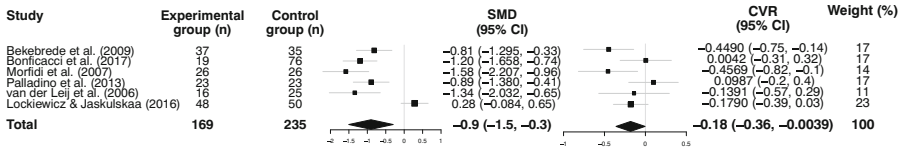
Fig. 4 Meta-analyses on foreign language: **a** receptive vocabulary knowledge, **b** spoken word production, **c** phonological awareness, **d** letter knowledge, and **e** word reading. Note. *SMD* = standardized mean difference; *CVR* = natural logarithm of the ratio of coefficients of variation. (Nakagawa et al. 2015)

The exact values and forest plots reflecting these results are attached in the Supplemental Materials (see S11 to S13).

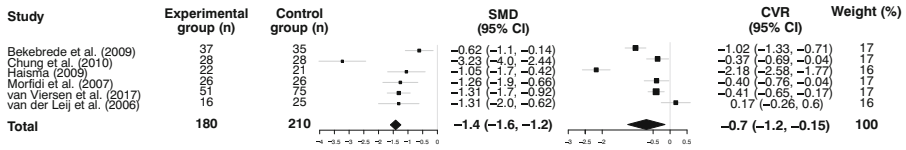
Reporting Bias

The influence of reporting bias could only be investigated with respect to the results of the meta-analysis on the SMDs for foreign language word reading due to limited data. Results of a funnel plot analysis with the trim and fill method by Duval and Tweedie (2000a, b) and Duval (2005), as implemented in the *metafor* package in R (Viechtbauer 2010), showed no evidence of reporting bias. The null hypothesis that the number of missing studies of the funnel plot was zero could not be rejected for the right ($p = 0.25$) or for the left side of the funnel plot ($p = 0.50$; Viechtbauer 2010). The corresponding funnel plot is attached in the Supplemental Materials (see S14).

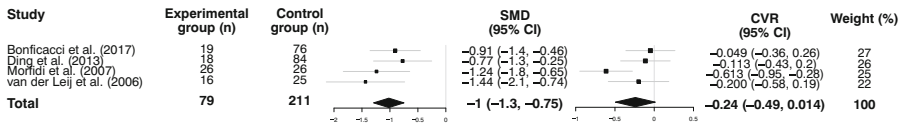
a.Foreign language nonword reading



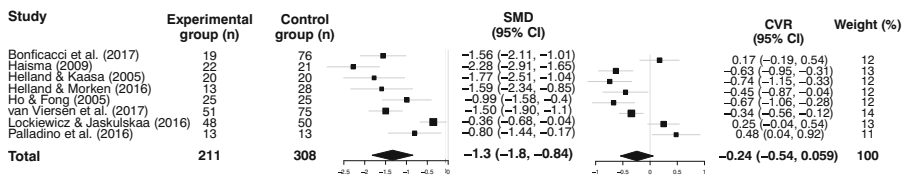
b.Foreign language orthographic knowledge



c.Foreign language reading comprehension



d.Foreign language spelling



e.Foreign language translation

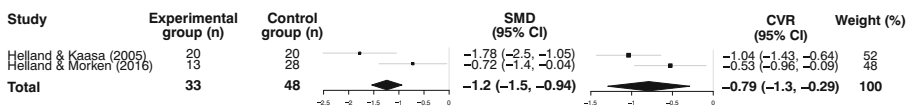


Fig. 5 Meta-analyses on foreign language: **a** nonword reading, **b** orthographic knowledge, **c** reading comprehension, **d** spelling, and **e** translation. Note. *SMD* = standardized mean difference; *CVR* = natural logarithm of the ratio of coefficients of variation. (Nakagawa et al. 2015)

Sensitivity Analysis

In order to assess if meta-analyses were influenced by our risk of bias assessment, we repeated the meta-analyses including the data reported by de Bree and Unsworth (2014), the only study that had been excluded due to a serious risk of bias in the selection of participants. This study contributed information to the outcome measures receptive vocabulary knowledge, word reading, nonword reading, and orthographic knowledge. Results revealed a difference between the analysis with and without the data from de Bree and Unsworth (2014) for receptive vocabulary knowledge, but not for word reading, nonword reading and orthographic knowledge. For receptive vocabulary knowledge, the overall difference in average performance between groups was not significant without de Bree and Unsworth (2014), but reached significance when this study was included. The overall difference between performance variation between groups was not interpretable due to significant heterogeneity with and without de Bree and Unsworth (2014). Overall effects for word reading, nonword reading,

and orthographic knowledge could not be interpreted due to significant heterogeneity with and without de Bree and Unsworth (2014). Results are detailed in the Supplemental Materials (see S15).

Discussion

The structure of this section follows the guidelines provided by the Cochrane Collaboration (Higgins and Green 2008). First, we present a summary of our main results, followed by an analysis of the overall completeness and applicability of the evidence that was summarized. Second, we assess the overall quality of the evidence for each foreign language outcome measure according to the Grades of Recommendation, Assessment, Development and Evaluation guidelines (GRADE, Schünemann et al. 2008). Lastly, we portray implications for practice and future research.

Summary of the Main Results

Figure 6 summarizes the main results of this review.

From a total of 2010 study records initially identified, only 16 (< 1%) met inclusion criteria for the current review. Fifteen study reports displayed low to moderate risk of bias and were thus entered into the meta-analyses. Only one study was excluded due to serious risk of bias. We extracted data on 15 foreign language outcome measures for this review. Meta-analyses could not be conducted for the following five measures due to insufficient data: foreign language discrimination of speech sounds, production of speech sounds, sentence comprehension, sentence production, and short-term memory.

In contrast, we computed separate meta-analyses for the remaining ten foreign language outcome measures: receptive vocabulary knowledge, spoken word production, phonological awareness, letter knowledge, word reading, nonword reading, orthographic knowledge, reading comprehension, spelling, and translation. Results on overall *SMDs* revealed significant between-study heterogeneity for spoken word production, word reading, nonword reading, orthographic knowledge, spelling, and translation, and so the interpretation of these values was not possible. To investigate the source of between-study heterogeneity, we planned to compute moderator analyses on 11 moderators that we assumed could have an impact on foreign language performance in children/adolescents with poor literacy skills. However, this was not possible for foreign language spoken word production, nonword reading, orthographic knowledge, spelling, and translation due to insufficient data. We were able to conduct moderator analyses on foreign language word reading, although the information reported in the included studies was only enough to investigate the impact of three of the 11 moderators addressed in this review (i.e., similarity between native and foreign language, onset age of foreign language instruction, and age at assessment). Information was insufficient for the remaining eight moderators. Results showed that neither the similarity between native and foreign language, onset age of foreign language instruction nor age at assessment could explain the between-study heterogeneity that we found for foreign language word reading.

Overall *SMDs* for foreign language receptive vocabulary knowledge showed that children/adolescents with poor literacy skills on average achieved a similar performance as the control group. In contrast, their average performance was poorer on foreign language phonological awareness, letter knowledge, and reading comprehension as compared with the control group.

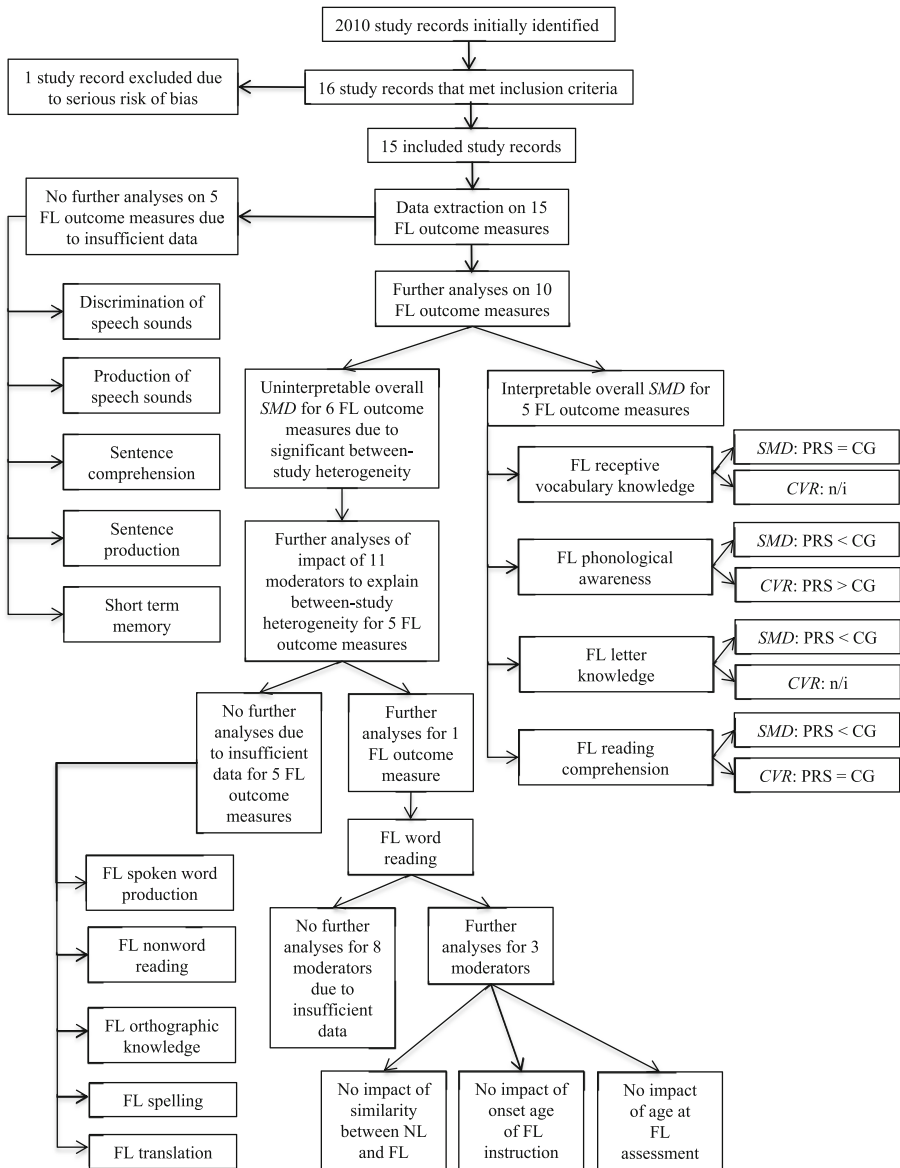


Fig. 6 Summary of main results of this review. Note. *FL* = foreign language; *SMD* = overall standard mean difference; *CVR* = overall coefficient of variation; *PRS* = poor readers/spellers; *CG* = control group; *NL* = native language; *n/i* = not interpretable due to significant between-study heterogeneity

Complementary meta-analyses on the difference of performance variation (*CVR*) revealed that the performance of individual poor readers/spellers in phonological awareness varied significantly more than that of control participants. In contrast, performance in reading comprehension varied to a similar extent across both participant groups. Comparisons of overall performance variation between poor readers/spellers and control participants could not be

determined for foreign language receptive vocabulary knowledge and letter knowledge, as *CVR* results reflected significant between-study heterogeneity.

Overall Completeness and Applicability of Evidence

Overall completeness and applicability of the evidence summarized in this review is limited in three ways. First, the information summarized in this review only refers to some foreign language outcome measures (i.e., receptive vocabulary knowledge, word production, phonological awareness, letter knowledge, word reading, nonword reading, orthographic knowledge, reading comprehension, spelling, and translation). In contrast, insufficient information was found for other foreign language outcome measures (i.e., foreign language discrimination and production of speech sounds, sentence comprehension, and production skills). Hence, the findings are restricted to a small set of foreign language skills.

Second, foreign language attainment of children/adolescents with poor literacy skills was investigated with native speakers of a variety of Indo-European and non-Indo-European native languages. However, information on some of the most widely spoken native languages such as English and Spanish is currently missing. Furthermore, while no language limits were introduced in the search criteria, the foreign language assessed in all studies was English. This limits the conclusions that can be drawn from this review with respect to other languages. Considering that the English orthography has been characterized as an “outlier orthography” due to its irregular grapheme–phoneme mappings (Share 2008), it is unclear how generalizable these results are to more “regular” orthographies.

Lastly, and most importantly, significant heterogeneity between study effects seriously limited the interpretation of available evidence for several foreign language outcome measures (i.e., foreign language spoken word production, word reading, nonword reading, orthographic knowledge, spelling, and translation). While most studies indicated a trend toward a lower foreign language attainment of children/adolescents with poor literacy skills as compared with the control group, the magnitude of this effect varied significantly from study to study. Where does this heterogeneity come from? In the registered protocol of this review, we anticipated 11 moderators that could represent potential sources of heterogeneity. Yet, due to limited data, we were only able to investigate the impact of three moderators (i.e., similarity between native and foreign language, onset age of foreign language instruction, and age at assessment) on only one of 15 foreign language measures (i.e., foreign language word reading). While none of these three moderators contributed to explaining the observed between-study heterogeneity, it is possible that some of the other eight moderators, for which we did not have enough data, influenced the results.

Quality of the Evidence and Potential Biases in the Review Process

To assess the overall quality of the evidence for each of the foreign language outcome measures, we applied the GRADE guidelines (Schünemann et al. 2008). Following the guidelines, we began by judging all studies as “low quality of evidence,” because they were all nonrandomized trials. The evidence for each of the foreign language outcome measures could later be changed (i.e., upgraded or downgraded) following the criteria of Schünemann et al. (2008). We detail the specific reasons for each decision below.

For foreign language reading comprehension, we upgraded the evidence to a level of moderate quality of evidence, due to the overall large *SMD* in the absence of significant

heterogeneity between studies. Furthermore, children/adolescents with poor and typical literacy skills varied to a similar extent in their performance. In contrast, we maintained a judgment of low quality for the evidence on foreign language phonological awareness and letter knowledge. Despite obtaining large overall *SMDs* in the absence of significant between-study variance, it was unclear to what extent *SMDs* were representative of the performance of individual poor reader/speller participants. In the case of foreign language letter knowledge, the reasons for this was a heterogeneous *CVR*. In contrast, for foreign language phonological awareness, results showed a higher performance variation in the poor reader/speller participant group than in the control group.

Finally, we downgraded the quality of the evidence on foreign language receptive vocabulary knowledge, spoken word production, word reading, nonword reading, orthographic knowledge, spelling, and translation to very low. For foreign language receptive vocabulary knowledge, we conducted a sensitivity analysis with and without the data of de Bree and Unsworth (2014). This was the only study that was excluded due to a serious risk of bias in the selection of participants. The results of the sensitivity analysis were inconsistent. Therefore, the evidence on foreign language receptive vocabulary knowledge in children/adolescents with poor literacy skills should be interpreted with caution. For foreign language spoken word production, word reading, nonword reading, orthographic knowledge, spelling, and translation, overall effects were not interpretable due to significant between-study variance. Moderator analyses for foreign language word reading did not contribute to explaining the observed variability, and so results should also be interpreted with caution. No potential publication biases were identified.

Implications for Practice

The results of this review provide evidence that children/adolescents with poor literacy skills on average show a lower attainment than their peers with typical literacy skills in foreign language phonological awareness, letter knowledge, and reading comprehension. However, we also found evidence of higher performance variation in foreign language attainment of poor readers/spellers than of control participants for phonological awareness. The sources of this variability have so far not been addressed by current research. Therefore, although children/adolescents with poor literacy skills seem to have a greater risk of experiencing foreign language learning difficulties, this might not be true for every individual child or adolescent with poor literacy skills. The common belief that children/adolescents with poor literacy skills show a lower foreign language attainment than their peers with typical literacy skills cannot be confirmed by the results of this review. Parents, teachers, and clinicians should keep in mind that an individual student with poor literacy skills might be just as successful as other students with typical literacy skills. Instead of relying on the false common belief that all poor readers/spellers will struggle in learning a foreign language, foreign language attainment should be closely monitored and support put in place when necessary.

Implications for Research

Available evidence on foreign language attainment in children/adolescents with poor literacy skills shows several limitations. Most importantly, future research should aim to better understand individual differences in foreign language attainment of children/adolescents with poor literacy skills. First, an investigation of the impact of moderators related to participant

characteristics might aid in better understanding the variability observed in foreign language attainment in children/adolescents with poor literacy skills. Past research has described heterogeneous native language profiles of children/adolescents with poor literacy skills that are likely to contribute to individual differences also in foreign language attainment (Bishop and Snowling 2004; Friedmann and Coltheart 2016; McArthur et al. 2013; Moll and Landerl 2009). Similarly, participants' linguistic background has been related to individual differences in foreign language attainment and is likely to be a source of the heterogeneity of foreign language attainment in poor readers/spellers reflected/observed in the results of this review (Nair et al. 2016; Tremblay and Sabourin 2012).

Second, information on the frequency, duration, and onset age of foreign language instruction should be reported. This would make it possible to assess the impact of these variables in future meta-analyses. Furthermore, studies are needed to investigate foreign language attainment in children/adolescents with poor literacy skills learning a foreign language other than English. Although English is undoubtedly the most frequently instructed foreign language worldwide, we need to know if the difficulties observed in children/adolescents with poor literacy skills in learning English as a foreign language extend to other foreign languages. Related to this, future studies should also assess children/adolescents with poor and typical literacy skills that speak some of the most frequently spoken languages, such as English and Spanish, as native language. This would allow for a more representative picture of foreign language attainment in poor readers/spellers for a large number of the world's population.

Finally, future meta-analyses focusing on heterogeneous populations, such as children/adolescents with poor literacy skills, should include computations related to variation of outcome measures to capture individual differences. The common practice of only computing overall effect sizes based on central tendency measures such as the SMD between groups can result in misleading answers to practically relevant research questions. For example, in the current review, only relying on overall SMDs would have led to the conclusion that children/adolescents with poor literacy skills show a lower foreign language attainment than their peers with typical literacy skills. However, the fact that foreign language performance varied more in children/adolescents with poor literacy skills than in control participants emphasizes that this conclusion might not apply to a significant proportion of poor reader/spellers.

Conclusions

This report provides the first systematic synthesis of the available evidence addressing the question if children/adolescents with poor literacy skills show a lower foreign language attainment than their peers with typical literacy skills. This represents a unique contribution, because the success achieved by poor readers/spellers in different foreign language outcome measures as compared with typical readers/spellers has so far only been investigated in individual studies. We therefore contribute a basis to understand the results of individual studies in a broader context and increase statistical power to estimate overall effects (Borenstein et al. 2009). Furthermore, this systematic review provides an overview of the available evidence on this issue which can be useful to researchers in guiding future studies, as well as parents, teachers/specialists, and policy makers in guiding educational decision-making.

Acknowledgments We would like to thank Deanna Francis and Nicholas Badcock for sharing their experience in conducting a systematic review and introducing us to key methodological resources. Furthermore, we would like to thank John Elias for reviewing our search strategy and guiding us in the selection and usage of databases.

Code Availability The code used to compute the analyses mentioned in this report are freely available on the Open Science Framework at <https://osf.io/np97f/>.

Author's Contributions Alexa von Hagen: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, visualization, writing—original draft, and writing—review and editing. Saskia Kohnen: conceptualization, investigation, methodology, and writing—review and editing. Nicole Stadie: conceptualization, investigation, methodology, and writing—review and editing.

Funding Open Access funding provided by Projekt DEAL. The work presented in this manuscript was carried out as part of Alexa von Hagen's PhD candidature in the Erasmus Mundus joint International Doctorate for Experimental Approaches to Language and Brain (IDEALAB) and was funded by the European Commission within the action nr. 2015-1603/001-001-EMJD (Framework Partnership Agreement 2012–2025).

Data Availability The supplementary materials and data mentioned in this report are freely available on the Open Science Framework at <https://osf.io/np97f/>.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

*References marked with an asterisk indicate studies included in the systematic review

- Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: listener perception versus linguistic scrutiny. *Language Learning*, 59(2), 249–306. <https://doi.org/10.1111/j.1467-9922.2009.00507.x>.
- Barbiero, C., Lonciari, I., Montico, M., Monasta, L., Penge, R., Vio, C., Tressoldi, P. E., Ferluga, V., Bigoni, A., Tullio, A., Carozzi, M., & Rofani, L. (2012). The submerged dyslexia iceberg: how many school children are not diagnosed? Results from an Italian study. *PLoS One*, 7(10), e48082. <https://doi.org/10.1371/journal.pone.0048082>.
- *Bekebrede, J., van der Leij, A., & Share, D. L. (2009). Dutch dyslexic adolescents: phonological-core variable-orthographic differences. *Reading and Writing*, 22(2), 133–165. <https://doi.org/10.1007/s11145-007-9105-7>.
- Bialystok, E. (1997). The structure of age: in search of barriers to second language acquisition. *Second Language Research*, 13(2), 116–137. <https://doi.org/10.1191/026765897677670241>.

- Bialystok, E. (2012). The impact of bilingualism on language and literacy development. In W. C. Ritchie & T. K. Bhatia (Eds.), *The handbook of bilingualism and multilingualism* (2nd ed., pp. 624–648). London: Blackwell.
- Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied PsychoLinguistics*, 19(01), 69–85. <https://doi.org/10.1017/S0142716400010584>.
- Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: interactions among languages and writing systems. *Scientific Studies of Reading*, 9(1), 43–61. https://doi.org/10.1207/s1532799xssr0901_4.
- Bishop, D. V., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: same or different? *Psychological Bulletin*, 130(6), 858–886. <https://doi.org/10.1037/0033-2909.130.6.858>.
- *Bonifacci, P., Canducci, E., Gravagna, G., & Palladino, P. (2017). English as a foreign language in bilingual language-minority children, children with dyslexia and monolingual typical readers. *Dyslexia*, 23(2), 181–206. <https://doi.org/10.1002/dys.1553>.
- Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.
- *Chung, K. K. H., & Ho, C. S. H. (2010). Second language learning difficulties in Chinese children with dyslexia: what are the reading-related cognitive skills that contribute to English and Chinese word reading? *Journal of Learning Disabilities*, 43(3), 195–211. <https://doi.org/10.1177/0022219409345018>.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). New York: Academic.
- Connor, U. (1996). *Contrastive rhetoric: cross-cultural aspects of second language writing*. Cambridge: Cambridge University Press.
- de Bot, K. (1992). A bilingual production model: Levelt's "speaking" model adapted. *Applied Linguistics*, 13(1), 1–24.
- de Bot, K., Paribakht, T. S., & Wesche, M. B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition: evidence from ESL reading. *Studies in Second Language Acquisition*, 19(3), 309–329.
- *de Bree, E., & Unsworth, S. (2014). Dutch and English literacy and language outcomes of dyslexic students in regular and bilingual secondary education. *Dutch Journal of Applied Linguistics*, 3(1), 62–81. <https://doi.org/10.1075/dujal.3.1.04bre>.
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22(04), 499–533.
- DeKeyser, R., Alfi-Shabtab, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied PsychoLinguistics*, 31(03), 413–438. <https://doi.org/10.1017/S0142716410000056>.
- *Ding, Y., Guo, J. P., Yang, L. Y., Zhang, D., Ning, H., & Richman, L. C. (2013). Rapid automatized naming and immediate memory functions in Chinese children who read English as a second language. *Journal of Learning Disabilities*, 46(4), 347–362. <https://doi.org/10.1177/0022219411424209>.
- Dixon, L. Q., Zhao, J., Shin, J. Y., Wu, S., Su, J. H., Burgess-Brigham, R., Gezer, M. U., & Snow, C. (2012). What we know about second language acquisition: a synthesis from four perspectives. *Review of Educational Research*, 82(1), 5–60. <https://doi.org/10.3102/0034654311433587>.
- Duval, S. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: prevention, assessment and adjustments* (pp. 128–144). West Sussex: Wiley.
- Duval, S. J., & Tweedie, R. L. (2000a). A nonparametric 'trim and fill' method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449), 89–98. <https://doi.org/10.1080/01621459.2000.10473905>.
- Duval, S. J., & Tweedie, R. L. (2000b). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>.
- Elliott, J. G., & Grigorenko, E. L. (2014). *The dyslexia debate*. New York: Cambridge University Press.
- Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences*, 99(1), 529–534. <https://doi.org/10.1073/pnas.012611199>.
- Friedmann, N., & Coltheart, M. (2016). Types of developmental dyslexia. In A. Baron & D. Ravid (Eds.), *Handbook of communication disorders: theoretical, empirical, and applied linguistics perspectives*. De Gruyter Mouton: Berlin.
- Friedmann, N., & Lukov, L. (2008). Developmental surface dyslexias. *Cortex*, 44(9), 1146–1160. <https://doi.org/10.1016/j.cortex.2007.09.005>.
- Friedmann, N., & Rahamim, E. (2007). Developmental letter position dyslexia. *Journal of Neuropsychology*, 1(2), 201–236. <https://doi.org/10.1348/174866407X204227>.

- Frost, R. (2012). A universal approach to modeling visual word recognition and reading: not only possible, but also inevitable. *Behavioral and Brain Sciences*, 35(5), 310–329. <https://doi.org/10.1017/S0140525X12000635>.
- Goulandris, N. E. (2003). *Dyslexia in different languages: cross-linguistic comparisons*. London: Whurr Publishers.
- *Haisma, J. (2009). Dyslexic subtypes and literacy skills in L2 opaque English: Met Name op het Gebied van het Leren van Verbaal Gedrag/taal. *Toegepaste Taalwetenschap in Artikelen*, 81(1), 65–74. <https://doi.org/10.1075/twvia.81.07hai>.
- *Helland, T., & Kaasa, R. (2005). Dyslexia in English as a second language. *Dyslexia (Chichester, England)*, 11(1), 41–60. <https://doi.org/10.1002/dys.286>.
- *Helland, T., & Morken, F. (2016). Neurocognitive development and predictors of L1 and L2 literacy skills in dyslexia: a longitudinal study of children 5–11 years old. *Dyslexia*, 22(1), 3–26. <https://doi.org/10.1002/dys.1515>.
- Helland, T., Plante, E., & Hugdahl, K. (2011). Predicting dyslexia at age 11 from a risk index questionnaire at age 5. *Dyslexia*, 17(3), 207–226. <https://doi.org/10.1002/dys.432>.
- Higgins, J. P., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions: Cochrane book series*. Chichester: Wiley.
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, 327(7414), 557–560. <https://doi.org/10.1136/bmj.327.7414.557>.
- *Ho, C. S. H., & Fong, K. M. (2005). Do Chinese dyslexic children have difficulties learning English as a second language? *Journal of Psycholinguistic Research*, 34(6), 603–618. <https://doi.org/10.1007/s10936-005-9166-1>.
- International Dyslexia Association. (2012). Definition of dyslexia. Available at <https://dyslexiaida.org/definition-of-dyslexia/> [accessed 20 Oct 2016].
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21(1), 60–99. [https://doi.org/10.1016/0010-0285\(89\)90003-0](https://doi.org/10.1016/0010-0285(89)90003-0).
- Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: the orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Advances in psychology, Orthography, phonology, morphology, and meaning* (Vol. 94, pp. 67–84). Oxford: North-Holland.
- Kohnen, S., Nickels, L., Castles, A., Friedmann, N., & McArthur, G. (2012). When ‘slime’ becomes ‘smile’: developmental letter position dyslexia in English. *Neuropsychologia*, 50(14), 3681–3692. <https://doi.org/10.1016/j.neuropsychologia.2012.07.016>.
- Kohnen, S., Colenbrander, D., Krajenbrink, T., & Nickels, L. (2015). Assessment of lexical and non-lexical spelling in students in grades 1–7. *Australian Journal of Learning Difficulties*, 20(1), 15–38. <https://doi.org/10.1080/19404158.2015.1023209>.
- Kohnen, S., Nickels, L., Geigis, L., Coltheart, M., McArthur, G., & Castles, A. (2018). Variations within a subtype: developmental surface dyslexias in English. *Cortex*, 106, 151–163. <https://doi.org/10.1016/j.cortex.2018.04.008>.
- Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, 12(1), 48–54. <https://doi.org/10.1080/19404158.2015.1023209>.
- Landerl, K., Wimmer, H., & Frith, U. (1996). The impact of orthographic consistency on dyslexia: a German-English comparison. *Cognition*, 63(3), 315–334.
- Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. New York: Oxford University Press.
- *Łockiewicz, M., & Jaskulska, M. (2016). Difficulties of Polish students with dyslexia in reading and spelling in English as L2. *Learning and Individual Differences*, 51, 256–264. <https://doi.org/10.1016/j.lindif.2016.08.037>.
- McArthur, G., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., Banales, E., & Castles, A. (2013). Getting to grips with the heterogeneity of developmental dyslexia. *Cognitive Neuropsychology*, 30(1), 1–24. <https://doi.org/10.1080/02643294.2013.784192>.
- Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: a meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, 140(2), 409–433. <https://doi.org/10.1037/a0033890>.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>.
- Moll, K., & Landerl, K. (2009). Double dissociation between reading and spelling deficits. *Scientific Studies of Reading*, 13(5), 359–382. <https://doi.org/10.1080/10888430903162878>.

- Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., et al. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction, 29*, 65–77. <https://doi.org/10.1016/j.learninstruc.2013.09.003>.
- *Morfidi, E., Van Der Leij, A., De Jong, P. F., Scheltinga, F., & Bekebrede, J. (2007). Reading in two orthographies: a cross-linguistic study of Dutch average and poor readers who learn English as a second language. *Reading and Writing, 20*(8), 753–784. <https://doi.org/10.1007/s11145-006-9035-9>.
- Müller, B., Ahnert, P., Burkhardt, J., Brauer, J., Czepezauer, I., Quente, E., Boltze, J., Wilcke, A., & Kirsten, H. (2014). Genetic risk variants for dyslexia on chromosome 18 in a German cohort. *Genes, Brain and Behavior, 13*(3), 350–356. <https://doi.org/10.1111/gbb.12118>.
- Nair, V. K., Biedermann, B., & Nickels, L. (2016). Consequences of late bilingualism for novel word learning: evidence from Tamil–English bilingual speakers. *International Journal of Bilingualism, 20*(4), 473–487. <https://doi.org/10.1177/1367006914567005>.
- Nakagawa, S., Poulin, R., Mengersen, K., Reinhold, K., Engqvist, L., Lagisz, M., & Senior, A. M. (2015). Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution, 6*(2), 143–152. <https://doi.org/10.1111/2041-210X.12309>.
- Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: evidence from component skills. *Language and Cognitive Processes, 18*(4), 443–468. <https://doi.org/10.1080/01690960344000008>.
- Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Oxford, England: Cambridge University Press.
- *Palladino, P., Bellagamba, I., Ferrari, M., & Cornoldi, C. (2013). Italian children with dyslexia are also poor in reading English words, but accurate in reading English pseudowords. *Dyslexia, 19*(3), 165–177. <https://doi.org/10.1002/dys.1456>.
- *Palladino, P., Cismondo, D., Ferrari, M., Ballagamba, I., & Cornoldi, C. (2016). L2 spelling errors in Italian children with dyslexia. *Dyslexia, 22*(2), 158–172. <https://doi.org/10.1002/dys.1522>.
- Saito, K., & Hanzawa, K. (2016). Developing second language oral ability in foreign language classrooms: the role of the length and focus of instruction and individual differences. *Applied PsychoLinguistics, 37*(04), 813–840. <https://doi.org/10.1177/1362168816679030>.
- Schünemann, H. J., et al. (2008). Interpreting results and drawing conclusions. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions: Cochrane book series* (pp. 187–241). Chichester: Wiley.
- Senior, A. M., Gosby, A. K., Lu, J., Simpson, S. J., & Raubenheimer, D. (2016). Meta-analysis of variance: an illustration comparing the effects of two dietary interventions on variability in weight. *Evolution, Medicine, and Public Health, 2016*(1), 244–255. <https://doi.org/10.1093/emphe/eow020>.
- Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*(2), 143–174. <https://doi.org/10.1348/000712603321661859>.
- Share, D. L. (2008). On the Anglocentricities of current reading research and practice: the perils of overreliance on an “outlier” orthography. *Psychological Bulletin, 134*(4), 584–615. <https://doi.org/10.1037/0033-2909.134.4.584>.
- Shaywitz, S. E., Morris, R., & Shaywitz, B. A. (2008). The education of dyslexic children from childhood to young adulthood. *Annual Review of Psychology, 59*(1), 451–475. <https://doi.org/10.1146/annurev.psych.59.103006.093633>.
- Siegel, L. S. (1988). Definitional and theoretical issues and research on learning disabilities. *Journal of Learning Disabilities, 21*(5), 264–266.
- Siegel, L. S. (2007). Perspectives on dyslexia. *Paediatrics & Child Health, 11*, 581–588.
- Snowling, M. J., Duff, F., Petrou, A., Schiffeldrin, J., & Bailey, A. M. (2011). Identification of children at risk of dyslexia: the validity of teacher judgements using ‘phonetic phases’. *Journal of Research in Reading, 34*(2), 157–170. <https://doi.org/10.1111/j.1467-9817.2011.01492.x>.
- Sotiropoulos, A., & Hanley, J. R. (2017). Developmental surface and phonological dyslexia in both Greek and English. *Cognition, 168*, 205–216. <https://doi.org/10.1016/j.cognition.2017.06.024>.
- Sparks, R. L. (2016). Myths about foreign language learning and learning disabilities. *Foreign Language Annals, 49*(2), 252–270. <https://doi.org/10.1111/flan.12196>.
- Sprenger-Charolles, L., Siegel, L. S., Jimenez, J. E., & Ziegler, J. C. (2011). Prevalence and reliability of phonological, surface, and mixed profiles in dyslexia: a review of studies conducted in languages varying in orthographic depth. *Scientific Studies of Reading, 15*(6), 498–521. <https://doi.org/10.1080/10888438.2010.524463>.
- Sterne, J. A. C., Higgins, J. P. T., Elbers, R. G., Reeves, B. C., & the development group for ROBINS-I. (2016). *Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I): detailed guidance*, updated 12 October 2016. Available at <http://www.riskofbias.info> [accessed 25 Oct 2016].

- Tremblay, M. C., & Sabourin, L. (2012). Comparing behavioral discrimination and learning abilities in monolinguals, bilinguals and multilinguals. *The Journal of the Acoustical Society of America*, *132*(5), 3465–3474. <https://doi.org/10.1121/1.4756955>.
- *van der Leij, A., & Morfidi, E. (2006). Core deficits and variable differences in Dutch poor readers learning English. *Journal of Learning Disabilities*, *39*(1), 74–90. <https://doi.org/10.1177/00222194060390010701>.
- *van Viersen, S., de Bree, E. H., Verdam, M., Krikhaar, E., Maassen, B., van der Leij, A., & de Jong, P. F. (2017). Delayed early vocabulary development in children at family risk of dyslexia. *Journal of Speech, Language, and Hearing Research*, *60*(4), 937–949. https://doi.org/10.1044/2016_JSLHR-L-16-0031.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>.
- Wight, M. C. S. (2015). Students with learning disabilities in the foreign language learning environment and the practice of exemption. *Foreign Language Annals*, *48*(1), 39–55. <https://doi.org/10.1111/flan.12122>.
- Wright, W. E. (2013). Bilingual education. In W. C. Ritchie & T. K. Bhatia (Eds.), *The handbook of bilingualism and multilingualism* (2nd ed., pp. 624–648). London: Blackwell.
- *Zhou, Y., McBride-Chang, C., Law, A. B. Y., Li, T., Cheung, A. C. Y., Wong, A. M. Y., & Shu, H. (2014). Development of reading-related skills in Chinese and English among Hong Kong Chinese children with and without dyslexia. *Journal of Experimental Child Psychology*, *122*, 75–91. <https://doi.org/10.1016/j.jecp.2013.12.003>.
- Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: language-specific or universal? *Journal of Experimental Child Psychology*, *86*(3), 169–193. [https://doi.org/10.1016/S0022-0965\(03\)00139-5](https://doi.org/10.1016/S0022-0965(03)00139-5).
- Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faisca, L., Saine, N., Lyytinen, H., Vaessen, A., & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: a cross-language investigation. *Psychological Science*, *21*(4), 551–559. <https://doi.org/10.1177/0956797610363406>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Alexa von Hagen¹ · Saskia Kohnen² · Nicole Stadie³

Saskia Kohnen
saskia.kohnen@mq.edu.au

Nicole Stadie
nstadie@uni-potsdam.de

¹ Competence Centre School Psychology Hesse, Department of Educational Psychology, Goethe-University Frankfurt, PEG Building (HP 68), Theodor-W.-Adorno-Platz 6, 60629 Frankfurt am Main, Germany

² Department of Cognitive Science, Australian Hearing Hub, Macquarie University, 16 University Avenue, Sydney, NSW 2109, Australia

³ Department Linguistik, Humanwissenschaftliche Fakultät, Universität Potsdam, Haus 14, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany