

## Supplemental materials

## S1. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
<b>TITLE</b>			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	p.1
<b>ABSTRACT</b>			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	p.2
<b>INTRODUCTION</b>			
Rationale	3	Describe the rationale for the review in the context of what is already known.	p.3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	pp.4-7
<b>METHODS</b>			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	p.12
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	pp.4-7 & 12-13
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	p.13
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	pp.13-14 & S3
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	pp.13-14
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	p.14
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	p.14 & S3
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	p.14-15 & S4
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	p.16-17

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY SKILLS

Section/topic	#	Checklist item	Reported on page #
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I <sub>2</sub> ) for each meta-analysis.	pp.16-17
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	p.18
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	p.18
<b>RESULTS</b>			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	pp.19 & S8
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	pp.19-20 & S6-S7
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	pp.21-25 & S9
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	pp.25-27
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	pp.25-27
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	p.28 & S14
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	pp.27-28 & S11-S13, S15
<b>DISCUSSION</b>			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	pp.29-31
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	pp.31-33
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	pp.33-35

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY SKILLS

FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Separate declaration page

*Note.* Extracted from Moher, Liberati, Tetzlaff & Altman (2009).

**S2.** Difference between registered protocol and final report.

1. We revised the introduction section in order to improve overall clarity. Changes focused on wording and order of the presented information, as well as the inclusion of supporting references. No new information was added.

2. Foreign language outcome measures were considered primary outcomes of this review, instead of moderators, as presented in the protocol. The reason for this modification was that at the protocol stage we had planned to perform one overall meta-analysis including all foreign language outcome measures. During the review process we decided to complete separate meta-analyses for each foreign language outcome measure, to capture the available evidence in a more detailed way. Therefore, in the final report, the foreign language outcomes measures are presented as primary outcomes, instead of moderators, as detailed in the protocol.

3. We re-structured the classification of moderators presented in the protocol. The broad classification between moderators related to participant characteristics, foreign language instruction and foreign language assessment was maintained. However, onset age of foreign language instruction and language pairing between native and foreign language are now considered foreign language instruction moderators and not participant characteristics moderators. Furthermore, age at foreign language assessment is included as a foreign language assessment moderator in this final report. We decided to perform these changes to improve overall clarity of the structure of data extraction and analysis.

4. The foreign language outcome measure reading was broken down into the following four measures: (a) letter knowledge, (b) word reading, (c) nonword reading and (d) orthographic knowledge. With this change, we aimed to capture more detail with respect to the available evidence on foreign language reading subskills in children/adolescents with poor literacy skills.

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY SKILLS

5. We re-named the foreign language outcome measure "vocabulary knowledge" as "receptive vocabulary knowledge" to make the contrast with "spoken word production" clearer and avoid potential misunderstandings.

6. In addition to synthesizing available data through standardized mean differences (*SMD*), as described in the protocol, we decided to also complete meta-analyses on the natural logarithm of the ratio of coefficients of variation (*CVR*) between participant groups (Nakagawa et al., 2015). This allowed us to capture the average difference in the performance of both participant groups, but also to assess performance variation.

### **S3. Search strategies**

#### **OVID databases, searched on 10<sup>th</sup> February 2017**

1. literacy difficult\$.ti. or literacy difficult\$.ab.
2. decoding difficult\$.ti. or decoding difficult\$.ab.
3. reading difficult\$.ti. or reading difficult\$.ab.
4. spelling difficult\$.ti. or spelling difficult\$.ab.
5. literacy deficit.ti. or literacy deficit.ab.
6. decoding deficit.ti. or decoding deficit.ab.
7. reading deficit.ti. or reading deficit.ab.
8. spelling deficit.ti. or spelling deficit.ab.
9. literacy disorder.ti. or literacy disorder.ab.
10. decoding disorder.ti. or decoding disorder.ab.
11. spelling disorder.ti. or spelling disorder.ab.
12. literacy disabilit\$.ti. or literacy disabilit\$.ab.
13. decoding disabilit\$.ti. or decoding disabilit\$.ab.
14. reading disabilit\$.ti. or reading disabilit\$.ab.
15. spelling disabilit\$.ti. or spelling disabilit\$.ab.
16. literacy delay.ti. or literacy delay.ab.
17. decoding delay.ti. or decoding delay.ab.
18. reading delay.ti. or reading delay.ab.
19. spelling delay.ti. or spelling delay.ab.
20. literacy impairment.ti. or literacy impairment.ab.
21. decoding impairment.ti. or decoding impairment.ab.
22. reading impairment.ti. or reading impairment.ab.
23. spelling impairment.ti. or spelling impairment.ab.
24. literacy problem\$.ti. or literacy problem\$.ab.
25. decoding problem\$.ti. or decoding problem\$.ab.
26. reading problem\$.ti. or reading problem\$.ab.
27. spelling problem\$.ti. or spelling problem\$.ab.
28. literacy dysfunction.ti. or literacy dysfunction.ab.
29. decoding dysfunction.ti. or decoding dysfunction.ab.
30. reading dysfunction.ti. or reading dysfunction.ab.
31. spelling dysfunction.ti. or spelling dysfunction.ab.
32. dyslexi\$.ti. or dyslexi\$.ab.



## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY SKILLS

OR (literacy AND problem)) OR (spelling AND impairment)) OR (reading AND impairment)) OR (decoding AND impairment)) OR (literacy AND impairment)) OR (spelling AND delay)) OR (reading AND delay)) OR (decoding AND delay)) OR (literacy AND delay)) OR (spelling AND disability)) OR (reading AND disability)) OR (decoding AND disability)) OR (literacy AND disability)) OR (spelling AND disorder)) OR (reading AND disorder)) OR (decoding AND disorder)) OR (literacy AND disorder)) OR (spelling AND deficit)) OR (reading AND deficit)) OR (decoding AND deficit)) OR (literacy AND deficit)) OR (spelling difficulties OR spelling difficulty)) OR (reading difficulties OR reading difficulty)) OR (decoding difficulties OR decoding difficulty)) OR literacy difficulties)

### **Web of Science, searched on 26<sup>th</sup> February 2017**

TOPIC: (("literacy difficult\*" OR "decoding difficult\*" OR "reading difficult\*" OR "spelling difficult\*" OR "literacy deficit" OR "decoding deficit" OR "reading deficit" OR "spelling deficit" OR "literacy disorder" OR "decoding disorder" OR "reading disorder" OR "spelling disorder" OR "literacy disability" OR "decoding disability" OR "reading disability" OR "spelling disability" OR "literacy delay" OR "decoding delay" OR "reading delay" OR "spelling delay" OR "literacy impairment" OR "decoding impairment" OR "reading impairment" OR "spelling impairment" OR "literacy problem" OR "decoding problem" OR "reading problem" OR "spelling problem" OR "literacy dysfunction" OR "decoding dysfunction" OR "reading dysfunction" OR "spelling dysfunction" OR dyslexi\* OR dysgraphi\*) AND ("second language" OR "foreign language" OR bilingua\* OR "additional language"))

Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.



# FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

## S4. Data extraction form customized in Covidence

### Foreign language attainment of children and adolescents with poor literacy skills: a systematic review protocol

*Data extraction form in Covidence*

#### Identification

Add Note

#### Study details

Sponsorship Source

Country

Setting

*Poor readers/spellers recruited in schools or in clinical settings*

Comments

#### Author's contact details

Author's name

Institution

Email

Address

#### Additional Identification data

Add another field

*If we suspect that the study is a duplicate, we will add a corresponding note here.*

# FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

## Methods

Add Note

### Design

Other

Group comparison

- Parallel group ← These two default options will not be used.
- Crossover

### Additional Methods data

Add another field

*If we are unsure, if the study design is a group comparison, we will detail this here.*

## Population

Add Note

### Inclusion criteria

*Information on the children's school grade and school type will be detailed here.*

### Exclusion criteria

### Group differences

*The selection criteria for the experimental group will be detailed here. We will collect information on the literacy measure used to determine poor or average literacy skills, the instruments and cut-off points that were used. Furthermore, exclusion criteria such as sensory impairments or general developmental delays will be mentioned here.*

### Additional Population data

Add another field

## Baseline Characteristics

Add Characteristic

# FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

Characteristic	Experimental group	Control group
age: M (SD)		
boys		
girls		
monolinguals		
bilinguals		
plurilinguals		
nonverbal reasoning		
L1 spoken language		
L1 word spelling		
L1 pseudoword spelling		
L1 word reading accuracy		
L1 nonword reading accuracy		
L1 word reading speed		
L1 pseudoword reading speed		
L1 reading comprehension		

## Interventions

*Information on the L1-L1 combination will be detailed as a note.*

Add Note

Add Intervention

Experimental group



Control group



## Description of Interventions

Add a descriptor

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

Characteristic	Experimental group	Control group
Age at FL onset		
Frequency of FL classes per week		
Duration of class in min		
Total amount of FL instruction		

### Outcomes

Add Note

Add Outcome

To add an outcome, you must have at least one intervention.

FL total words read			
Endpoint			
	mean	SD	N
EXPERIMENTAL GROUP			
CONTROL GROUP			

Add a timepoint

FL pseudowords read			
Endpoint			
	mean	SD	N
EXPERIMENTAL GROUP			
CONTROL GROUP			

Add a timepoint

*As many outcome tables as necessary can be added here, according to the outcome measures presented in each study.*

# FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

## S5. Risk of bias assessment tool

**ROBINS-I form**  
Adapted from Sterne et al. (2016)

Domain	Signalling question	Response option
Risk of bias due to confounding	1.1 Is there potential for confounding of the effect of intervention in this study?	Y/ PY/ <u>PN</u> / <u>N</u>
	If N/PN to 1.1: the study can be considered to be at low risk of bias due to confounding and no further signalling questions need to be considered.	
	If Y/PY to 1.1: answer the following signalling questions	
	1.2 Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	<u>Y</u> / <u>PY</u> / PN/ N/ NI
	1.3 Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	<u>Y</u> / <u>PY</u> / PN/ N/ NI
	<b>Risk of bias judgement</b>	
	no confounding expected	Low
	(i) confounding expected, all known important confounding domains appropriately measured and controlled for <u>and</u> (ii) reliability and validity of measurement of important domains were sufficient, such that we do not expect serious residual confounding.	Moderate
	(i) at least one known important domain was not appropriately measured, or not controlled for <u>or</u> (ii) reliability or validity of measurement of an important domain was low enough that we expect serious residual confounding.	Serious
	(i) confounding inherently not controllable or (ii) the use of negative controls strongly suggest unmeasured confounding.	Critical
No information on whether confounding might be present.	No information	

Note. Y= Yes; PY = Probably Yes; PN = Probably No; N = No; NI = No Information

Domain	Signalling question	Response option
Risk of bias in selection of participants into the study	2.1 Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention?	Y/ PY/ <u>PN</u> / <u>N</u> / NI
	2.2 Do start of follow-up and start of intervention coincide for most participants?	<u>Y</u> / <u>PY</u> / PN/ N/ NI
	<b>Risk of bias judgement</b>	
	(i) All participants who would have been eligible for the target trial were included in the study <u>and</u> (ii) for each participant, start of follow up and start of intervention coincided.	Low
	(i) Selection into the study may have been related to intervention and outcome <u>and</u> the authors used appropriate methods to adjust for the selection bias; <u>or</u> (ii) start of follow up and start of intervention do not coincide for all participants; <u>and</u> (a) the proportion of participants for which this was the case was too low to induce important bias; <u>or</u> (b) the authors used appropriate methods to adjust for the selection bias; <u>or</u> (c) the review authors are confident that the rate (hazard) ratio for the effect of intervention remains constant over time.	Moderate
	(i) Selection into the study was related (but not very strongly) to intervention and outcome; <u>and</u> this could not be adjusted for in analyses; <u>or</u> (ii) start of follow up and start of intervention do not coincide <u>and</u> a potentially important amount of follow-up time is missing from analyses; <u>and</u> the rate ratio is not constant over time.	Serious
	(i) Selection into the study was very strongly related to intervention and outcome; <u>and</u> this could not be adjusted for in analyses; <u>or</u> (ii) a substantial amount of follow-up time is likely to be missing from analyses; <u>and</u> the rate ratio is not constant over time.	Critical
	No information is reported about selection of participants into the study or whether start of follow up and start of intervention coincide.	No information

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

Domain	Signalling question	Response option	
Risk of bias in classification of interventions	3.1 Were intervention groups clearly defined?	<u>Y</u> / <u>PY</u> / PN/ N/ NI	
	3.2 Was the information used to define intervention groups recorded at the start of the intervention?	<u>Y</u> / <u>PY</u> / PN/ N/ NI	
	3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	Y/ <u>PY</u> / <u>PN</u> / <u>N</u> / NI	
	<b>Risk of bias judgement</b>		
	(i) Intervention status is well defined; <u>and</u> (ii) intervention definition is based solely on information collected at the time of intervention.	Low	
	(i) Intervention status is well defined; <u>and</u> (ii) some aspects of the assignments of interventions status were determined retrospectively.	Moderate	
	(i) Intervention status is not well defined; <u>or</u> (ii) major aspects of the assignments of intervention status were determined in a way that could have been affected by knowledge of the outcome.	Serious	
	(Unusual) An extremely high amount of misclassification of intervention status, e.g. because of unusually strong recall biases.	Critical	
	No definition of intervention or no explanation of the source of information about intervention status is reported.	No information	

*Note.* To avoid confusion with the term 'foreign language instruction', we changed the original term 'risk of bias in classification of interventions' used by the ROBINS I scale (Sterne et al., 2016), we used the term 'risk of bias in classification of participant group' throughout the manuscript.

Domain	Signalling question	Response option	
Risk of bias due to deviations from intended interventions	4.1 Were there deviations from the intended intervention beyond what would be expected in usual practice?	Y/ <u>PY</u> / <u>PN</u> / <u>N</u> / NI	
	4.2 Were these deviations from intended intervention unbalanced between groups <u>and</u> likely to have affected the outcome?	Y/ <u>PY</u> / <u>PN</u> / <u>N</u> / NI	
	4.3 Were important co-interventions balanced across intervention groups?	<u>Y</u> / <u>PY</u> / PN/ N/ NI	
	<b>Risk of bias judgement</b>		
	(i) Any deviations from intended intervention reflected usual practice; <u>or</u> (ii) any deviations from usual practice were unlikely to impact on the outcome. The important co-interventions were balanced across intervention groups, <u>and</u> there were no deviations from the intended interventions (in terms of implementation or adherence) that were likely to impact on the outcome.	Low	
	(i) There were deviations from usual practice, but their impact on the outcome is expected to be slight; <u>or</u> (ii) the important co-interventions were not balanced across intervention groups, or there were deviations from the intended interventions (in terms of implementation and/or adherence) that were likely to impact on the outcome; <u>and</u> the analysis was appropriate to estimate the effect of starting and adhering to intervention, allowing for deviations (in terms of implementation, adherence and co-intervention) that were likely to impact the outcome.	Moderate	
	(i) There were deviations from usual practice that were unbalanced between the intervention groups and likely to have affected the outcome. (ii) The important co-interventions were not balanced across intervention groups, or there were deviations from the intended interventions (in terms of implementation and/or adherence) that were likely to impact on the outcome; <u>and</u> (iii) the analysis was not appropriate to estimate the effect of starting and adhering to intervention, allowing for deviations (in terms of implementation, adherence and co-interventions) that were likely to impact on the outcome.	Serious	

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

Domain	Signalling question	Response option
Risk of bias due to deviations from intended interventions	(i) There were substantial deviations from usual practice that were unbalanced between the intervention groups and likely to have affected the outcome. (ii) There were substantial imbalances in important co-interventions across intervention groups, or there were substantial deviations from the intended interventions (in terms of implementation and/or adherence) that were likely to impact on the outcome; <i>and</i> (iii) the analysis was not appropriate to estimate the effect of starting and adhering to intervention, allowing for deviations (in terms of implementation, adherence and co-interventions) that were likely to impact on the outcome.	Critical
	No information is reported on whether there is deviation from the intended intervention.	No information

Domain	Signalling question	Response option	
Risk of bias due to missing data  <i>Applicable to each outcome</i>	5.1 Were outcome data available for all, or nearly all participants (>90%)?	<u>Y</u> / <u>PY</u> / <u>PN</u> / <u>N</u> / <u>NI</u>	
	5.2 Were participants excluded due to missing data on intervention status?	<u>Y</u> / <u>PY</u> / <u>PN</u> / <u>N</u> / <u>NI</u>	
	5.3 Were participants excluded due to missing data on other variables needed for the analysis?	<u>Y</u> / <u>PY</u> / <u>PN</u> / <u>N</u> / <u>NI</u>	
	5.4 Are the proportions of participants and reasons for missing data similar across interventions?	<u>Y</u> / <u>PY</u> / <u>PN</u> / <u>N</u> / <u>NI</u>	
	5.5 Is there evidence that results were robust to the presence of missing data?	<u>Y</u> / <u>PY</u> / <u>PN</u> / <u>N</u> / <u>NI</u>	
	<b>Risk of bias judgement</b>		
	(i) Data were reasonably complete; <i>or</i> (ii) proportions of and reasons for missing participants were similar across intervention groups; <i>or</i> (iii) the analysis addressed missing data and is likely to have removed any risk of bias.	Low	
	(i) Proportions of and reasons for missing participants differ slightly across intervention groups; <i>and</i> (ii) the analysis is unlikely to have removed the risk of bias arising from the missing data.	Moderate	
	(i) Proportions of missing participants differ substantially across interventions; or reasons for missingness differ substantially across interventions; <i>and</i> (ii) the analysis is unlikely to have removed the risk of bias arising from the missing data; <i>or</i> missing data were addressed inappropriately in the analysis; or the nature of the missing data means that the risk of bias cannot be removed through appropriate analysis.	Serious	
	(i) (Unusual) There were critical differences between interventions in participants with missing data; <i>and</i> (ii) missing data were not, or could not, be addressed through appropriate analysis.	Critical	
No information is reported about missing data or the potential for data to be missing.	No information		

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

Domain	Signalling question	Response option	
Risk of bias in measurement of outcomes  <i>Applicable to each outcome</i>	6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Y/ PY/ <u>PN/N/</u> NI	
	6.2 Were outcome assessors aware of the intervention received by study participants?	Y/ PY/ <u>PN/N/</u> NI	
	6.3 Were the methods of outcome assessment comparable across intervention groups?	<u>Y/PY/</u> PN/N/ NI	
	6.4 Were any systematic errors in measurement of the outcome related to intervention received?	Y/ PY/ <u>PN/N/</u> NI	
	<b>Risk of bias judgement</b>		
	(i) The methods of outcome assessment were comparable across intervention groups; <u>and</u> (ii) the outcome measure was unlikely to be influenced by knowledge of the intervention received by study participants (i.e. is objective) or the outcome assessors were unaware of the intervention received by study participants; <u>and</u> (iii) any error in measuring the outcome is unrelated to intervention status.	Low	
	(i) The methods of outcome assessment were comparable across intervention groups; <u>and</u> (ii) the outcome measure is only minimally influenced by knowledge of the intervention received by study participants; <u>and</u> (iii) any error in measuring the outcome is only minimally related to intervention status.	Moderate	
	(i) The methods of outcome assessment were not comparable across intervention groups; <u>or</u> (ii) the outcome measure was subjective (i.e. vulnerable to influence by knowledge of the intervention received by study participants; <u>and</u> the outcome was assessed by assessors aware of the intervention received by study participants; <u>or</u> (iii) error in measuring the outcome was related to intervention status.	Serious	
	The methods of outcome assessment were so different that they cannot reasonably be compared across intervention groups.	Critical	
	No information is reported about the methods of outcome assessment.	No information	

Domain	Signalling question	Response option	
Risk of bias in selection of the reported result  <i>Applicable to each outcome</i>	7.1 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple outcome <i>measurements</i> within the outcome domain?	Y/ PY/ <u>PN/N/</u> NI	
	7.2 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple <i>analyses</i> of the intervention-outcome relationship?	Y/ PY/ <u>PN/N/</u> NI	
	7.3 Is the reported effect estimate likely to be selected, on the basis of the results, from different <i>subgroups</i> ?	Y/ PY/ <u>PN/N/</u> NI	
	<b>Risk of bias judgement</b>		
	There is clear evidence (usually through examination of a pre-registered protocol or statistical analysis plan) that all reported results correspond to all intended outcomes, analyses and sub cohorts.	Low	
	(i) The outcome measurements and analyses are consistent with an <i>a priori</i> plan; or are clearly defined and both internally and externally consistent; <u>and</u> (ii) there is no indication of selection of the reported analysis among multiple analyses; <u>and</u> (iii) there is no indication of selection of the cohort or subgroups for analysis and reporting on the basis of the results.	Moderate	
	(i) Outcomes are defined in different ways in the methods and results section, or in different publications of the study; <u>or</u> (ii) there is a high risk of selective reporting among multiple analyses; <u>or</u> (iii) the cohort or subgroup is selected from a larger study for analysis and appears to be reported on the basis of the results.	Serious	
	(i) There is evidence or strong suspicion of selective reporting of results; <u>and</u> (ii) the unreported results are likely to be substantially different from the reported results.	Critical	
	There is too little information to make a judgement (for example if only an abstract is available for the study).	No information	









## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

	Bekebrede et al. (2009)	Bonifacci et al. (2017)	Chung et al. (2010)	de Bree & Unsworth (2014)	Ding et al. (2013)	Haisma (2009)	Helland & Kaasa (2005)	Helland & Morken (2016)	Ho & Fong (2005)	Lockiewicz & Jaskulskaa (2016)	Morfidi et al. (2007)	Palladino et al. (2013)	Palladino et al. (2016)	van der Leij & Morfidi (2006)	van Viersen et al. (2017)	Zhou et al. (2014)	Total of studies with this information
Foreign language instruction																	
Onset age of foreign language instruction		?		?		?						?	?				
Early childhood: onset age before 6 years			X						X							X	3
Late childhood: onset age from 6-11 years					X		X	X		X	X						5
Adolescence: onset age from 12-17 years	X													X	X		3
Transition from adolescence to adulthood: onset age from 18 years onwards																	0
Foreign language assessment																	
Age at foreign language assessment																	
Early childhood: before 6 years of age																	0
Late childhood: from 6-11 years of age		X	X		X			X	X		X					X	7
Adolescence: from 12-17 years of age	X			X		X	X			X		X	X	X	X		9
Transition from adolescence to adulthood: onset age from 18 years onwards																	0

*Note.* X = allocated to this subgroup; ? = no information; n/a = not applicable; nNL = Native language; r = recognizing letters, ordering letters and moving letters between words;

cFL = Foreign language.

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

### S7. Foreign language outcome measures of included studies.

	Bekebrede et al. (2009)	Bonifacci et al. (2017)	Chung et al. (2010)	de Bree et al. (2014)	Ding et al. (2013)	Haisma (2009)	Helland et al. (2005)	Helland et al. (2016)	Ho et al. (2005)	Lockiewicz et al. (2016)	Morfidi et al. (2007)	Palladino et al. (2013)	Palladino et al. (2016)	van der Leij et al. (2006)	van Viersen et al. (2017)	Zhou et al. (2014)	Total of studies
Discrimination of speech sounds																	0
Production of speech sounds																	0
Receptive vocabulary knowledge	X			X					X		X			X		X	6
Spoken word production			X		X				X		X			X			5
Sentence comprehension							X										1
Sentence production							X										1
Short term memory									X								1
Phonological awareness			X						X		X			X			4
Letter knowledge			X								X						2
Word reading	X	X	X	X	X		X	X	X	X	X	X		X	X	X	14
Nonword reading	X	X		X						X	X	X		X			7
Orthographic knowledge	X		X	X		X					X			X	X		7
Reading comprehension		X			X						X			X			4
Spelling		X				X	X	X	X	X			X		X		8
Translation							X	X									2
Total of measures	2	3	3	2	3	2	5	3	6	2	5	1	1	5	2	2	

Note. X = includes data on this foreign language outcome measure

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

### S8. Reasons for excluding studies at full text screening phase

Authors	Exclusion reason
Borodkin et al. (2014)	Participants in post secondary education <sup>a</sup>
Letnes (2014)	Participants in post secondary education <sup>a</sup>
Ferrari et al. (2012)	Poor literacy skills not 1 SD, year or grade below expected level <sup>b</sup>
Jared et al. (2011)	Poor literacy skills not 1 SD, year or grade below expected level <sup>b</sup>
Kahn-Horwitz et al. (2006)	Poor literacy skills not 1 SD, year or grade below expected level <sup>b</sup>
Keung et al. (2009)	Poor literacy skills not 1 SD, year or grade below expected level <sup>b</sup>
Sparks et al. (1998)	Poor literacy skills not 1 SD, year or grade below expected level <sup>b</sup>
Farukh et al. (2016)	Literacy performance measure by self- or teacher reports <sup>c</sup>
Ghonsooly et al. (2010)	Literacy performance measure by self- or teacher reports <sup>c</sup>
Meng et al. (2016)	Literacy performance measure by self- or teacher reports <sup>c</sup>
Russak et al. (2015)	Literacy performance measure by self- or teacher reports <sup>c</sup>
vanDaal et al. (1999)	Literacy performance measure by self- or teacher reports <sup>c</sup>
Caglar-Ryeng et al. (2010)	No comparison with controls on foreign language measure <sup>d</sup>
Crombie et al. (1997)	No comparison with controls on foreign language measure <sup>d</sup>
Björn et al. (2013)	Allocation to participant group based on foreign language performance <sup>e</sup>
Kalindi et al. (2015)	Allocation to participant group based on foreign language performance <sup>e</sup>
McBride-Chang et al. (2012)	Allocation to participant group based on foreign language performance <sup>e</sup>
Pfenninger et al. (2015)	Allocation to participant group based on foreign language performance <sup>e</sup>
Alanis et al. (2005)	Access to foreign language outside of instruction context <sup>f</sup>
Geva et al. (1994)	Access to foreign language outside of instruction context <sup>f</sup>
Gottardo (2002)	Access to foreign language outside of instruction context <sup>f</sup>
Gottardo et al. (2008)	Access to foreign language outside of instruction context <sup>f</sup>
Gupta et al. (2007)	Access to foreign language outside of instruction context <sup>f</sup>
Guzman-Orth et al. (2013)	Access to foreign language outside of instruction context <sup>f</sup>
Haigh et al. (2011)	Access to foreign language outside of instruction context <sup>f</sup>
Hedman (2012)	Access to foreign language outside of instruction context <sup>f</sup>
Hutchinson et al. (2004)	Access to foreign language outside of instruction context <sup>f</sup>
Kieffer (2014)	Access to foreign language outside of instruction context <sup>f</sup>
Kline et al. (1972)	Access to foreign language outside of instruction context <sup>f</sup>
Lallier et al. (2014)	Access to foreign language outside of instruction context <sup>f</sup>
Lesaux et al. (2007)	Access to foreign language outside of instruction context <sup>f</sup>
Limbos (2006)	Access to foreign language outside of instruction context <sup>f</sup>

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

---

Lipka et al. (2007)	Access to foreign language outside of instruction context <sub>f</sub>
Willis (2002)	Access to foreign language outside of instruction context <sub>f</sub>
Kormos et al. (2010)	No assessment of oral or written foreign language performance <sub>g</sub>
Sauve (2009)	No assessment of oral or written foreign language performance <sub>g</sub>
Alanis et al. (2005)	Duplicated study report <sub>h</sub>
Bonifacci et al. (2017)	Duplicated study report <sub>h</sub>
Borodkin et al. (2014)	Duplicated study report <sub>h</sub>
Chung et al. (2010)	Duplicated study report <sub>h</sub>
Farukh et al. (2016)	Duplicated study report <sub>h</sub>
Farukh et al. (2016)	Duplicated study report <sub>h</sub>
Gupta et al. (2008)	Duplicated study report <sub>h</sub>
Helland (2008)	Duplicated study report <sub>h</sub>
Helland et al. (2016)	Duplicated study report <sub>h</sub>
Helland et al. (2016)	Duplicated study report <sub>h</sub>
Lipka et al. (2012)	Duplicated study report <sub>h</sub>
McBride-Chang et al. (2012)	Duplicated study report <sub>h</sub>
Pfenninger et al. (2016)	Duplicated study report <sub>h</sub>
Tong et al. (2017)	Duplicated study report <sub>h</sub>
van derLeij et al. (2006)	Duplicated study report <sub>h</sub>
Wiss (1993)	No acces to full text
Amner (1933)	No empirical study report
Dulude (2012)	No empirical study report
Finelli-Thomsen et al. (2012)	No empirical study report
Gonzales et al. (1981)	No empirical study report
Kovelman et al. (2016)	No empirical study report
Lodej (2016)	No empirical study report
Nijakowska (2009)	No empirical study report
Nijakowska (2010)	No empirical study report
Schneider (2009)	No empirical study report
AquinoAndersen et al. (2016)	Qualitative analysis
Ghazaleh et al. (2011)	Qualitative analysis
Lockhart-Pedersen (2013)	Qualitative analysis
Marogna (2013)	Qualitative analysis
Ni'mah (2016)	Qualitative analysis

---

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

---

Szaszkiewicz (2013)	Qualitative analysis
Gamper (2013)	Single case study
Valdois et al. (2014)	Single case study

---

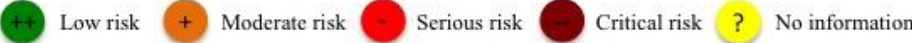
*Note.* Studies are ordered according to selection signaling questions and other criteria. <sup>a</sup>Selection criteria signaling question 1 (see Appendix 1); <sup>b-g</sup>Selection criteria signaling questions 3-9 (see Appendix 1); <sup>h</sup>Exactly the same report or duplicated report from the same sample.



# FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

## S9. Risk of bias assessment of included studies

	Confounding	Selection of participants	Classification of interventions	Deviatio from intended interventions	Missing data	Measurement of outcomes	Selection of reported results	Overall
Bekebrede et al. (2009)	+	+	++	++	?	++	+	+
Bonifacci et al. (2017)	+	+	++	++	?	++	+	+
Chung & Ho (2010)	+	+	++	++	?	++	+	+
de Bree & Unsworth (2014)	+	-	++	++	?	++	+	-
Ding et al. (2013)	+	+	+	++	?	++	+	+
Haisma (2009)	+	+	++	++	?	++	+	+
Helland & Kaasa (2005)	+	+	++	++	?	++	+	+
Helland & Morken ( 2016)	+	++	+	++	++	++	+	+
Ho & Fong (2005)	+	+	++	++	?	++	+	+
Lockiewicz & Jaskulskaa (2016)	+	+	++	++	?	++	+	+
Morfidi et al. (2007)	+	+	++	++	?	++	+	+
Palladino et al. (2013)	+	+	++	++	?	++	+	+
Palladino et al. (2016)	+	+	++	++	?	+	+	+
van der Leij & Morfidi (2006)	+	+	++	++	?	++	+	+
van Viersen et al. (2017)	+	+	++	++	++	++	+	+
Zhou et al. (2014)	+	+	++	++	?	++	+	+



## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

### S10. Data transformations

After completing the data extraction and risk of bias assessment for the 16 studies included in this report, we downloaded all the information from Covidence and completed the following steps:

1. In six studies, children/adolescents with poor literacy skills were allocated to subgroups on the basis of relevant native or foreign language measures. In these cases, we merged information into one data-point per study, in the following way:

Study report	Data transformations
Bekebrede et al. (2009)	Data from the subgroup of children with dyslexia with good (ORTH+) and poor (ORTH-) orthographic knowledge were merged into one group.
Haisma (2009)	We merged data from a subgroup of participants with phonological and surface dyslexia.
Helland & Kaasa (2005)	Data was available for two subgroups of children with dyslexia with poor and good foreign language comprehension skills as well as for the overall experimental group. Therefore, we simply excluded the subgroup data and relied on the complete group data.
van der Leij & Morfidi (2006)	We merged data from a subgroup of poor readers with good and poor foreign language orthographic knowledge.
van Viersen et al. (2017)	We merged data from a subgroup of dyslexic and gifted dyslexic participants.

2. A similar situation emerged with respect to the inclusion of more than one control group in seven studies. In these cases, we proceeded as follows:

Study report	Data transformations
Bonifacci et al. (2017)	We excluded data from an additional control group of children with typical literacy skills with minority bilingual background.
Chung & Ho (2010)	In addition to age-matched control participants, this study included a group of reading-matched control participants. In order to focus on the comparison with age-matched controls, we excluded reading-matched information (for a similar approach see McArthur et al., 2013; McDougall, Borowsky, MacKinnon & Hymel, 2005).
Ding et al. (2013)	This study included two control groups of good and average readers. As this distinction was not made in other studies, we merged the data of both control groups.
Palladino et al. (2016)	We excluded data from an additional control group of children with English language difficulties.
van Viersen et al. (2017)	This study included two control groups with typically developing and gifted participants. As this distinction was not made in other studies, we merged the data of both control groups.
Zhou et al. (2014)	Just as for Chung & Ho (2010), we excluded data from a reading-matched control group and only included data from the age-matched control participants in further analyses.

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

3. With respect to foreign language outcome measures, we used the  $M$ ,  $SD$  and  $n$  reported for each participant group to calculate  $SMDs$  (with Hedges' correction) and variances following the computations suggested by Bornsetin et al. (2009) on p.26-28, as well as  $CVRs$  and corresponding variances following Nakagawa et al. (2015). Furthermore, 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 by checking the original data reported in the study. If measures were based on the occurrence of errors (instead of accuracy rates), the sign of the  $M$  was reversed (Borenstein et al., 2009). This occurred in the following cases:

Study report	Data transformations
Chung & Ho (2010)	Rapid letter naming task measured in reaction time.
Ding et al. (2013)	Rapid letter naming task measured in reaction time. Passage reading errors.
Ho & Fong (2005)	Rapid picture naming task measured in reaction time.
Lockiewicz et al. (2016)	Pseudoword reading speed measured in reaction time. Phonological, orthographic, other correct word and blank gap errors in spelling task.
Morfidi et al. (2007)	Rapid automatized naming task measured in reaction time.
Palladino et al. (2016)	Spelling errors
van der Leij & Morfidi (2006)	Rapid automatized naming task measured in reaction time.

4. In addition, for each study data from similar tasks were grouped together under broader categories according to the 15 foreign language outcome measures we focused on in this review (see Table 1 in main text). For this purpose, we aggregated  $SMDs$ ,  $CVRs$  and variances following the computations suggested on p.230 of Bornstein et al. (2009) assuming a correlation coefficient of  $r = .5$  between tasks. Details of the specific tasks that we merged are provided below:

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

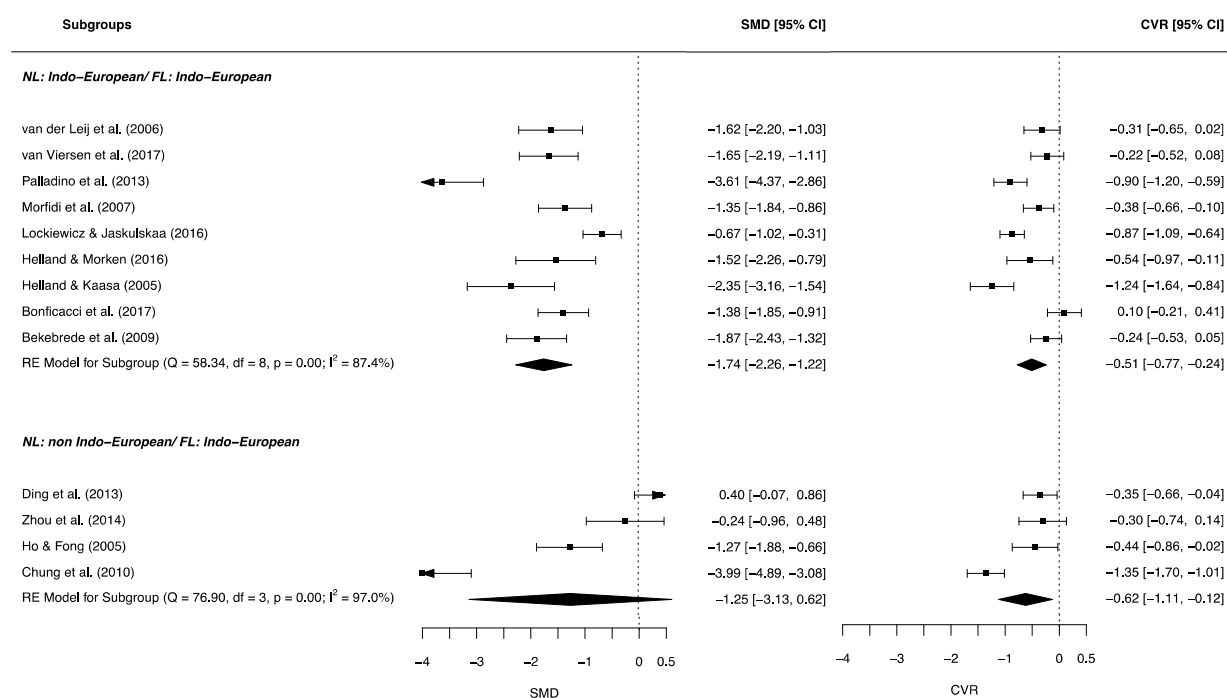
Study report	Data transformations
Bekebrede et al. (2009)	Word reading: flashed word production and word reading fluency
Bonifacci et al. (2017)	Word reading: accuracy and speed data Nonword reading: accuracy and speed data
Chung & Ho (2010)	Reading comprehension: sentence and passage reading comprehension
Ding et al. (2013)	Phonological awareness: rhyme detection and phoneme deletion
Haisma (2009)	Word reading: passage reading fluency and errors
Helland & Kaasa (2005)	Spelling: total, regular and irregular words score
	Sentence comprehension: total and subscores for narrative, interrogative, negative, passive sentences and inversion
	Sentence production: model sentences total and subscores for morphology, syntax and semantics, narrative, interrogative, negative sentences and inversion plus daily conversation and picture story task
Ho & Fong (2005)	Phonological awareness: onset detection and phoneme deletion
Lockiewicz et al. (2016)	Word reading: accuracy and speed data Nonword reading: accuracy and speed data
	Spelling: total errors and phonological, orthographic, other correct word and blank gap errors
Morfidi et al. (2007)	Spoken word production: Serial rapid naming and semantic fluency tasks
	Word reading: speeded and unspeeded word reading task, as well as text reading accuracy
Palladino et al. (2013)	Word reading: total, absolute and relative accuracy score, as well as speed
	Nonword reading: total, absolute and relative accuracy score, as well as speed
Palladino et al. (2016)	Spelling: correct words, Z-score, Italian and English rules two and three syllable words phonological and non-phonological errors
van der Leij & Morfidi (2006)	Spoken word production: serial rapid naming and semantic fluency
	Word reading: unspeeded and speeded word reading, as well as text reading accuracy

5. The same procedure was followed for longitudinal studies reporting more than one data-point per outcome measure (Borenstein et al., 2009), as detailed below:

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

Study report	Data transformations
van der Leij & Morfidi. (2006)	This study reported more than one data-point on the same orthographic choice task. Therefore, we merged this information into one data-point.
van Viersen et al. (2017)	This study reported more than one data-point on the same task. Therefore, we merged the data-points collected for the same English as a foreign language task at different developmental stages. However, we excluded the information on the participants' performance on German and French as a second foreign language, because all other studies only focused on foreign language outcome measures that represented the participants' first foreign language.
Zhou et al. (2014)	This study reported more than one data-point on the same task. We excluded information from the first and second test point of this longitudinal study, because participants' literacy status was only determined at the third test point. Due to this moderate risk of bias in the classification to intervention status (see risk of bias assessment section for details), we only included the third data-point of this study.

## S11. Moderator analysis on language pairing for foreign language word reading

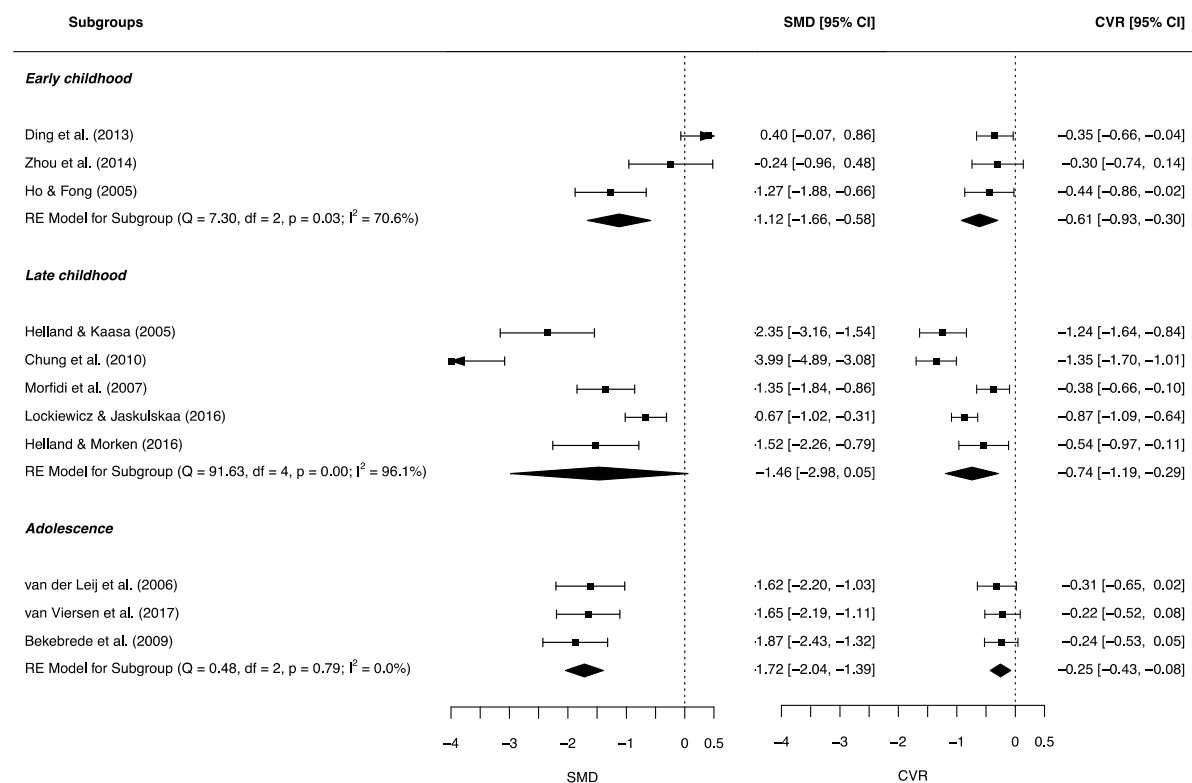


Note. *SMD* = Standardized Mean Difference; *CVR* = Natural logarithm of the ratio of coefficients of variation (Nakagawa et al., 2015).

We compared the results of separate random effects modeling meta-analyses for studies with two types of language pairings: (a) a native Indo-European language with an alphabetic writing system combined with a foreign language with the same characteristics and (b) a native non Indo-European language with an ideographic writing system combined with a Indo-European foreign language with an alphabetic writing system. No significant impact of this moderator was detected neither for the *SMD*, nor for the *CVR* of the foreign language word reading performance between children/adolescents with poor and typical literacy skills (*SMD*:  $Z_{Diff}^* = 0.75$ ,  $p = .45_{ns}$ ; *CVR*:  $Z_{Diff}^* = 0.42$ ,  $p = .67_{ns}$ ).

# FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

## S12. Moderator analysis on onset age of foreign language instruction for foreign language word reading

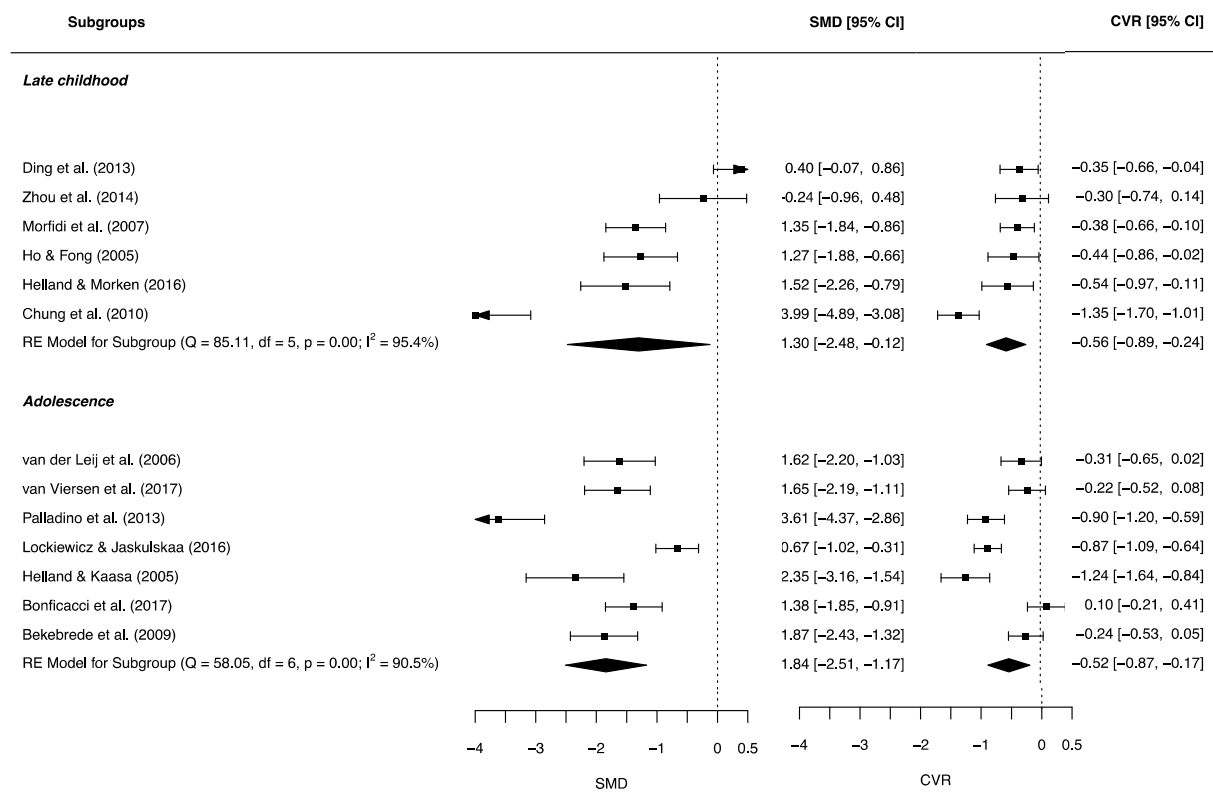


Note. SMD = Standardized Mean Difference; CVR = Natural logarithm of the ratio of coefficients of variation (Nakagawa et al., 2015).

Subgroups of studies assessing participants with an onset age of foreign language instruction in early childhood (before age 6), late childhood (age 6 to 11) and adolescence (age 12 to 17) were compared. Results revealed no significant difference between subgroups for SMD, or for CVR (early childhood vs. late childhood: SMD:  $Z_{Diff}^* = 0.43$ ,  $p = .66_{ns}$ ; CVR:  $Z_{Diff}^* = 0.45$ ,  $p = .65_{ns}$ ; early childhood vs. adolescence: SMD:  $Z_{Diff}^* = 2.02$ ,  $p = .04_{ns}$ ; CVR:  $Z_{Diff}^* = 2.23$ ,  $p = .02_{ns}$ ; late childhood vs. adolescence: SMD:  $Z_{Diff}^* = 0.35$ ,  $p = .72_{ns}$ ; CVR:  $Z_{Diff}^* = 2.14$ ,  $p = .03_{ns}$ ).

# FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

## S13. Moderator analysis on age at foreign language assessment for foreign language word reading



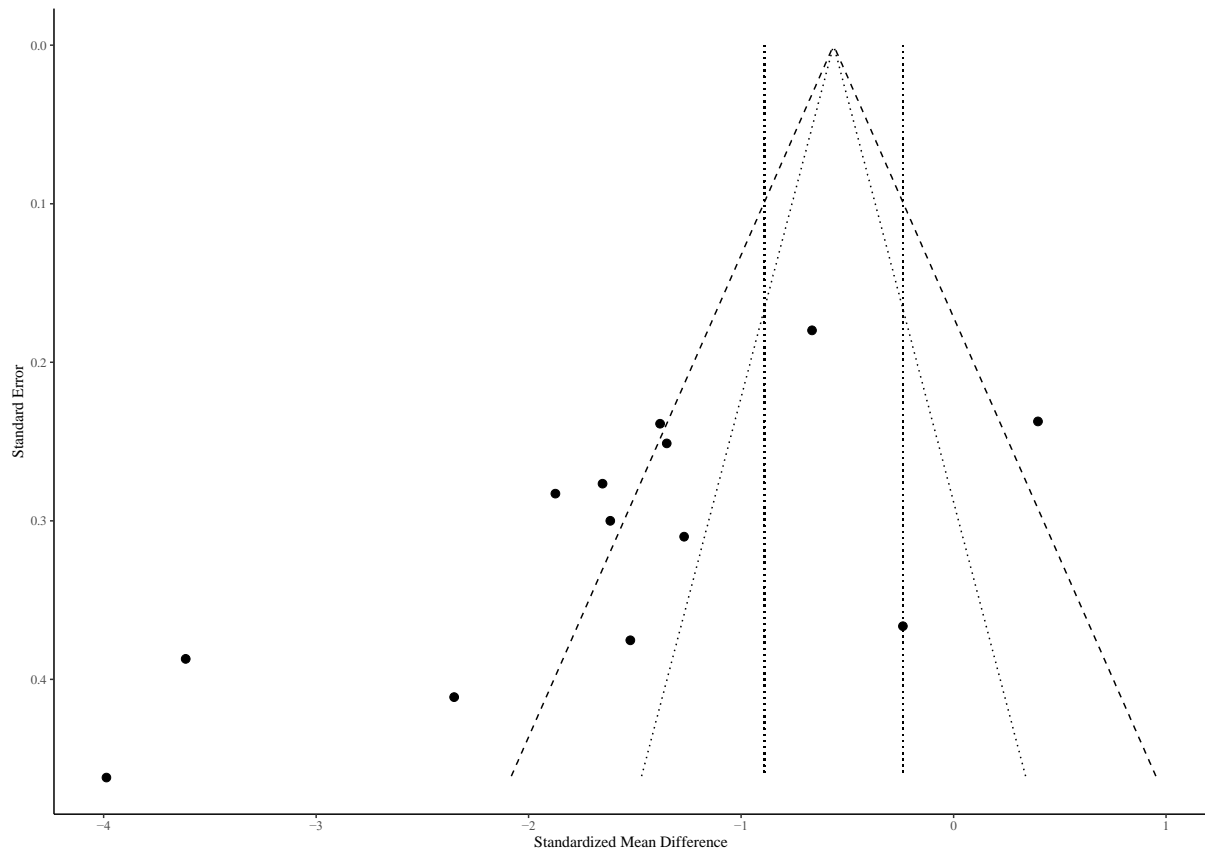
Note. *SMD* = Standardized Mean Difference; *CVR* = Natural logarithm of the ratio of coefficients of variation (Nakagawa et al., 2015).

We compared the overall effects of a subgroup of studies that assessed participants in late childhood (age 6 to 11) and adolescence (age 12 to 17). No significant differences between subgroups were found neither for the *SMD* or the *CVR* (*SMD*:  $Z_{Diff^*} = 0.87, p = .38_{ns}$ ; *CVR*:  $Z_{Diff^*} = 0.18, p = .85_{ns}$ ).



# FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

## S14. Funnel plot of studies measuring foreign language word reading ( $n=13$ )



*Note.* Each dot represents one study measuring foreign language word reading.

## FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY

### S15. Results of sensitivity analysis.

	Mean difference								Variance difference							
	without de Bree & Unsworth (2014)				with de Bree & Unsworth (2014)				without de Bree & Unsworth (2014)				with de Bree & Unsworth (2014)			
	<i>SMD</i>	<i>Z</i>	<i>Q</i>	<i>I</i> <sub>2</sub>	<i>SMD</i>	<i>Z</i>	<i>Q</i>	<i>I</i> <sub>2</sub>	<i>CVR</i>	<i>Z</i>	<i>Q</i>	<i>I</i> <sub>2</sub>	<i>CVR</i>	<i>Z</i>	<i>Q</i>	<i>I</i> <sub>2</sub>
Receptive vocabulary	-0.47	-2.59 <sub>ns</sub>	7.40 <sub>ns</sub>	45.98%	-0.42	-2.84*	7.83 <sub>ns</sub>	36.12%	-0.28	-1.70 <sub>ns</sub>	16.43*	75.66%	-0.23	-1.58 <sub>ns</sub>	18.53*	73.02%
Word reading	-1.59	-5.48*	155.30*	92.27%	-1.53	-5.78*	157.46*	91.74%	-0.54	-4.63*	76.93*	84.40%	-0.56	-5.14*	79.19*	83.58%
Nonword reading	-0.90	-2.93*	43.91*	88.61%	-0.84	-3.21*	44.39*	86.48%	-0.18	-2.00 <sub>ns</sub>	9.85*	49.22%	-0.18	-2.35 <sub>ns</sub>	9.83 <sub>ns</sub>	38.95%
Orthographic knowledge	-1.42	-15.66*	31.17*	83.96%	-1.37	-5.52*	31.29*	80.82%	-0.70	-2.49 <sub>ns</sub>	81.67*	93.87%	-0.66	-2.75 <sub>ns</sub>	82.52*	92.73%

*Note.* *SMD* = Standard Mean Difference; *CVR* = natural logarithm of the ratio of coefficients of variance (Nakagawa et al., 2015). The significance level was  $p < .05$  for the *Q* statistic and  $p < .005$  for the *Z* statistic (Bonferroni correction for 10 comparisons). Cells marked in grey show overall effects that could not be interpreted due to the presence of significant between study heterogeneity

# FOREIGN LANGUAGE ATTAINMENT AND POOR LITERACY