Bilingualism: Language and Cognition

cambridge.org/bil

Applying meta-analysis to research on bilingualism: An introduction

Luke Plonsky, Ekaterina Sudina and Yuhang Hu

Department of English, Northern Arizona University, Flagstaff, Arizona, USA

Review Article

Cite this article: Plonsky L, Sudina E, Hu Y (2021). Applying meta-analysis to research on bilingualism: An introduction. *Bilingualism: Language and Cognition* **24**, 819–824. https://doi.org/10.1017/S1366728920000760

Received: 1 February 2020
Revised: 31 October 2020
Accepted: 5 November 2020
First published online: 27 January 2021

Keywords:
meta-analysis; research synthesis; effect sizes; bilingualism; quantitative research methods

Address for correspondence:
Luke Plonsky, Email: lukeplonsky@gmail.com

Abstract

Meta-analysis overcomes a number of the limitations of traditional literature reviews (Norris & Ortega, 2006). Consequently, the use of meta-analysis as a synthetic technique has been applied across a range of scientific disciplines in recent decades. This paper seeks to formally introduce the potential of meta-analysis to the field of bilingualism. In doing so, we first describe a number of advantages to the meta-analytic approach such as greater systematicity, objectivity, and transparency relative to narrative reviews. We also outline the major stages in conducting a meta-analysis, highlighting critical considerations encountered at each stage. These include (a) domain definition, (b) coding scheme development and implementation, (c) analysis, and (d) interpretation. The focus, however, is on providing a conceptual introduction rather than a full-length tutorial. Meta-analyses in bilingualism and nearby fields are referred to throughout in order to illustrate the points being made.

Conceptual motivation

Researchers in bilingualism, as in other social and behavioral sciences, have traditionally brought together findings in individual domains in the form of (narrative) literature reviews. Unfortunately, such an approach introduces a great deal of opacity as well as a number of potential flaws, biases, and limitations at all stages – from the collection of studies to the interpretation of their outcomes to the synthesis of findings across studies on relationships of interest. Such issues are doubtlessly at play in an often-contentious and always-complex field such as bilingualism. Meta-analysis endeavors to overcome many of these limitations by embracing a scientific approach to the process of reviewing existing literature (i.e., one that strives for systematicity, objectivity, and transparency). This paper seeks to provide a formal and conceptual introduction to meta-analysis – a procedure for aggregating findings across multiple studies that address a common question – for the field of bilingualism. For a tutorial on the more practical side of conducting a meta-analysis, see Plonsky & Oswald, 2015. We begin by highlighting some of the core attributes and advantages inherent to the meta-analytic approach.

One major benefit afforded by the systematicity and objectivity of meta-analysis is seen in the sample of studies that is synthesized. The meta-analyst endeavors to carry out an exhaustive search for relevant research such that the final sample approximates if not equals the population of studies within the domain of interest. More thorough sampling also allows for greater statistical power, greater generalizability of findings, and a more comprehensive view of accumulated findings within the domain. Traditional reviews, by contrast, are much more idiosyncratic thus allowing for gaps and biases in the corpus of evidence.

The benefits of scientific rigor are also evident in the data collection process. Meta-analysts treat each study as a 'participant' who is surveyed using a coding scheme designed to extract all relevant substantive and methodological features as well as study outcomes. Such coding allows for the systematic analysis of numerous and potentially multivariate relationships while also reducing if not eliminating reliance on the fallible memories or note-taking systems of reviewers. It is this feature of meta-analysis that would allow researchers in bilingualism to comprehensively account for unique sample attributes, for example, or for the many unique measures that might be employed for a given construct. The coding process also requires that the meta-analyst produce operational definitions that can be coded for reliably across the sample, thus potentially introducing a level of scientific rigor and transparency to theoretically challenging notions such as heritage learner status, implicit vs. explicit learning, and different levels of proficiency.

A third hallmark of meta-analysis is the use of standardized indices (i.e., effect sizes) to estimate the relationships of interest both overall and as moderated by the substantive and methodological features that are coded. Literature reviews, by contrast, have often relied on tests of statistical inference and the flawed practice of null hypothesis significance testing, which are inherently less precise, less stable, and less informative than effect sizes (e.g., Plonsky, 2015).

Finally, related to the use of effect sizes are the meta-analytic principles of estimation-thinking and synthetic-mindedness (Cumming, 2014). The "synthetic research ethic"

© The Author(s), 2021. Published by Cambridge University Press

CAMBRIDGE UNIVERSITY PRESS

(Norris & Ortega, 2006, p. 4), embodied in part by meta-analysis, involves recognizing that no single study can provide a conclusive answer to any question worth asking (Tryon, 2016). Part of doing so involves an understanding of the error that is always present around our results; to ignore such error is both disingenuous and arguably unethical. We urge researchers to consider the implications of these principles not only when reviewing previous literature but throughout the research cycle and in all the roles we fill (e.g., authors, reviewers, editors, researcher trainers) in an effort to more fully advance our scientific understanding of bilingualism.

Brief description of meta-analyses to date in bilingualism

Given the benefits described in the previous section, it is not surprising that researchers in a wide range of fields – from ecology to medicine to education – have turned to meta-analysis as the means par excellence for synthesizing findings across studies (e.g., Cooper & Hedges, 2009; Ioannidis, 2016). Applications of meta-analysis are now common in the applied language sciences as well, such as in second-language acquisition (see Plonsky, 2017) and, in recent years, in the realm of bilingualism.

Table 1 presents an overview of research syntheses and meta-analyses on bilingualism. As shown in the middle column, a range of major topics are represented. The far-right column indicates the overall (meta-analytic) effects from each study. Approximately half of the studies included here have been concerned with aggregating correlations. Peng et al. (2018), for example, extracted and combined observed correlations from 197 studies of the relationship between working memory and reading comprehension, revealing a mean correlation among bilinguals of $r = .30$. Likewise, based on a sample of 59 unique reports, Jeon and Yamashita (2014) meta-analyzed the relationships between second-language (L2) reading comprehension and a number of related skills including (a) L2 grammar knowledge ($r = .85$), L2 vocabulary knowledge ($r = .79$), and L2 decoding ($r = .56$).

Other meta-analyses in this sample were interested in understanding mean differences between groups, which are generally expressed by a standardized mean difference index such as Cohen's *d* or Hedges *g*. For example, Adesope, Lavin, Thompson, and Ungerleider's (2010) meta-analysis of the cognitive benefits of bilingualism observed on the basis of 63 studies ($N = 6{,}022$) that bilinguals outperform monolinguals on cognitive tasks such as problem-solving, on average, by approximately .4 standard deviations ($g = .41$). We discuss strategies for interpreting effect sizes below.

Finally, in order to present a more inclusive view of the breadth of synthetic techniques, we have also included in Table 1 examples of a 'scoping review' (Visonà & Plonsky, 2020), a 'systematic review' (Hambly, Wren & McLeod, 2013), and a methodological synthesis (Plonsky, Marsden, Crowther, Gass & Spinner, 2020).

Major stages in meta-analysis

We have thus far presented meta-analysis in purely conceptual and straightforward terms: primary studies are collected and coded to obtain overall effects within a given domain. In reality, as with primary research, numerous choices must be made throughout the meta-analytic process, each of which is likely to influence study outcomes (Boers, Bryfonski, Faez, McKay, in press; Norris & Ortega, 2007; Oswald & Plonsky, 2010). In the section that follows, we briefly outline the major stages and some of the decisions they entail. Our intention is not to provide a tutorial, however. For guidance on how to conduct a meta-analysis, see Cooper (2016) and, in the context of the language sciences, Plonsky and Oswald (2015).

Defining the domain and searching for primary studies

In the first stage of a meta-analysis, researchers outline the domain of research that will be the focus of study and decide on designs and variables of interest. It should be emphasized that meta-analytic results are shaped by both the way constructs have been conceptualized and operationalized in primary studies as well as by the scope of the domain in question (i.e., broad and inclusive versus narrow and more specific). For instance, Adesope et al.'s (2010) meta-analysis included only those studies that recruited 'balanced' bilinguals (i.e., equally well-versed in both languages), while studies with L2 learners (i.e., sometimes dubbed 'sequential bilinguals') and/or participants with language impairment were not deemed eligible. Branum-Martin et al. (2012), however, focused exclusively on bilingual children, whereas Hambly et al. (2013) meta-analyzed studies involving both bilingual and multilingual children, those with and without speech sound disorders. Donnelly et al.'s (2015) meta-analysis defined bilingual participants more broadly, including those who attained comparable proficiency levels in both languages and those who used both target languages at least 40% of the time in daily life. Given these unique definitions and operationalizations, it is not surprising that the findings of these reviews differ substantially. We have described the domain here in terms of target populations. However, it is also certainly the case that design features and data collection instruments, among other features, might also be considered in defining the domain of interest.

Having determined the research domain and scope, researchers proceed to the literature search process. As a guiding principle, wider-ranging searches are likely to capture a more comprehensive and thus more precise and more generalizable view of the domain in question. Options abound for conducting such searches, some obvious (library- and web-based databases, references in previous reviews) and others less so (e.g., websites of prominent authors, conference programs, technical reports, direct contact with individual authors) (see Delaney & Tamás, 2018; Plonsky & Brown, 2015).

As candidate reports are examined, an explicit but likely expanding set of eligibility criteria must be applied to determine which studies will be included. It is critical to document this stage of the process and, ideally, to involve multiple reviewers in the decisions of which studies to include. By doing so, the team both reduces false negatives and allows for additional transparency in the form of agreement rates on study selection (Stoll et al., 2019). For example, Adesope et al.'s (2010) search yielded an initial pool of 5,185 articles. After excluding duplicate and ineligible articles based on abstract readings, 157 articles were retained, and inter-coder reliability reached a Cohen's Kappa of .88. The final round of screening involved reading the full texts of 157 articles, and 39 articles representing a total of 63 studies were then included in the meta-analysis (Cohen's Kappa = .92). Branum-Martin et al. (2012) searched both English and Chinese databases and ended up with a sample of 38 primary studies that met the inclusion criteria, two of which were unpublished dissertations and three were articles published in Chinese; however, no information was provided on inter-coder reliability

https://doi.org/10.1017/S1366728920000760 Published online by Cambridge University Press

Table 1. Selected research syntheses and meta-analyses in bilingualism

Authors	Topic (*K**)	Effect size**s
Adesope et al. (2010)	Cognitive benefits of bilingualism (63)	*g* = .41
Avery & Marsden (2019)	Sensitivity to grammatical information (57)	*d* = .20
Bowles (2010)	Reactivity during think-alouds (14)	*d* = 1.16
Branum-Martin et al. (2012)	Cross-linguistic phonological awareness (38)	*r*s = .34-.88
Gunnerud et al. (in press)	Cognitive benefits of bilingualism among children (100)	*g* = .06
Grundy & Timmer (2017)	Bilingualism and working memory (27)	*r* = .20
Hambly et al. (2013)	Bilingualism and speech production (66)	N/A
Jeon & Yamashita (2014)	Correlates of L2 reading comprehension (59)	*r*s = .56, .79, .85
Kalandadze et al. (2019)	Metaphor comprehension, autism, and task type (14)	*g* = .63
Lauro & Schwartz (2017)	Bilingual lexical access (26)	*g* = -.48
Lehtonen et al. (2018)	Bilingualism and executive functioning (152)	*g* = -.08
Li (2016)	Aptitude and L2 learning (66)	*r* = .49
Melby-Lervåg & Lervåg (2011)	Cross-linguistic transfer (47)	*r*s = .16, .54, .44, .60, .24, .46
Melby-Lervåg & Lervåg (2014)	L1 vs. L2 reading (82)	*d*s = -.62, -1.12, -.12
Mukadam et al. (2017)	Bilingualism and cognitive decline (13)	OR = .96
Peng et al. (2018)	Reading and working memory (197)	*r* = .30
Plonsky et al. (2020)	Judgment tasks (302)	*d*s = .14, 1.35
Qureshi (2016)	Age effects (26)	*d* = .46; *r* = -.4
Shin (2020)	Working memory and reading (24)	*r*s = .21, .29, .34
Teimouri et al. (2019)	Anxiety and L2 achievement (96)	*r* = -.36
Visonà & Plonsky (2020)	Arabic as a heritage language (34)	N/A

Notes. **K* denotes the number of primary studies in the sample; ***d* and *g* both represent standardized mean differences; *r* = correlations; OR = odds ratio

during the search process or on the possible presence of publication bias(es).

As shown in Table 1, meta-analytic samples (denoted as *K*) vary widely. There is no strict minimum number of studies to include. However, as with primary research, larger samples ($K > 20$) are preferred because they are more likely to yield stable estimates. In sum, in addition to being comprehensive, the literature search strategy must be concisely yet transparently summarized in the write-up.

Data collection (coding)

The second major stage involves developing a coding scheme that allows for key attributes of primary studies to be documented along with their corresponding effect sizes. The features to be coded depend on the research domain and research questions but can be generally classified broadly as (a) study descriptors (e.g., author(s), title, and other identifiers; characteristics of the sample; aspects of the research design; measures and instrumentation; features associated with methodological quality and transparency) and (b) study outcomes (i.e., effect sizes such as correlation coefficients, Cohen's *d* values, and odds ratios). The coding sheet must be based on a solid understanding of the substantive domain including pertinent variables and methodological practices. It is also necessary to pilot the instrument and to modify it based on the emerging characteristics across primary studies.

Furthermore, it is advisable to recruit and train one or more additional coders to increase the accuracy of coding, with the help of a coding manual (e.g., Marsden et al., 2018; Melby-Lervåg & Lervåg, 2014). When doing so, the researchers should calculate and report an estimate of inter-coder agreement (e.g., Cohen's Kappa, or κ) overall and for each category in the coding sheet (see Norouzian, in press). In Lehtonen et al. (2018), for example, κ was in the range between .83 and 1.00; in Peng et al. (2018), inter-coder reliability ranged from .95 to .98. Of note, if the sample of primary studies is large, researchers may opt to code twice only a set of studies. For example, in Hambly et al. (2013), a sample of 14 studies out of 66 underwent double coding, and the overall inter-coder reliability was 86%; Melby-Lervåg and Lervåg (2011) double-coded all studies in the sample, but calculated inter-coder reliability for only 30% of the sample; Plonsky et al. (2020) double-coded 15% of the sample of 302 studies, which exceeds the often recommended minimum number of 20 studies (Lipsey & Wilson, 2001).

To promote transparency and accuracy in research reporting, researchers are encouraged to make their codebook available as an appendix and/or online (e.g., on IRIS [iris-database] as in Marsden et al., 2018) and to carefully explain their coding strategies as well as difficulties encountered along the way. For instance, participants' language proficiency is often reported idiosyncratically across studies. A meta-analyst might, therefore, prepare a set of decision rules to allow for more consistent and transparent coding of this variable. Missing data almost invariably come into play as well. The meta-analyst must decide in such cases whether primary studies with unreported features will be excluded (e.g., Adesope et al., 2010) or whether missing data will be imputed or, more likely, requested from primary authors,

https://doi.org/10.1017/S1366728920000760 Published online by Cambridge University Press

a strategy employed by a number of meta-analyses in the field of bilingualism (Lauro & Schwartz, 2017; Lehtonen et al., 2018; Melby-Lervåg & Lervåg, 2014; Mukadam et al., 2017; Peng et al., 2018). We encourage authors who do so to provide the response rate for the sake of greater transparency (see, e.g., Nicklin & Plonsky, 2020). Whether or not the missing data are provided, in such a situation, meta-analysts might find themselves feeling constrained by the shortcomings or even confounds in the designs and reporting practices of primary research.

To conclude, the quality of the instrument and the accuracy of the coded data exert a substantial influence on meta-analytic results. Therefore, it is vital that the coding scheme includes variables and values pertinent to the research questions posed and that concurrent double coding is performed in a consistent and reliable fashion.

Analysis

After all effect sizes have been compiled or calculated, the meta-analysis proper (i.e., the aggregation of primary effects) can take place. In theory, this process is fairly simple: the synthesist calculates the average of the effect sizes found in the sample and its corresponding variance. In practice, however, a number of decisions must be made concerning, for example, whether and how to account for data dependencies that arise when a single study includes multiple groups/conditions, measures, and/or testing points. It is also common to weight study effects by sample size (e.g., Li, 2010) or by inverse variance (Qureshi, 2016) such that those with less sampling error contribute more to the meta-analytic mean. Corrections for statistical artifacts such as measurement error (reliability) and range restriction can also be applied.

Related to effect size weighting is the decision of model selection (fixed vs. random effects). The fixed effects model assumes that studies included in the meta-analysis are sampled from populations which have one fixed or 'true' effect size. Any deviations from that value are therefore assumed to be due to sampling error alone. By contrast, the random effects model allows for the presence of systematic variability in observed effects due to moderators. A full discussion of these models is outside the scope of this paper. We argue, however, that a random effects model is likely more appropriate for bilingualism researchers due to the complexities of language learning, usage, and so forth (Oswald & Plonsky, 2010). Research has also indicated that the real-world data in the social sciences are likely to have variable population parameters, making the random effects model preferred (Field & Gillett, 2010).

Regardless of any weighting procedures that are applied, the 'grand mean' meta-analysis produces an estimate of an overall relationship – typically a difference between conditions or a correlation between variables – several examples of which are found in Table 1. However, we are often just as or even more interested in the variability in effects around that mean. In the next step, moderator analysis, the meta-analyst examines substantive and methodological features in relation to (i.e., as predictors of) study outcomes. Consider Qureshi's (2016) meta-analysis of age effects, for example. The overall difference of $d = .46$ between early and late bilinguals was strongly moderated by whether the participants were living in a second ($d = .68$) vs. foreign language ($d = -.09$) environment.

Finally, inherent to many domains is the potential for bias in available effects. Publication bias, also referred to as the 'file-drawer problem', often occurs because studies with statistically significant results are more likely to be published (Cooper, 2016; Field & Gillett, 2010). When such a bias is present, the sample of observed effects is likely to present an overestimate of the population effect. Several strategies can be applied to minimize (pre-emptively), estimate, and reduce the presence of bias. These include more thorough searches to obtain unpublished and 'gray' literature and diagnostic tools such as funnel plots as seen in Lehtonen et al. (2018) and Melby-Lervåg and Lervåg (2011, 2014). A wide range of analytical and statistical tools are also available such as comparing effect sizes from published and unpublished studies (e.g., Avery & Marsden, 2019), calculating the fail-safe N statistic (Grundy & Timmer, 2017), p-curve (Mahowald et al., 2016), and others (see Rothstein, Sutton & Borenstein, 2005).

Table 2. Field-specific benchmarks for interpreting effect sizes (d, r, R^2) in L2 research

Effect size	Small-ish (25^{th} percentile)	Medium-ish (50^{th} percentile)	Large-ish (75^{th} percentile)
d (between)	0.40	0.70	1.00
d (within)	0.60	1.00	1.40
r	0.25	0.40	0.65
R^2	0.18	0.32	0.51

Interpretation

Parallel to a primary study, the final stage in meta-analysis involves interpreting outcomes. Here, too, a number of considerations come into play. Making sense of the effect sizes resulting from overall and moderator analyses can feel somewhat subjective. As a starting point, meta-analysts in bilingualism might consider existing benchmarks. Plonsky and Oswald (2014) generated a distribution of observed d and r values in L2 research based on a sample of 91 meta-analyses and 346 primary studies. Similarly, as part of a methodological synthesis of the use of multiple regression, Plonsky and Ghanbar (2018) aggregated R^2 values from a sample of 541 regression analyses found in 171 published reports. Both studies then proposed tentative but field-specific benchmarks for interpreting the different effect sizes of interest, as shown in Table 2.

We want to emphasize that such benchmarks are nothing more than a starting point for gauging the magnitude of effects within the field. There are a number of additional factors that should also be taken into consideration when interpreting meta-analytic effects. These include, for example, theoretical and/or practical significance (e.g., implications for health or educational policy), comparable domains, change over time in the domain's theoretical development and/or methodological practices, attenuation due to statistical artifacts (e.g., measurement error, range restriction), publication bias(es), ceiling effects, and over/under-sampling among certain populations (see related discussions in Avery & Marsden, 2019; Brysbaert, 2019; Plonsky & Oswald, 2014).

Conclusion

We have sought in this paper to both raise awareness of the potential of meta-analysis and to lay out some of the many decision points that meta-analysts necessarily encounter. In doing so,

https://doi.org/10.1017/S1366728920000760 Published online by Cambridge University Press

we have argued that meta-analysis represents a powerful approach for synthesizing findings across primary studies that improves on the challenges facing more traditional reviews. It is for these and other reasons that we anticipate applications of meta-analysis will continue to increase in tandem with continued expansion and accumulation of findings in the field of bilingualism.

References

Adesope OO, Lavin T, Thompson T and Ungerleider C (2010) A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research* **80**, 207–245.

Avery N and Marsden E (2019) A meta-analysis of sensitivity to grammatical information during self-paced reading: Towards a framework of reference for reading time effect sizes. *Studies in Second Language Acquisition* **41**, 1055–1087.

Boers F, Bryfonski L, Faez F and McKay T (in press) A call for cautious interpretation of meta-analytic reviews. *Studies in Second Language Acquisition*.

Bowles M (2010) Features that make a task amenable to think-aloud: A meta-analysis of studies investigating the validity of think-alouds on verbal tasks. In *The think-aloud controversy in second language research* (Chapter 3). New York: Routledge.

Branum-Martin L, Tao S, Garnaat S, Brunta F and Francis DJ (2012) Meta-analysis of bilingual phonological awareness: Language, age, and psycholinguistic grain size. *Journal of Educational Psychology* **104**, 932–944.

Brysbaert M (2019) How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition* **2**, 1–38.

Cooper H (2016) *Research synthesis and meta-analysis: A step-by-step approach* (5th ed.). Thousand Oaks, CA: Sage.

Cooper H and Hedges LV (2009) Research synthesis as a scientific process. In Cooper H, Hedges LV and JC Valentine (eds), *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Russell Sage Foundation.

Cumming G (2014) The new statistics: Why and how. *Psychological Science* **25**, 7–29.

Delaney A and Tamás PA (2018) Searching for evidence or approval? A commentary on database search in systematic reviews and alternative information retrieval methodologies. *Research Synthesis Methods* **9**, 124–131.

Donnelly S, Brooks PB and Homer BD (2016) Examining the bilingual advantage on conflict resolution asks: A meta-analysis. In Noelle DC, Dale R, Warlaumont AS, Yoshimi J, Matlock T, Jennings CD and Maglio PP (eds), Proceedings of the 37th Annual Meeting of the Cognitive Science Society (pp. 596–601). Cognitive Science Society.

Field AP and Gillett R (2010) How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology* **63**, 665–694.

Grundy JG and Timmer K (2017) Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research* **33**, 325–340.

Gunnerud HL, ten Braak D, Reikerås EKL, Donolato E and Melby-Lervåg M (in press) Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin*. http://dx.doi.org/10.1037/bul0000301

Hambly H, Wren Y and McLeod S (2013) The influence of bilingualism on speech production: A systematic review. *International Journal of Language & Communication Disorders* **48**, 1–24.

Ioannidis JPA (2016) The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Milbank Quarterly* **94**, 485–514.

Jeon EH and Yamashita J (2014) L2 reading Comprehension and its correlates: A meta-analysis. *Language Learning* **64**, 160–212.

Kalandadze T, Bambini V and Næss K-A (2019) A systematic review and meta-analysis of studies on metaphor comprehension in individuals with autism spectrum disorder: Do task properties matter? *Applied Psycholinguistics* **40**, 1421–1454.

Lauro J and Schwartz AI (2017) Bilingual non-selective lexical access in sentence contexts: A meta-analytic review. *Journal of Memory and Language* **92**, 217–233.

Lehtonen M, Soveri A, Laine A, Järvenpää J, de Bruin A and Antfolk J (2018) Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin* **144**, 394–425. doi:10.1037/bul0000142

Li S (2010) The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning* **60**, 309–365.

Li S (2016) The construct validity of language aptitude. *Studies in Second Language Acquisition* **38**, 801–842.

Lipsey DB and Wilson MW (2001) *Practical meta-analysis*. Sage.

Mahowald K, James A, Futrell R and Gibson E (2016) A meta-analysis of syntactic priming in language production. *Journal of Memory and Language* **91**, 5–27.

Marsden E, Thompson S and Plonsky L (2018) A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics* **39**, 861–904.

Melby-Lervåg M and Lervåg A (2011) Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading* **34**, 114–135.

Melby-Lervåg M and 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**, 409–433.

Mukadam N, Sommerlad A and Livingston G (2017) The relationship of bilingualism compared to monolingualism to the risk of cognitive decline or dementia: A systematic review and meta-analysis. *Journal of Alzheimer's Disease* **58**, 45–54.

Nicklin C and Plonsky L (2020) Outliers in L2 research: A synthesis and data re-analysis from self-paced reading. *Annual Review of Applied Linguistics* **40**, 26–55.

Norouzian R (in press) Inter-rater reliability in second language meta-analyses: The case of categorical moderators. *Studies in Second Language Acquisition*.

Norris JM and Ortega L (2006) The value and practice of research synthesis for language learning and teaching. In Norris JM and Ortega L (eds), *Synthesizing research on language learning and teaching*. Philadelphia: John Benjamins, pp. 3–50.

Norris JM and Ortega L (2007) The future of research synthesis in applied linguistics: Beyond art or science. *TESOL Quarterly* **41**, 805–815.

Oswald FL and Plonsky L (2010) Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics* **30**, 85–110.

Peng P, Barnes M, Wang CC, Wang W and Li S (2018) A meta-analysis on the relation between reading and working memory. *Psychological Bulletin* **144**, 48–76.

Plonsky L (2015) Statistical power, p values, descriptive statistics, and effect sizes: A "back-to-basics" approach to advancing quantitative methods in L2 research. In Plonsky L (ed), *Advancing quantitative methods in second language research*. New York, NY: Routledge, pp. 23–45.

Plonsky L (2017) Quantitative research methods. In Loewen S and Sato M (eds), *The Routledge handbook of instructed second language acquisition*. New York, NY: Routledge, pp. 505–521.

Plonsky L and Ghanbar H (2018) Multiple regression in L2 research: A methodological synthesis and guide to interpreting R^2 values. *Modern Language Journal* **102**, 713–731.

Plonsky L, Marsden E, Crowther D, Gass S and Spinner P (2020) A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research* **36**, 583–621..

Plonsky L and Oswald FL (2014) How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning* **64**, 878–912.

Plonsky L and Oswald FL (2015) Meta-analyzing second language research. In Plonsky L (ed), *Advancing quantitative methods in second language research*. New York, NY: Routledge, pp. 106–128.

Rothstein HR, Sutton AJ and Borenstein M (eds) (2005) *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester: John Wiley.

Qureshi MA (2016) A meta-analysis: Age and second language grammar acquisition. *System* **60**, 147–160.

https://doi.org/10.1017/S1366728920000760 Published online by Cambridge University Press

Shin J (2020) A meta-analysis of the relationship between working memory and L2 reading comprehension: Does task type matter? *Applied Psycholinguistics* **41**, 873–900.

Stoll CRT, Izadi S, Fowler S, Green P, Suls J and Colditz GA (2019) The value of a second reviewer for study selection in systematic reviews. *Research Synthesis Methods* **10**, 539–545.

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

Tryon WW (2016) Replication is about effect size: Comment on Maxwell, Lau, and Howard (2015). *American Psychologist* **71**, 236–237.

Visonà M and Plonsky L (2020) Arabic as a heritage language: A scoping review. *International Journal of Bilingualism* **24**, 559–615.

https://doi.org/10.1017/S1366728920000760 Published online by Cambridge University Press