Bilingualism: Language and Cognition

cambridge.org/bil

Towards a credibility revolution in bilingualism research: Open data and materials as stepping stones to more reproducible and replicable research

Cylcia Bolibaugh[1], Norbert Vanek[2] and Emma J. Marsden[1]

[1]University of York, York, UK and [2]University of Auckland, Auckland, NZ

Review Article

Cite this article: Bolibaugh C, Vanek N, Marsden EJ (2021). Towards a credibility revolution in bilingualism research: Open data and materials as stepping stones to more reproducible and replicable research. *Bilingualism: Language and Cognition* 1–6. https://doi.org/10.1017/S1366728921000535

Received: 23 February 2021
Revised: 16 July 2021
Accepted: 16 July 2021

Keywords:
open data; open code; open materials; reproducibility; age effects; bilingual advantage

Address for correspondence:
Cylcia Bolibaugh, Centre for Research in Language Learning and Use, Department of Education, University of York, YO10 5DD, United Kingdom.
E-mail: cylcia.bolibaugh@york.ac.uk

Abstract

The extent to which findings in bilingualism research are contingent on specific analytic choices, experimental designs, or operationalisations, is currently unknown. Poor availability of data, analysis code, and materials has hindered the development of cumulative lines of research. In this review, we survey current practices and advocate a credibility revolution in bilingualism research through the adoption of minimum standards of transparency. Full disclosure of data and code is necessary not only to assess the reproducibility of original findings, but also to test the robustness of these findings to different analytic specifications. Similarly, full provision of experimental materials and protocols underpins assessment of both the replicability of original findings, as well as their generalisability to different contexts and samples. We illustrate the review with examples where good practice has advanced the agenda in bilingualism research and highlight resources to help researchers get started.

Introduction

A recent commentary on the bilingual advantage in executive function (Duñabeitia & Carreiras, 2015) optimistically concludes that *veritas est temporis filia*, truth is the daughter of time. The phrase captures the notion that the scientific enterprise is cumulative, and though false pistes might be taken, these are ultimately corrected. Nonetheless, there are reasons to hold a more sober view (Ioannidis, 2012). As Duñabeita and Carreiras highlight, one precondition for progress is an unbiased publishing system in which the robustness of research is the primary criterion for publication. Another is the complete disclosure of all steps and processes underlying published outputs. Unfortunately, complete disclosure has been the exception rather than the norm (Young, Ioannidis & Al-Ubaydli, 2008).

Bilingualism research, and some areas within bilingualism research in particular, have not made the progress that one might expect, given 'a global research effort of unprecedented magnitude' (Hartsuiker, 2015, p.336). In the present piece, we discuss ways in which minimum standards of methodological transparency, necessary for both reproducibility and replicability[1], can overcome the crisis of confidence in bilingualism research. We argue that these minimum standards are not only necessary to distinguish between 'helpful' and 'unhelpful' replication attempts (National Academies of Sciences & Medicine, 2019) and thus build a cumulative scientific enterprise, but that they also enable a series of methodological innovations that have the potential to accelerate the research cycle. To briefly preview our argument, full disclosure of data and code is necessary not only to assess the reproducibility of original findings, but also to test the robustness of these findings to different analytic specifications. Similarly, full provision of experimental materials and protocols underpins assessment of both the replicability of original findings, as well as their generalisability to different contexts and samples. We illustrate each section of the review with recent impactful examples and follow with pointers for those looking to share their data and code, and materials and protocols.

Open data and analytic code

Sharing of data and code (such as R scripts, or SPSS syntax that can be generated through the graphical user interface) underpins computational reproducibility, and is necessary for the verification of individual studies, but also confers other benefits which we elaborate below.

[1]We use the following definitions from National Academies of Science and Medicine (2019) throughout: "Reproducibility means … obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis. Replicability means obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data".

CAMBRIDGE UNIVERSITY PRESS

Computational reproducibility

In many cases, exact replication of a study can be prohibitive or difficult. The reasons underlying this difficulty may be related to the characteristics of a particular sample of participants (e.g., Kindertransport survivors in Schmid, 2002; adult international adoptees in Pallier, Dehaene, Poline, LeBihan, Argenti, Dupoux & Mehler, 2003), or the design of the study itself (e.g., the Barcelona Age Factor which exploited a change in curricular language provision; Muñoz, 2006), among other factors. Longitudinal and panel studies (e.g., Xavier Vila, Ubalde, Bretxa & Comajoan-Colomé, 2018) may be particularly difficult to replicate. In these cases, an "attainable minimum standard" (Peng, 2011) for verifying scientific claims is via an assessment of the computational reproducibility of the analyses.

Providing the data and computer code necessary to re-run analyses and re-create the results in published outputs can be key to catching potentially harmful errors at an early stage. Surveys of statistical errors at the reporting stage (Nuijten, Hartgerink, van Assen, Epskamp & Wicherts, 2016), as well as the coding stage (Ziemann, Eren & El-Osta, 2016) have found that these appear in up to half of sampled articles, and frequently have implications for the substantive conclusions drawn (see Herndon, Ash & Pollin, 2014 for a notable coding error).

The extent of computational reproducibility within bilingualism research is currently unknown, but efforts from adjoining disciplines may be indicative of general trends. Plonsky, Egbert and Laflair (2015) solicited datasets from 255 candidate studies published between 2002 and 2012 in *Language Learning* and *Studies in Second Language Acquisition*, and received 37 (approximately 15%). Two similar studies reported only slightly higher figures in journals with mandatory data sharing policies: Stodden, Seiler and Ma (2018) estimated that 44% of the 204 articles they sampled from *Science* had at least some recoverable data and code, and that 26% of the sample were potentially reproducible. Hardwicke, Mathur, MacDonald, Nilsonne, Banks, Kidwell, Hofelich Mohr, Clayton, Yoon, Henry Tessler, Lenne, Altman, Long and Frank (2018) found that nearly half of articles sampled from *Cognition* (85/174) had datasets which were likely to be reusable. The authors were able to reproduce published values in 63% of a subset of these articles, though author assistance was needed for half the cases. Thus despite growing numbers of calls for sharing of data as a matter of course, the realities of data sharing in related disciplines suggest that it is still relatively uncommon, and the actual reproducibility of results likely to be low.

Though reanalyses of existing studies in bilingualism are relatively few to date, they have the potential to make significant impact. One early example is Vanhove's (2013) reanalysis of data from DeKeyser, Alfi-Shabtay and Ravid (2010), using piecewise regression to test the long-contested relationship between age of acquisition and ultimate attainment. Results pointed to a need to qualify earlier conclusions since a discontinuity in age effects was only found in one of the two datasets reanalysed. Evaluating the technical validity of earlier statistical approaches brought a twofold benefit. It highlighted the problem of arbitrary binning of continuous variables, and emphasised the usefulness of reanalysing existing studies by moving beyond linear statistics where curvilinear approaches are more suitable.

Analytic robustness

Beyond assuring the verifiability of results, the sharing of data and code enables a more stringent test of the robustness of published findings to different specifications of analysis. Researchers who prepare a data set for analysis must make a series of decisions regarding which data to combine, transform, or exclude. In a given study, for example, a researcher may need to decide whether and how to combine aspects of language experience and use into a single bilingualism quotient, which indices of executive function tasks to use as predictors, and how to treat outliers in response times. Choices such as these are frequently referred to as researcher degrees of freedom (Simmons, Nelson & Simonsohn, 2011). While many such choices appear methodologically or substantively arbitrary, they can be consequential to the inferences drawn. A recent study asking 29 teams of analysts to independently answer a research question given the same data set (Silberzahn, Uhlmann, Martin, Anselmi, Aust, Awtrey, Bahník, Bai, Bannard, Bonnier, Carlsson, Cheung, Christensen, Clay, Craig, Dalla Rosa, Dam, Evans, Flores Cervantes, Fong, Gamez-Djokic, Glenz, Gordon-McKeon, Heaton, Hederos, Heene, Mohr, Hofelich Högden, Hui, Johannesson, Kalodimos, Kaszubowski, Kennedy, Lei, Lindsay, Liverani, Madan, Molden, Molleman, Morey, Mulder, Nijstad, Pope, Pope, Prenoveau, Rink, Robusto, Roderique, Sandberg, Schlüter, Schönbrodt, Sherman, Sommer, Sotak, Spain, Spörlein, Stafford, Stefanutti, Tauber, Ullrich, Vianello, Wagenmakers, Witkowiak, Yoon & Nosek, 2018) concluded that 'significant variation in the results of analyses of complex data may be difficult to avoid, even by experts with honest intentions' (p.338).

Looking to meta-research in related disciplines can inform us about the robustness of analyses in bilingualism. Plonsky et al. (2015) followed their survey of data availability in *Language Learning* and *Studies in Second Language Acquisition* with an assessment of the robustness of the subset of studies with usable data; when they applied a testing method that made different assumptions (viz., bootstrapping), they found that a quarter of previously significant focal tests were no longer significant. A different approach to assessing robustness was taken by Steegen, Tuerlinckx, Gelman and Vanpaemel (2016), who constructed a series of datasets by iterating through all reasonable choices in data processing. By repeating their analysis over these differently constructed datasets (more than 100 reanalyses), the authors demonstrated the power of a multiverse analysis to 'reduce the problem of selective reporting by making the fragility or robustness of the results transparent, and … [identify] the most consequential choices' (p. 707).

A similar approach was recently adopted by Poarch, Vanhove and Berthele, (2019), who carried out a multiverse analysis of the bilingual executive function advantage in bidialectals. By documenting a range of possible analyses when varying data exclusion criteria, and the coding of the flanker and Simon effects, the authors illustrated the potential effects of subjective choices on result interpretations. This study is a particularly useful example of good practice in the context of substantial variation across studies on the effects of bilingualism on executive function.

Research synthesis and planning

A final benefit of providing data and code alongside published outputs concerns the development of research syntheses, and the planning of future research. Aggregating findings across a line of research is typically carried out through meta-analyses of summary effects from primary studies, yet the basic information required to compute effects is often missing from primary reports (Larson-Hall & Plonsky, 2015). A culture of archiving data will not only increase

the number of studies included in future meta-analyses, but also enable more sophisticated research syntheses using either trial or participant level data (see the special issue of *Psychological Methods*, Curran, 2009; Glass, 2000). The power of this approach to detect small effects, and hence adjudicate between inconsistent findings, can be seen in a study by Nicenboim, Vasishth and Rösler (2019) addressing the recent large scale, multisite 'failure to replicate' anticipatory effects in language comprehension (Nieuwland, Politzer-Ahles, Heyselaar, Segaert, Darley, Kazanina, Von Grebmer Zu Wolfsthurn, Bartolozzi, Kogan, Ito, Mézière, Barr, Rousselet, Ferguson, Busch-Moreno, Fu, Tuomainen, Kulakova, Husband, Donaldson, Kohu, Rueschemeyer & Huettig, 2018). In a meta-analysis with trial-level data, the authors found evidence for a clear, but small effect of prediction, that only emerged when analysed across multiple studies. More realistic estimation of effect sizes will further enable researchers to consider what effect sizes might be considered relevant, and shift to planning of studies powered to detect the 'smallest effect size of interest' (Lakens, Scheel & Isager, 2018). Asking researchers to consider what effect sizes can be studied reliably may also mitigate future 'decline effects' like that identified by de Bruin and Della Sala (2015) in the bilingual advantage literature. The decline effect refers to a phenomenon whereby strong initial evidence for a novel effect diminishes as a line of research develops. De Bruin and Della Salla attribute the decline effect to a combination of statistical regression to the mean, and difficulties in publishing small or null effects.

Good practice in reproducibility

The examples discussed above highlight ways in which integrating reproducibility into bilingualism research has helped the field make theoretical advances. Nonetheless, they are not particularly illuminating to the researcher looking to share their data and analysis code now. An overview of issues involved in making research data available for dissemination can be found in the data sharing primer from UKRN (Towse et al., 2020). Further tangible guidance is available in recently published tutorials such as Klein, Hardwicke, Aust, Breuer, Danielsson, Hofelich Mohr, Ijzerman, Nilsonne, Vanpaemel and Frank (2018), as well as the inaugural issue of Advances in Methods and Practices in Psychological Science (Challenges in Making Data Available, 2018). Here, we briefly signpost some additional resources that can help implement the key principles of organisation, documentation, automation and dissemination necessary for reproducibility.

The simplest way to ensure the reproducibility of a research project is to plan for it from the beginning. This is the approach taken by the Project Tier Protocol (https://www.projecttier.org/), an opinionated framework that provides a clear template and workflow for creating and documenting a reproducible research project. The Project Tier protocols are a good entry point for researchers working with commercial analysis software such as SPSS, Stata, or SAS; they contain guidance on how to manually create meta-data, data codebook, and read-me files that supplement the syntax files available from these packages – and ensure that the distinction between processed data and raw or original data is preserved.

For researchers working in open source software environments like the R computing language (R Core Team, 2013), a number of packages that assist reproducible project management are available. One comprehensive package, Workflowr (Blischak, Carbonetto & Stephens, 2019), combines literate programming and version control with reproducibility checks, and is aimed at those with minimal experience with version control systems. Beyond R, Code Ocean (Clyburne-Sherin, Fei & Green, 2019) (https://codeocean.com/) provides online modular containers for a large number of widely used software environments along with code and data, and runs in a browser. CodeOcean is useful for helping researchers without experience of using dedicated containerisation software to manage their code dependencies and guard against parts of their analysis 'breaking' as software packages are updated; additionally each capsule is assigned a DOI to ensure that it is persistently findable.

Open materials and protocols

The availability of data elicitation materials and study protocols underpins the development of systematic lines of research. When materials are available, researchers can evaluate the comparability of constructs and their operationalisations across studies. Establishing the commensurability of data elicitation measures also allows researchers to analyse pooled data across studies, in Integrative Data Analyses, an alternative to meta-analyses (Bauer & Hussong, 2009). Finally, open materials and protocols are especially important for the planning of replication studies. Replication studies play a central role in the accumulation of evidence for or against a hypothesis (Leek & Peng, 2015), and, when preregistered and conducted at scale (e.g., Morgan-Short, Marsden, Heil, Issa, Leow, Mikhaylova, Mikołajczak, Moreno, Slabakova & Szudarski, 2018), may present the least biased way of estimating effects: a recent comparison of 15 meta-analyses to multi-site, pre-registered replications on the same topics found that meta-analyses systematically inflated effect sizes even after corrective measures had been taken (Kvarven, Strømland & Johannesson, 2019).

As is the case with sharing of data and code, existing meta-research suggests that materials and protocols in bilingualism research are not yet routinely archived or shared. In a methodological synthesis of the use of self-paced reading in studies investigating adult bilingual participants, Marsden, Thompson and Plonsky (2018) found that only 4% of 71 eligible studies had full materials available, and 77% gave just one brief example of stimuli. A survey of instrument availability across three journals in second language research found that only 17% of instruments were available between 2009 and 2013 (Derrick, 2016). Likewise, Hardwicke, Wallach, Kidwell, Bendixen, Crüwell, & Ioannidis (2020), sampling a broader range of social science literature between 2014–2017, found that materials availability was indicated for only 11% of 151 sampled studies, and protocols availability for none. The lack of detailed protocols is particularly worrying in light of findings that researchers believe that unreported lab practices may influence the outcomes of their research (Brenninkmeijer, Derksen & Rietzschel, 2019).

Unfortunately, the current lack of transparency regarding instrumentation and protocols presents an important threat to the quality of replication efforts. A synthesis of replication studies in second language learning (Marsden, Morgan-Short, Thompson & Abugaber, 2019) found that only 3 of the original 67 studies that were replicated had provided all of their materials. In the absence of full reporting of materials and instructions, non-replications become contentious rather than informative, generating debate around the fidelity of the replication attempt rather than an understanding of the limiting conditions of an effect (e.g., Grundy & Bialystok, 2019).

From this admittedly low base, a growing number of initiatives and individual examples of good practice are addressing the conditions underpinning replicability. Firstly, care has been paid to theorising and measuring language proficiency (Kaushanskaya, Blumenfeld & Marian, 2019), language exposure (Anderson, Mak, Chahi & Bialystok, 2018), and language dominance (Dunn & Fox Tree 2009); this care is now being extended to examine constructs and tasks in executive function (e.g., Paap & Greenberg, 2013, Poarch & Van Hell, 2019). More generally, materials availability is increasing. Digital objects associated with published reports in bilingualism research can now be found in generalist (e.g., Figshare, the Open Science Framework), and discipline specific repositories (e.g., the IRIS Repository of Instruments for Research into Second Languages). As a community supported repository archiving instruments, materials and stimuli for research into second and foreign languages, IRIS now also hosts special collections of instruments (e.g., 63 self-paced reading tasks). Finally, replicability and reproducibility have become priorities for a growing number of bilingualism researchers, e.g., Poort and Rodd (2018)'s publically accessible project archiving data elicitation materials, protocols, data, and analysis scripts exemplifies the systematic and transparent reporting necessary for future close replication. Beyond the efforts of individual researchers, a recent call for registered replications of second language studies with non-academic participant samples (Andringa & Godfroid, 2019) is systematically addressing questions around the contextual generalisability of L2 research. Similar efforts will be needed to more explicitly consider the role of bilinguals' histories of language learning and use (Mishra, 2018).

Good practice in replicability

In order to replicate a research study, one needs the full set of stimuli (e.g., pictures, participant instructions, software setup, test items, response options, distractors) used to elicit the data. As this level of detail is usually more information than is conventionally accepted in a publication methods section, archiving all non-proprietary material in a public repository, and linking the material to the publication itself is an important first step. Practical guidance on sharing materials can be found in a recent tutorial from the founders of Databrary (Gilmore, Lorenzo Kennedy & Adolph, 2018).

Researchers have a number of choices regarding where to host their materials. While many behavioural tasks can now be shared in task specific repositories (e.g., PsychoPy, jsPsych, and lab.js experiments can be shared on the Pavlovia platform, pavlovia.org), and other researchers may share materials on their own websites or general repositories like the Open Science Foundation, there is a further tangible benefit to also archiving protocols, instruments and materials in domain specific repositories such as IRIS. Domain specific materials repositories increase the comparability of sources of data; for example, once uploaded to IRIS, materials are associated with rich, searchable meta-data, with parameters for Research Area, Instrument Type, Data Type, Participant Type, Language Feature, among many others. These collections in turn enable meta-research on constructs and methods, such as that exemplified by Marsden et al. (2018)'s methodological synthesis of the use of self-paced reading in second language research.

While archiving data elicitation materials is an important and relatively straightforward step, it may not be sufficient. Going forward, a key shortcoming to address is the lack of standardised formats to document data elicitation procedures. A method which may have promise, and which is being trialled in conjunction with Stage 1 Registered Reports, is the use of video recording of study protocols (Heycke and Spitzer, 2019; Spitzer and Heycke, 2020). The potential of this approach can be seen in the Databrary repository, which not only specifically encourages the archiving of video documentation of study procedures, participant instructions, apparatuses and testing contexts, but also provides tools to code, quantify and systematically compare differences across studies (Gilmore & Adolph, 2017).

Recommendations going forward

This review has attempted to illustrate something every researcher knows: the lifecycle of any research study is beset by a series of decisions, many of which are essentially arbitrary, whose consequences are usually unknown. Debates regarding tasks, coding, and analysis seldom arise, except when inconsistencies and failures to replicate threaten previously established findings. Compounding these issues, our current publication practices neither prioritise nor straightforwardly accommodate complete disclosure of research procedures.

We have argued that one simple remedy with the potential to minimise unhelpful sources of non-replicability is to ensure that published reports are accompanied by the archiving, and public release where possible, of study materials, protocols, data and analysis scripts. Of course, transparency does not guarantee quality, and further recommendations exist, including the need to make sure that data adhere to FAIR principles (Wilkinson, Dumontier, Aalbersberg, Appleton, Axton, Baak, Blomberg, Boiten, da Silva Santos, Bourne, Bouwman, Brookes, Clark, Crosas, Dillo, Dumon, Edmunds, Evelo, Finkers, Gonzalez-Beltran, Gray, Groth, Goble, Grethe, Heringa, 't Hoen, Hooft, Kuhn, Kok, Kok, Lusher, Martone, Mons, Packer, Persson, Rocca-Serra, Roos, van Schaik, Sansone, Schultes, Sengstag, Slater, Strawn, Swertz, Thompson, Van Der Lei, Van Mulligen, Velterop, Waagmeester, Wittenburg, Wolstencroft, Zhao & Mons, 2016), that results can be reproduced with the code provided, and that analyses are preregistered (with Chambers, 2013; or without peer review) – but we believe that full methodological transparency represents an initial, attainable minimum standard.

Researchers may hesitate to release their instruments, data and code for a number of reasons (Houtkoop, Chambers, Macleod, Bishop, Nichols & Wagenmakers, 2018), among them the worry that scrutiny will uncover mistakes. As increasingly sophisticated analyses and complex experimental paradigms become more common, this is unavoidable. A credibility revolution in bilingualism research will require a culture in which mistakes are viewed as inevitable, and practices are designed to collectively mitigate their impact (Rouder, Haaf & Snyder, 2019).

References

Anderson JA, Mak L, Chahi AK and Bialystok E. (2018) The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods* **50**, 250–263.

Andringa S and Godfroid A. (2019) Call for Participation. *Language Learning* **69**, 5–10. https://doi.org/10.1111/lang.12338

Bauer DJ and Hussong AM. (2009) Psychometric approaches for developing commensurate measures across independent studies: traditional and new models. *Psychological Methods* **14**, 101–125. https://doi.org/10.1037/a0015583

Blischak JD, Carbonetto P and Stephens M. (2019) Creating and sharing reproducible research code the workflowr way. *F1000Research* **8**, 1749. https://doi.org/10.12688/f1000research.20843.1

Brenninkmeijer J, Derksen M and Rietzschel E. (2019) *Informal Laboratory Practices in Psychology*. https://doi.org/10.1525/collabra.221

Challenges in Making Data Available. (2018) In *Advances in Methods and Practices in Psychological Science* (special section; Vol. 1, Issue 1).

Chambers CD. (2013) Registered reports: a new publishing initiative at Cortex. *Cortex* **49**, 609–610. https://doi.org/10.1016/j.cortex.2012.12.016

Clyburne-Sherin A, Fei X and Green SA. (2019) Computational Reproducibility via Containers in Psychology. *Meta-Psychology* **3**. https://doi.org/10.15626/MP.2018.892

Curran PJ (ed.) (2009) Special Issue: Multi-Study Methods for Building a Cumulative Psychological Science. In *Psychological Methods* **14**, 2, https://psycnet.apa.org/PsycARTICLES/journal/met/14/2

Czapka S, Wotschack C, Klassert A and Festman J. (2019) A path to the bilingual advantage: Pairwise matching of individuals. *Bilingualism: Language and Cognition* 1–11. https://doi.org/10.1017/S1366728919000166.

de Bruin A and Della Sala S (2015) The decline effect: How initially strong results tend to decrease over time. *Cortex* **73**, 375–377. https://doi.org/10.1016/j.cortex.2015.05.025

DeKeyser R, Alfi-Shabtay I and Ravid D. (2010) Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics* **31**, 413–438. https://doi.org/10.1017/S0142716410000056

Derrick DJ. (2016) Instrument Reporting Practices in Second Language Research. *TESOL Quarterly* **50**, 132–153. https://doi.org/10.1002/tesq.217

Duñabeitia JA and Carreiras M. (2015) The bilingual advantage: Acta est fabula? *Cortex* **73**, 371–372. https://doi.org/10.1016/j.cortex.2015.06.009

Dunn AL and Fox Tree JE. (2009) A quick, gradient Bilingual Dominance Scale. *Bilingualism: Language and Cognition* **12**, 273–289. https://doi.org/10.1017/S1366728909990113

Gilmore RO and Adolph KE. (2017) Video can make behavioural science more reproducible. *Nature Human Behavior* **2017**. https://doi.org/10.1038/s41562-017-0128.

Gilmore RO, Lorenzo Kennedy J and Adolph KE. (2018) Practical Solutions for Sharing Data and Materials From Psychological Research. *Advances in Methods and Practices in Psychological Science* **1**, 121–130. https://doi.org/10.1177/2515245917746500

Glass GV. (2000, January). *Meta-Analysis at 25*. https://www.gvglass.info/papers/meta25.html

Grundy JG and Bialystok E. (2019) When a "Replication" Is Not a Replication. Commentary: Sequential Congruency Effects in Monolingual and Bilingual Adults. *Frontiers in Psychology* **10**, 797. https://doi.org/10.3389/fpsyg.2019.00797

Hardwicke TE, Wallach JD, Kidwell MC, Bendixen T, Crüwell S and Ioannidis JPA. (2020) An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014-2017). *Royal Society Open Science* **7**, 190806. https://doi.org/10.1098/rsos.190806

Hardwicke TE, Mathur MB, MacDonald K, Nilsonne G, Banks GC, Kidwell MC, Hofelich Mohr A, Clayton E, Yoon EJ, Henry Tessler M, Lenne RL, Altman S, Long B and Frank MC. (2018) Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science* **5**, 180448. https://doi.org/10.1098/rsos.180448

Hartsuiker RJ. (2015) Why it is pointless to ask under which specific circumstances the bilingual advantage occurs. *Cortex* **73**, 336–337. https://doi.org/10.1016/j.cortex.2015.07.018

Herndon T, Ash M and Pollin R. (2014) Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics* **38**, 257–279. https://doi.org/10.1093/cje/bet075

Heycke T and Spitzer L. (2019) Screen Recordings as a Tool to Document Computer Assisted Data Collection Procedures. *Psychologica Belgica* **59**, 269–280. https://doi.org/10.5334/pb.490

Houtkoop BL, Chambers C, Macleod M, Bishop DVM, Nichols TE and Wagenmakers E-J. (2018) Data Sharing in Psychology: A Survey on Barriers and Preconditions. *Advances in Methods and Practices in Psychological Science* **1**, 70–85. https://doi.org/10.1177/2515245917751886

Ioannidis JPA. (2012) Why Science Is Not Necessarily Self-Correcting. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* **7**, 645–654. https://doi.org/10.1177/1745691612464056

Kaushanskaya M, Blumenfeld HK and Marian V. (2019) The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition* 1–6. https://doi.org/10.1017/S1366728919000038.

Klein O, Hardwicke TE, Aust F, Breuer J, Danielsson H, Hofelich Mohr A, Ijzerman H, Nilsonne G, Vanpaemel W and Frank MC. (2018) A Practical Guide for Transparency in Psychological Science. *Collabra: Psychology* **4**, 20. https://doi.org/10.1525/collabra.158

Kvarven A, Strømland E and Johannesson M. (2019) Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-019-0787-z. Published online by Springer Nature, 23 December 2019

Lakens D, Scheel AM and Isager PM. (2018) Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science* **1**, 259–269. https://doi.org/10.1177/2515245918770963

Larson-Hall J and Plonsky L. (2015) Reporting and Interpreting Quantitative Research Findings: What Gets Reported and Recommendations for the Field. *Language Learning* **65**, 127–159. https://doi.org/10.1111/lang.12115

Leek JT and Peng RD. (2015) Opinion: Reproducible research can still be wrong: Adopting a prevention approach. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 1645–1646. https://doi.org/10.1073/pnas.1421412111

Marsden EJ, Morgan-Short K, Thompson S and Abugaber D. (2019) Replication in Second Language Research: Narrative and Systematic Reviews and Recommendations for the Field. *Language Learning* **68**, 321–391. https://doi.org/10.1111/lang.12286

Marsden EJ, Thompson S and Plonsky L. (2018) A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics* **39**, 861–904. https://doi.org/10.1017/S0142716418000036

Mishra RK. (2018) *Bilingualism and Cognitive Control*. Springer.

Morgan-Short K, Marsden E, Heil J, Issa B, Leow RP, Mikhaylova A, Mikołajczak S, Moreno N, Slabakova R and Szudarski P. (2018) Multi-site replication in SLA research: Attention to form during listening and reading comprehension in L2 Spanish. *Language Learning* **68**, 392–437. https://doi.org/10.1111/lang.12292

Muñoz C. (2006) *Age and the Rate of Foreign Language Learning*. Multilingual Matters. https://play.google.com/store/books/details?id=1C_-zfVkmOkC

National Academies of Sciences & Medicine. (2019) *Reproducibility and Replicability in Science*. The National Academies Press. https://doi.org/10.17226/25303

Nicenboim B, Vasishth S and Rösler F. (2019) *Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly available data*. https://doi.org/10.31234/osf.io/2atrh. Available online February 28, 2019.

Nieuwland MS, Politzer-Ahles S, Heyselaar E, Segaert K, Darley E, Kazanina N, Von Grebmer Zu Wolfsthurn S, Bartolozzi F, Kogan V, Ito A, Mézière D, Barr DJ, Rousselet GA, Ferguson HJ, Busch-Moreno S, Fu X, Tuomainen J, Kulakova E, Husband EM, Donaldson DI, Kohu Z, Rueschemeyer S.-A. and Huettig F (2018) Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife* **7**. https://doi.org/10.7554/eLife.33468

Nuijten MB, Hartgerink CHJ, van Assen M. A. L. M., Epskamp S and Wicherts JM (2016) The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods* **48**, 1205–1226. https://doi.org/10.3758/s13428-015-0664-2

Paap KR and Greenberg ZI. (2013) There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology* **66**, 232–258. https://doi.org/10.1016/j.cogpsych.2012.12.002

Pallier C, Dehaene S, Poline J.-B., LeBihan D, Argenti A.-M., Dupoux E and Mehler J. (2003) Brain imaging of language plasticity in adopted adults: can a second language replace the first? *Cerebral Cortex* **13**, 155–161. https://doi.org/10.1093/cercor/13.2.155

Peng RD. (2011) Reproducible research in computational science. *Science* **334**, 1226–1227. https://doi.org/10.1126/science.1213847

Plonsky L, Egbert J and Laflair GT. (2015) Bootstrapping in Applied Linguistics: Assessing its Potential Using Shared Data. *Applied Linguistics* **36**, 591–610. https://doi.org/10.1093/applin/amu001

Poarch GJ and Van Hell JG. (2019) Does performance on executive function tasks correlate? Evidence from child trilinguals, bilinguals, and second language learners. In IA Sekerina, L Spradlin and V Valian (eds.), *Bilingualism, executive function, and beyond: Questions and insights* (pp. 223–236). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.57.14poa

Poarch GJ, Vanhove J and Berthele R. (2019) The effect of bidialectalism on executive function. *International Journal of Bilingualism* **23**, 612–628. https://doi.org/10.1177/1367006918763132

Poort ED and Rodd JM. (2018, June 8). The cognate facilitation effect in bilingual lexical decision is influenced by stimulus list composition [Experiment 2]. Retrieved from osf.io/zadys

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rouder JN, Haaf JM and Snyder HK. (2019) Minimizing Mistakes in Psychological Science. *Advances in Methods and Practices in Psychological Science* **2**, 3–11. https://doi.org/10.1177/2515245918801915

Schmid MS. (2002) *First language attrition, use and maintenance: The case of German Jews in Anglophone countries*. John Benjamins Publishing Company.

Silberzahn R, Uhlmann EL, Martin DP, Anselmi P, Aust F, Awtrey E, Bahník Š, Bai F, Bannard C, Bonnier E, Carlsson R, Cheung F, Christensen G, Clay R, Craig MA, Dalla Rosa A, Dam L, Evans MH, Flores Cervantes I, Fong N, Gamez-Djokic M, Glenz A, Gordon-McKeon S, Heaton TJ, Hederos K, Heene M, Mohr AJ, Hofelich Högden F, Hui K, Johannesson M, Kalodimos J, Kaszubowski E, Kennedy DM, Lei R, Lindsay TA, Liverani S, Madan CR, Molden D, Molleman E, Morey RD, Mulder LB, Nijstad BR, Pope NG, Pope B, Prenoveau JM, Rink F, Robusto E, Roderique H, Sandberg A, Schlüter E, Schönbrodt FD, Sherman MF, Sommer SA, Sotak K, Spain S, Spörlein C, Stafford T, Stefanutti L, Tauber S, Ullrich J, Vianello M, Wagenmakers E. -J., Witkowiak M, Yoon S and Nosek BA (2018) Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices in Psychological Science* **1**, 337–356. https://doi.org/10.1177/2515245917747646

Simmons JP, Nelson LD and Simonsohn U. (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* **22**, 1359–1366. https://doi.org/10.1177/0956797611417632

Spitzer L and Heycke T. (2020) *Preregistration: Videos in peer-review of Registered Reports*. *PsychArchives*. https://doi.org/10.23668/PSYCHARCHIVES.3127

Steegen S, Tuerlinckx F, Gelman A and Vanpaemel W. (2016) Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* **11**, 702–712. https://doi.org/10.1177/1745691616658637

Stodden V, Seiler J and Ma Z. (2018) An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 2584–2589. https://doi.org/10.1073/pnas.1708290115

Towse J, Rumsey S, Owen N, Langford P, Jaquiery M and Bolibaugh C. (2020, October 30). Data Sharing: A primer from UKRN. https://doi.org/10.31219/osf.io/wp4zu

Vanhove J. (2013) The critical period hypothesis in second language acquisition: a statistical critique and a reanalysis. *PloS One* **8**, e69172. https://doi.org/10.1371/journal.pone.0069172

Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J.-W., da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, Van Der Lei J, Van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J and Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018. https://doi.org/10.1038/sdata.2016.18

Xavier Vila F, Ubalde J, Bretxa V and Comajoan-Colomé L (2018) Changes in language use with peers during adolescence: a longitudinal study in Catalonia. *International Journal of Bilingual Education and Bilingualism* 1–16. https://doi.org/10.1080/13670050.2018.1436517

Young NS, Ioannidis JPA and Al-Ubaydli O. (2008) Why current publication practices may distort science. *PLoS Medicine* **5**, e201. https://doi.org/10.1371/journal.pmed.0050201

Ziemann M, Eren Y and El-Osta A. (2016) Gene name errors are widespread in the scientific literature. *Genome Biology* **17**, 177. https://doi.org/10.1186/s13059-016-1044-7