



CRITICAL COMMENTARY

Replication studies in second language acquisition research: *Definitions, issues, resources, and future directions*

Introduction to the special issue

Kevin McManus 

Department of Applied Linguistics, Penn State University, University Park, PA 16802, USA
Email: kmcmanus@psu.edu

(Received 16 October 2024; Revised 26 November 2024; Accepted 28 October 2024)

Abstract

A long-standing concern in the field of second language acquisition is that replication studies are not only infrequent but also poorly designed, reported, and labeled. This special issue responds to an urgent need for action by showcasing eleven high-quality replication studies. In doing so, this collection highlights exemplary standards in replication study design and reporting. This introduction to the special issue provides readers with a point of reference for what replication research is, including why replication studies are needed, issues about originality and innovation in replication research, how replication studies can be designed and conducted, and recent advances and resources to support future replication efforts in the field. The introduction concludes with an overview of each study in the special issue, highlighting its main components and discussing how the replication strengthens the field and advances knowledge and understanding about the topic.

A long-standing concern in the field of second language acquisition (SLA) is that replication studies are infrequent (Polio & Gass, 1997; Porte, 2012; Santos, 1989; Tarone, Swain, & Fathman, 1976). This situation threatens the validity and reliability of SLA research as well as the growth of the discipline because confirmed trends in research cannot be distinguished from chance findings and artifacts of a specific research method (Mackey, 2012; Porte, 2013; Porte & McManus, 2019; Valdman, 1993). In addition, replication studies are often poorly designed, reported, and labeled (Marsden, Morgan-Short, Thompson, & Abugaber, 2018; McManus, 2022a; Polio, 2012; Yamashita & Neiriz, 2024), echoing broader concerns about low study quality, transparency, and the need for methodological reform (e.g., Plonsky, 2013, 2024; Riazi & Farsani, 2023; Sudina, 2021). The consequences of this chronic replication problem are severe because a lack of replication (a) harms the credibility of the field's key claims

and theories about SLA, (b) weakens attempts to inform language learning policy and pedagogy, and (c) impedes scientific progress.

This special issue responds to an urgent need for action by showcasing eleven high-quality replication studies. In addition to demonstrating how and why replication is essential to the growth and credibility of the discipline, this collection exemplifies what high-quality replication studies look like and how they can be designed, carried out, and reported. Before looking into the detail of the special issue's eleven replication studies, this introduction provides a point of reference for what replication research is, including why replication studies are needed, issues about originality and innovation in replication research, how replication studies can be designed and conducted, and recent advances and resources to support future replication efforts in the field.

What is replication research?

Replication is a research method used to confirm, consolidate, and extend knowledge and understanding in empirical fields of study (Brandt et al., 2014; Isager et al., 2023; Porte & McManus, 2019). A replication study works toward this goal by repeating a study's research design and methods with or without changes, collecting new data, and systematically comparing the previous study's results with those from the new study (Marsden et al., 2018; McManus, 2022b; Nosek & Errington, 2020). As a result, a replication study is a new empirical study with new data that purposefully addresses the same or similar research question and/or hypothesis as a previous study (McManus, 2023; National Academies of Sciences, Engineering, and Medicine, 2019; Schmidt, 2009). Because one aim of carrying out a replication study is to determine the extent to which repeating a previous study's design and methods leads to comparable patterns of results and conclusions, replication provides a systematic framework for reconsidering, refining, extending, and sometimes limiting prior research findings. This is why replication has long been considered an essential part of the research process (King & Mackey, 2016; Peterson & Panofsky, 2021; Porte, 2012). Replication studies are key to scientific progress precisely because they systematically revisit prior claims to confirm their stability and detect potential anomalies. It is through this essential process of confirming results, detecting anomalies, and understanding their nature that new lines of research can emerge.

Given these features of the replication research process, it is troubling that replication studies in SLA are still claimed to be infrequent and poorly designed, reported, and disseminated (Marsden et al., 2018; Polio & Gass, 1997; Porte, 2012). One outcome of this situation is that calls for replication are now relatively common across the social sciences (Language Teaching Review Panel, 2008; Plucker & Makel, 2021; Zwaan, Etz, Lucas, & Donnellan, 2018), with journal special issues highlighting the specific ways that replication studies support disciplinary growth (Godfroid & Andringa, 2023; Tschichold, 2023). In SLA, field-specific resources now exist to support the conduct of replication studies (Marsden et al., 2018; McManus, 2023, 2024b; Porte & McManus, 2019).

When conducting a replication study, the aim is to design and report a new empirical study that repeats a previous piece of research in some way, with systematic comparison throughout (Brown, 2012; Porte & McManus, 2019). By conducting a close replication study, in which almost all elements of the initial study's design and methods are retained without modification (Brandt et al., 2014; LeBel, McCarthy, Earp, Elson, & Vanpaemel, 2018; McManus, 2022b), the researcher's aim is to understand how an

intentional, carefully selected, well-motivated, and principled modification to the initial study's method impacts the results, if at all. For example, a replication that modifies a prior study's pedagogical intervention while retaining all other aspects of the research design and methods can be used to understand how that intentional modification influenced the study's findings and conclusions (e.g., McManus & Marsden, 2018). In this way, replication research allows the stability of a study's claims to be examined in a principled manner.

Even though determining the stability of previous results is recognized as an essential first step in developing any research program (Gass, Loewen, & Plonsky, 2021; Isager et al., 2023; Mackey, 2012; Schmidt, 2009), this feature is frequently absent in how SLA research is conducted and reported. In other words, systematically revisiting a prior study to confirm its results before extending that line of research is rarely practiced (but see Ellis, Hafeez, Martin, Chen, Boland, & Sagarra, 2014; Morgan-Short et al., 2018). Integrating replication into the research process (i.e., how we develop research questions, design studies, generate claims, and assess theories) is required to support the growth and longevity of the discipline in a principled and systematic way. For example, even though it is well understood that the amounts and types of prior knowledge and experiences that learners bring to learning a new language can substantially influence the trajectories of new language learning (Duff & Byrnes, 2019; Ellis, 2019; Lantolf, Poehner, & Swain, 2018), no single study can examine all contextual influences on learning outcomes (e.g., L2 proficiency, L1 background, prior immersion and instructed experiences, social networks). This is because each study can only ever provide a snapshot of learners' knowledge and use of language at that one point in time. Replication is needed to determine the stability of important and influential claims as well as to investigate variables unaccounted for in previous research. The more stable a claim turns out to be through replication, the more confident we can be that what has been observed and reported is a confirmed trend (rather than a chance finding, for instance). By not probing the stability of claims made in prior research through replication, we continue to build a discipline "on the basis of scanty evidence" (Polio & Gass, 1997, p. 500).

Originality and innovation in replication research

Unfortunately, claims that replication studies lack originality and/or innovation are not only common but they also misguided. Furthermore, these criticisms serve only to harm the discipline. Does repeating aspects of a prior study's research design and methods mean that replication studies lack originality and innovation (Martin & Clarke, 2017; Zwaan et al., 2018)? Regrettably, the field's practices reinforce this view. For example, some journals (e.g., *Applied Psycholinguistics*) distinguish between "original research article" and "replication research article," which, while likely unintentional, suggests that replication studies are not to be considered original. While concerns about originality are thought to have limited the conduct of replication studies in SLA research (Marsden et al., 2018; Porte, 2012; but see McManus, 2022a), it is misleading and incorrect to claim that replication studies cannot make original contributions.

Originality refers to studies designed to detect problems "in the existing archetype to the extent that it provokes other researchers and their understanding to further study and analyze so as to answer the questions raised" (Alajami, 2020, p. 4; see also Dean, Hender, Rodgers, & Santanen, 2006; Gordon, 2007). In other words, originality

describes work designed to advance knowledge and understanding by investigating questions and/or problems and/or detecting anomalies in prior work. In addition, original research brings together new and existing information, including, but not limited to, new research questions/hypotheses, previously-reported methods, and new results (Dirk, 1999; Guetzkow, Lamont, & Mallard, 2004; Shibayama & Wang, 2020). Based on this understanding, it is difficult to see how a replication study that meets these criteria could be claimed to lack originality. As a reminder, a replication study repeats a previous study's research design and methods with or without modification, collects new data, and systematically compares the earlier study's findings with those from the new research (Mackey, 2012; Nosek & Errington, 2020; Polio & Gass, 1997; Porte & McManus, 2019). In this way, a key aim of replication research is to confirm, consolidate, and extend knowledge and understanding within empirical fields of study by confronting existing claims and ideas using new evidence (Brandt et al., 2014; Isager et al., 2023; Long, 1993; Santos, 1989). Thus, claims that replication research is unoriginal are misguided and not consistent with contemporary understandings of originality.

Innovation is also vital to the growth and strengthening of the discipline, but it, too, is misunderstood as describing something completely new, untested, and radical in nature with incremental innovation not counting (Kahn, 2018; Nind, Wiles, Bengry-Howell, & Crow, 2013; Wiles, Crow, & Pain, 2011). Clearly, setting this as a target for a research study would be both impossible to achieve and undesirable, rendering most SLA research lacking innovation. Rather, innovation exists along a continuum, is not a binary phenomenon, and is defined by small, gradual changes to current approaches (Ledgerwood, 2014; Long, 1993). This view of innovation is consistent with contemporary understandings of innovation as a cumulative process in which new lines of research emerge by building on established knowledge. In the next section, I review some of the ways that innovation and originality apply to the design and conduct of replication studies in SLA, highlighting how we can better support scientific progress by using replication to systematically revisit prior claims to confirm their stability and detect potential anomalies. I conclude this section by stressing that we must eliminate outdated notions and misconceptions of replication as lacking originality and innovation. These attitudes are unhelpful, misguided, and harmful to the discipline.

Approaches to replication

Because a replication study can come in different forms, several categorizations have been developed to convey the amount of intentional change implemented in the replication (Brandt et al., 2014; LeBel et al., 2018; Marsden et al., 2018; Porte & McManus, 2019). In an *exact replication*, for example, the selected study's entire design, methods, and procedure are followed without alteration. In a *close replication*, one major variable is intentionally modified, while an *approximate replication* intentionally modifies no more than two major variables (for approaches to variable selection and modification in replication studies, see Brandt et al., 2014; LeBel et al., 2018; Porte & McManus, 2019). In a *conceptual replication*, however, almost all aspects of the previous study can be modified because the aim here involves "repetition of a test of a hypothesis or a result of earlier research with different methods" (Schmidt, 2009, p. 91). It is important to note, however, that drawing meaningful comparisons between studies (i.e., initial and replication) is difficult when several major modifications are integrated into the replication. This makes conceptual replications an inappropriate

choice for revisiting prior studies to confirm results. At the same time, exact replication in the social sciences is probably one of the most difficult, if not impossible, types of replication study to conduct (Stroebe & Strack, 2014). Even though it might be possible to locate and use the previous study's materials and analyses, keeping the data sample the same is very difficult when working with human subjects (Nosek & Errington, 2020). Of course, there could be similarities (e.g., L2 English learners in Austria and Germany), but several factors make the samples different (e.g., contexts, times, individual backgrounds).

In terms of using replication to revisit prior studies to confirm results and conclusions, close and approximate replication studies that include the least intentional modification are the most desirable because they allow for the greatest amount of systematic comparison (see Figure 1 for an overview of replication approaches).

The amount of modification between the initial study and the replication should always be reported in the study's title, abstract, and main text (Appelbaum et al., 2018; McManus, 2022a, 2024), with labels such as *close replication* or *approximate replication* as well as summary tables to document between-study differences and similarities. This means that a replication should be clearly identified as a replication study from the outset. Furthermore, it is important for authors to identify, categorize, and report elements of the replication that are (a) the same as, (b) close to, or (c) different from the initial study (Brandt et al., 2014; LeBel et al., 2018; Marsden et al., 2018; McManus, 2023). In a close replication study, for instance, in which one variable is intentionally modified and all other differences are minimal, a minor modification refers to elements of the replication that are very close to those in the initial study, whereas a major modification refers to variables that are different between the studies. In addition, a major modification will represent the research focus of the replication, whereas a minor modification should not. In a close replication investigating whether web-based instruction influences the learning outcomes reported in a classroom-based study, for example, modifying the modality of instruction from in-person to web-based would represent an element of the initial study's design that is different in the replication. In this example, the modality of the instruction (from in-person to web-based) can be categorized as a major variable modification. Communicating and justifying all between-study differences can also serve several benefits. For example, readers can judge if a modification appears justified or not and also where the replication is on the continuum from close to conceptual. Reporting all between study differences also allows future work to evaluate the boundary conditions on a particular effect and to develop guidance for evaluating magnitudes of change in replication research.

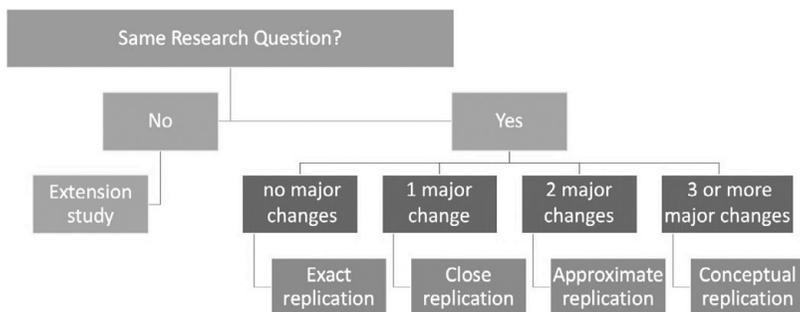


Figure 1. Overview of approaches to replication.

Supporting future replication efforts in the field

Supporting and improving the conduct and dissemination of replication studies is vital to the growth of the discipline. Doing so requires the field to develop and apply standards and expectations in design and reporting so that replication studies can meet their aims to confront and revisit existing understanding with new evidence. As a result, replication reporting must include the following components, as a minimum: a clear rationale for why the replication was carried out, full and transparent descriptions of differences and similarities in the design, methods, and results, and a framework for determining and evaluating the replicability of effect. Critically, however, these components are neither encouraged nor discussed in general reporting guidelines for empirical research in SLA, including research methodology textbooks (e.g., Mackey & Gass, 2022; Rose, McKinley, & Baffoe-Djan, 2020) and journal reporting guidelines (e.g., *Language Learning, Studies in Second Language Acquisition*; see McManus, 2024a). This is because the aims of most empirical studies (or extension studies) are different from those of replication studies. For one, extension studies aim to extend a current line of research in new ways (e.g., contexts, methods, populations), whereas a replication study aims to critically revisit one study to better understand its findings and impact on the discipline. This difference in aims impacts how the study is designed and reported, which explains why the information reported in replication studies is often piecemeal, incomplete, and difficult to understand and evaluate. There is thus a compelling need for disciplines to develop resources to support the design, conduct, and reporting of replication studies (e.g., Language Teaching Review Panel, 2008; Marsden et al., 2018; McManus, 2024a; Porte, 2012; Porte & McManus, 2019). In the next sections, I focus on three areas of current and future investment central to the growth of the discipline and the role of replication studies in that growth: reporting, design, and evaluation.

Replication reporting

In SLA research, improvements to the quality and transparency of reporting are needed (see Plonsky, 2013, 2024), and this situation has been noted for replication studies as well (Brown, 2012; Marsden et al., 2018; Polio & Gass, 1997; Porte & McManus, 2019). The task is to develop and implement reporting strategies to maximize the transparency and openness of a replication project, including, but not limited to, transparent study labeling, definition, and description of the replication approach, description and justification of between-study modifications, comparative reporting in the replication study's design and methods, and evaluation of the initial study's results in light of those reported in the replication. One resource contributing to innovation in this area includes a new standard for replication studies in applied linguistics (McManus, 2024a, see also 2023; Porte & McManus, 2019). This standard outlines basic expectations in replication reporting so that outcomes can be better understood and evaluated. Table 1, adapted from McManus (2024a), outlines the minimum requirements in replication reporting for each section of a publication. The replication studies included in this special issue closely align with this standard for replication reporting.

Replication design and study selection

In terms of supporting the design of high-quality replication studies, Brandt et al. (2014) and Isager et al. (2023) are two innovations that review and propose new ways for approaching the design of replication studies to maximize impact.

Table 1. Reporting expectations for replication studies in SLA research (adapted from McManus 2024a)

| Paper section | Features to be included |
|-----------------------------|--|
| Title | <ul style="list-style-type: none"> Identify topic, variables, and/or theoretical issues investigated State approach to replication (e.g., close replication, approximate replication) |
| Abstract | <ul style="list-style-type: none"> State motivation for replication, initial study selection, and the replication approach (e.g., close) Describe modifications in the replication |
| Introduction and background | <ul style="list-style-type: none"> Outline key findings with comparison to the initial study Justify why initial study was replicated Review initial study (design, method, findings), and other replications |
| Research questions | <ul style="list-style-type: none"> Define replication approach, describe and justify modifications Use initial study's research questions, if available, with modifications Describe similarities and differences between initial study and replication, also summarized in a table |
| Method | <ul style="list-style-type: none"> Justify sample size in the replication State if initial study's materials, coding, and analysis procedures were available, used, and/or modified Describe procedure for evaluating replicability of effects |
| Results | <ul style="list-style-type: none"> Report and compare results with initial study, including results, tables, figures Use supplementary materials for full and transparent reporting Revisit research questions |
| Discussion and conclusions | <ul style="list-style-type: none"> Discuss similarities/differences in findings Interpret replication results, accounting for modifications made Discuss limitations and propose future replications |
| Data and materials | <ul style="list-style-type: none"> Make all information necessary to understand, evaluate, and build on replication publicly available, including materials, data, coding and analysis procedures |

First, Brandt et al.'s (2014) "Replication Recipe" guides researchers through a series of questions and decisions for conducting a close replication study (see also McManus, 2023; Porte & McManus, 2019). In addition to spelling out a "convincing close replication par excellence" (e.g., defining the effects and methods that will be replicated a priori, following the initial study's design and methods as closely as possible), Brandt et al. (2014) developed a 36-question guide that researchers can use to design a close replication study. The 36 questions are grouped into the following topics: nature of the effect to be replicated, design of the replication study, differences and similarities between the initial and replication study, analysis and evaluation of the replication study, registering the replication, and reporting the replication study. This guide also exists as a preregistration template on the Open Science Foundation website (OSF, <https://osf.io/4jd46>), which can be used by researchers to preregister a replication study. For example, the *Replication Recipe* begins with several questions about the aims of the proposed replication study, including what the effect is to be replicated, why this effect needs to be replicated, what is the effect size, confidence interval, and sample size of the initial effect, and where was the initial study was conducted (e.g., data collection location, region, country). Together, these questions guide researchers through several important aspects of doing replication research by focusing on what finding(s) the study seeks to replicate, why the replication study is needed/important, what the (statistical) parameters of the initial finding were, and contextual features of the result.

In terms of designing the replication study, researchers are provided with several questions that promote reflection and planning before proceeding with the replication, including whether the initial study's materials are available, where the replication will

take place, what the target sample size is, and with justification. Indeed, addressing these questions before carrying out the replication study is important because it helps researchers consider the feasibility of conducting the replication study. For example, not having access to the initial study's materials can make conducting a replication difficult (but not impossible). Similarly, planning the target sample size helps determine to what extent the required number of participants can be accessed and recruited.

Perhaps one of the most useful elements of the Replication Recipe is documenting differences between the initial and replication studies, an essential component of designing, conducting, and reporting replication studies. The Replication Recipe includes nine questions, each with a response of exact, close, or different; for example:

- The similarities/differences in the instructions are: exact, close, or different
- The similarities/differences in the measures are: exact, close, or different
- The similarities/differences in the stimuli are: exact, close, or different
- The similarities/differences in the procedure are: exact, close, or different
- The similarities/differences in the location are: exact, close, or different
- The similarities/differences between participant populations are: exact, close, or different

Using these questions, researchers can determine to what extent the replication constitutes a close replication attempt (compared with an approximate or conceptual replication). In cases where “different” is selected for most responses, this means that the type of replication study being planned is conceptual (e.g., different instructions, measures, stimuli, location, participant populations).

In addition, the importance of deciding which specific study to replicate cannot be overlooked. To this end, Isager et al. (2023) proposed a decision model for replication selection under resource and knowledge constraints, which is useful for helping researchers think about the reasons for selecting a study for replication. This is because there are often multiple potential candidate studies in a topic/line of research that could be replicated, but constraints will necessarily prevent replication of all of them (e.g., time, resources). The question comes down to how one decides among them.

The literature has proposed several justifications for why a study might be selected for replication, including statistical (e.g., inconsistent results, low statistical power, low precision), theoretical (e.g., initial finding is theoretically important), methodological concerns (e.g., threats to validity), academic impact (e.g., frequently cited study), public/societal impact (e.g., findings used in popular press, education, policy; see Mackey, 2012; Nosek & Lakens, 2013; Porte, 2012, 2013). In terms of a consensus, it is broadly agreed that the claims/studies most in need of replication are those in which (a) the study's design and methods are sound, and (b) the study is important and/or valuable, but (c) the evidence is ambiguous (Brandt et al., 2014; Nosek & Errington, 2020; Porte & McManus, 2019). Thus, claims that are important/valuable with unambiguous evidence are less likely to need corroboration through replication.

Of course, a key question here is how to determine whether a study's design and methods are sound, if the study is valuable and/or important, and if the evidence underlying the claim is ambiguous or not. Even though there is no universally accepted means to answer these questions, there are steps researchers can consider. For instance, the academic, social, and theoretical impact of a study can inform decisions about the importance of the study (e.g., if the article is cited frequently in the field, and if the study's findings are discussed in textbooks). In addition, closely reviewing the analysis and results can provide information about the quality of the evidence provided (e.g., do

the findings come with small effect sizes, wide confidence intervals, and/or from small samples?). On this backdrop, Isager et al. (2023) recommended that researchers prioritize four factors when making decisions about what to replicate:

- Value/importance of the claim
- Uncertainty of the evidence underlying the claim
- Ability of the replication study to reduce uncertainty about the claim
- Costs and feasibility of executing a replication study

Replication evaluation

In addition to questions about replication reporting and design, an important consideration in replication research is determining how to evaluate the replicability of an effect. Or, otherwise said: How can results be compared to determine whether the replication has arrived at the same sets of results as the initial study? Marsden et al. (2018) reported that replications in SLA research mostly used the following methods: narrative comparison, mentioning the findings of the initial study, and dichotomous interpretation using *p*-values. These represent a good starting point, but researchers would benefit from considering additional methods for evaluating the replicability of an effect. One study that represents a particular innovation in this regard that can strengthen replication evaluation and reporting is Errington et al. (2021).

In this study, Errington et al. (2021) developed and examined several criteria to assess the replicability of preclinical research in cancer biology:

- Effect in the initial and replication studies patterns in the same direction and is statistically significant ($p < 0.05$)
- Initial effect size falls within the 95% confidence interval of the replication effect size
- Replication effect size falls within the 95% confidence interval of the initial effect size
- Replication effect size falls within the 95% prediction interval of the initial effect size
- Meta-analysis combining original and replication effect sizes is statistically significant ($p < 0.05$)

Overall, Errington et al. (2021) showed replication rates as follows: 47% for same direction and statistical significance, 25% for the original effect size being inside the 95% confidence interval (CI) of the replication, 48% for the replication effect size being inside the 95% CI of the original, 61% for the replication effect size being inside the 95% prediction interval, and 63% for a criterion based on a meta-analytic combination of the data from the initial experiment and the replication. In addition, Errington et al. (2021) found that null results reported in the initial study were more likely to replicate than positive results in the initial study. Errington et al.'s findings and their recommendations for evaluating replication are useful because they illustrate the many varied ways that results can be evaluated. They also suggest that the most stringent tests of replicability will include more than one assessment. Implementing even just one of these methods would be a solid start toward moving the field beyond narrative comparisons only.

The special issue

To date, replication studies in SLA research have been rare and of varying quality with a preference for conceptual over approximate and close replication designs. In addition,

the transparency and completeness of replication reporting have—in many instances—been low. This special issue responds to these limitations in replication design and reporting with eleven replication studies. Thus, this special issue aims to showcase and highlight exemplary standards in replication study design and reporting to benefit the field. Indeed, each study can be regarded as a role model usable by researchers and students in conducting their own replication studies and teaching others about replication design and reporting.

In compiling the special issue, a call for proposals was launched in December 2022, in which prospective contributors were invited to submit short proposals no longer than four pages that stated (i) the study to be replicated with justifications, (ii) the type of replication study to be completed, (iii) modifications to be implemented in the replication with justifications, (iv) design of the replication study, (v) data analysis plan, and (vi) potential impact. Forty-seven proposals were received, and each was independently evaluated by two reviewers (Kevin McManus and an SSLA editorial board member) using the criteria provided in the call. Following this process, nineteen proposals were selected for potential inclusion, subject to favorable evaluation following external peer review. At that point, authors were provided with recommendations for reporting replication studies as presented in Appelbaum et al. (2018) and Porte and McManus (2019), including:

- Report the type of replication in the title and abstract, as follows: exact, close, approximate, or conceptual.
- Report whether the replication has conditions, materials, and/or procedures that were not part of the initial study.
- Present justifications and descriptions for all additions and variable modifications.
- Summarize all changes between the initial and replication studies in a table.
- Clearly label results from additional or different analyses.
- State the criteria for deciding whether the initial study's results were replicated. Examples of criteria include statistical significance testing, effect sizes, confidence intervals, and Bayes factors in Bayesian methods.
- Include in the discussion a section titled “Future replication research” where suggestions for future replication studies building on the current replication study are made.

This process culminated in a special issue of eleven replication studies involving a variety of topics, populations, replication designs, motivations for replications, and initial studies that integrate open science practices (e.g., open data, open materials, preregistration). In the remainder of this introduction, I provide an overview of each study in this special issue, highlighting its main components and discussing how the replication study strengthens the field and advances knowledge and understanding about the topic.

Barlow et al. (2024) examined the ways that “native” and “non-native” listeners judge “foreign-accented” speech, contributing to an established line of SLA research suggesting that statements produced by “foreign-accented” speakers are less likely to be judged as true than those produced by “native speakers” (e.g., Hanzlíková & Skarnitzl, 2017; Lev-Ari & Keysar, 2010). While some studies suggest that these results might be attributable to processing difficulty, accent-based prejudice, or both, findings are mixed, at best, resulting in very little clarity and consensus on this topic. To address this limitation in the field, Barlow et al. (2024) carried out a close replication study of Boduch-Grabka and Lev-Ari (2021), a study that isolated the impact of processing

difficulty on veracity judgments by exposing listeners to “native” or “non-native” accents prior to completing a veracity judgment task. The close replication retained all aspects of the initial study’s design, with the addition of an explicit bias task to independently examine the contributions of processing difficulty and accent-based prejudice to listeners’ judgments of veracity. The initial study’s audio materials were accessed from a public repository (Open Science Foundation, <https://osf.io>), with additional materials provided by the authors via an email request from Barlow et al. Overall, the initial study’s findings were not replicated: “we did not reproduce the effect of accent on veracity judgments that have been previously reported (Boduch-Grabka & Lev-Ari, 2021; Lev-Ari & Keysar, 2010) or the Boduch-Grabka and Lev-Ari (2021) finding that prior exposure to Polish-accented English speech improves veracity judgments of Polish-accented statements.” Furthermore, results from the explicit bias task did not appear to shed further light on listeners’ judgments. In terms of why Barlow et al.’s findings did not replicate the initial study, the authors suggest that methodological differences, including “insufficient methodological detail provided by the original researchers,” between the studies may explain the different patterning of results (e.g., time frame differences, social, cultural, and political effects, participant recruitment, potential inclusion criteria differences). Going forward, Barlow et al. (2024) recommend that previous studies reporting an accent-based reduction in veracity judgments “may not be robust to replication” and that future replication studies are needed to determine the factors that influence the presence/absence of such effects in listening.

Gudmestad, Edmonds, Henderson, and Lindqvist (2024) examined the extent to which L2 learners’ interpretation of verbal moods in L2 Spanish varied as a function of L2 proficiency and L1 background through a close replication of Kanwit and Geeslin (2014), an important study that “initiated L2 research on the interpretation of variable structures and because it demonstrated that the traditional interpretations of verbal moods in adverbial clauses in Spanish are variable for learners and NSs.” To examine the role of L1 background in the replication, Gudmestad et al. (2024) recruited Swedish-speaking and French-speaking learners of L2 Spanish who completed the same verbal mood interpretation task and grammar test (to assess L2 proficiency) as the initial study, whereas Kanwit and Geeslin (2014) recruited English-speaking learners only. To assess proficiency effects, participants were grouped by (i) institutional course level and (ii) performance on the grammar test, leading to two groups Level 1 and Level 2. In terms of evaluating the replicability of the findings, Gudmestad et al. used three criteria: crosstabulations of responses, statistical significance in the replication, and the direction of effect estimates for statistically significant results in the replication. Overall, the findings broadly confirmed those reported in the initial study. However, some differences emerged from the variable modification (L1 background). For example, while French-speaking learners’ patterns of interpretation of verbal mood patterns did not shift substantially between proficiency levels for interpretation, Swedish-speaking learners showed patterns of interpretation more consistent with the initial study. As discussed in the replication, this pattern of results (among others) suggests caution in the initial study’s proposed developmental stages for mood because L1 background was found to be an important variable not examined in the initial study.

Hamada, Shimizu, Hoshino, Takaki, and Ushiro (2024) conducted a secondary meta-analysis of Jeon and Yamashita (2022), a frequently cited and pedagogically important study in the field, to (a) ascertain whether the initial study’s metanalytic findings can be reproduced and (b) explore the extension of their findings to the L2 SVR model using a meta-analytic structural equation modeling (MASEM) approach.

Furthermore, reproducing meta-analysis is widely agreed to be important (but seldom undertaken) to account for and explore unresolved questions from prior meta-analytic works (Plonsky, 2012; see also Boers, Bryfonski, Faez, & McKay, 2021). Hamada et al. (2024) worked toward this goal by using the same data set and meta-analytic calculations and procedures from the initial study. By adding MASEM (as an extension of the initial study), this secondary meta-analysis has the potential to provide a more nuanced, comprehensive, and robust account of effects on L2 reading outcomes than explored in the initial study (e.g., metalinguistic skills, language comprehension skills, word decoding skills). This is because the initial study used a correlation-based analysis that limits conclusions to a single effect, whereas MASEM allows for more than just one effect to be investigated simultaneously while also managing latent variables. In terms of evaluating the comparability of the results, Hamada et al. compared the average correlation coefficients and their statistical significance as well as standardized mean differences between the studies. Overall, some previously reported effects on reading comprehension were confirmed (e.g., decoding, orthographic knowledge), while others were not (e.g., L2 phonological awareness, L1 reading comprehension). The authors discuss the interpretation of the different sets of results as a potential consequence of data recoding.

Liu and Lu's (2024) close replication study of Römer and Berger (2019) addressed long-standing concerns with data sparsity in learner corpora and the resulting claims about group-level L2 developmental trajectories. Data sparsity, Liu and Lu discuss, refers to the small-scale nature of most existing learner corpora, which can compromise statistical analyses. The issue of data sparsity can also be found in Römer and Berger (2019), a theoretically and empirically important study in the field showing that the associations between verbs and verb argument constructions (VACs) used by German and Spanish learners of English move closer to L1 usage norms as the learners' proficiency increases. In that study, however, which involved correlation analyses of the first version of the EF-Cambridge Open Language Database (Alexopoulou, Geertzen, Korhonen, & Meurers, 2015) and the British National Corpus (BNC Consortium, 2007), a large number of combinations between verbs and VACs in the BNC were absent in the learner corpus. This issue was addressed with log transformation, an approach with several known limitations, especially for counting data with many zeros. Liu and Lu's replication study addressed this concern by retaining the initial study's overall design and methods, making one variable modification: using the *EFCAMDAT2 corpus*, an updated and larger version of the corpus used in the initial study. Also, to address analytical concerns, Liu and Lu (2024) carried out both correlational analyses (as in the initial study) as well as zero-inflated negative binomial regression analyses, a statistical approach that accounts for count data with overdispersion and excess zeros (Winter, Perlman, & Majid, 2018). Liu and Lu's findings replicated those reported in the initial study. For example, both studies reported that learners produced a wider variety of verbs in specific construction types (e.g., V about N) as their L2 proficiency increased. While not discussed in the initial study, Liu and Lu's (2024) analysis also indicated an influence of writing topics on the learners' use of verbs. In addition, Liu and Lu (2024) found that the use of zero-inflated negative binomial (ZINB) models to account for count data with overdispersion and excess zeros corroborated the correlation analyses, showing that (i) L1 verb-VAC frequency was significantly positively associated with L2 verb-VAC frequency and (ii) that increased overlap between the verb-VAC combinations used by L1 users and higher proficiency L2 learners. Taken together, this replication corroborated the initial study's findings while showcasing the advantages of ZINB models in analyzing corpus frequency data with overdispersion and excess zeros.

Huensch's (2024) preregistered close replication study of Darcy, Mora, and Daidone (2016), a pioneering study in the field of SLA that explored the relationships among general cognition and language abilities, examined relationships between L2 speech perception and production and inhibitory control. In that study, Darcy et al. (2016) investigated to what extent variability in L2 pronunciation outcomes could be explained by individual differences in inhibitory control. Using a retrieval-induced inhibition task and measures of perception and production, Darcy et al. found important relationships between inhibition and perception, with little impact of inhibition on production, a finding that might be explained by the type of inhibition task used. Huensch thus addressed this limitation by adding a different measure of inhibition (Stroop task, Simon task) to the study design and keeping all other aspects of the study unmodified. To support replication, the initial study's authors provided their materials and tasks on the Instrument Repository for Research in Second Language Studies (IRIS; see Marsden, Mackey, & Plonsky, 2016). To determine the replicability of an effect, means, and standard deviations from the descriptive statistics of the two studies were used to compute Hedge's g effect sizes and corresponding CIs. Hedge's $g < .40$ and corresponding CIs passing through zero were taken to indicate negligible differences. Statistical tests were interpreted using the directionality (positive vs. negative) and magnitude of the effect size. Overall, the initial study's findings were not replicated: "no strong, clear, or consistent relationship emerges between inhibitory control and L2 perception/production skills." Thus, Huensch explores several factors that may explain these findings, including a lack of statistical power in the initial study.

Given substantial evidence to date about the benefits of providing aural support during reading (or reading-while-listening) for vocabulary learning (e.g., R. Brown, Waring, & Donkaewbua, 2008; Vu & Peters, 2022), very little is known about the ways in which L1 background may interact with the effectiveness of this instructional method, especially when the transparency of grapheme–phoneme correspondences varies between L1 and L2. To this end, the Twilex Group (2024) conducted a preregistered close replication study of Malone (2018), a methodologically rigorous, transparent, and important study in research examining the effectiveness of reading-while-listening, which systematically integrated L1 differences into the study design. In the initial study, the main research goal was to compare the effectiveness of reading-while-listening (RWL) and reading-only (RO) on incidental vocabulary learning, including to what extent does learning depend on the amount of exposure (two vs. four exposures) and working memory (WM). The replication's major variable modification was the L1 background of the learners, achieved by recruiting L1 speakers of Chinese (Mandarin, Cantonese) and Germanic languages (German, Dutch), who were L2 learners of English. The initial study's participants spoke a range of L1s. Any other modifications to the initial study's design and methods were considered minor in scope (e.g., online instead of in-person data collection). The close replication study obtained all materials either through personal communication with the author of the initial study or from the IRIS database. Claims about the replicability of an effect were determined if the replication obtained the same direction and size of the correlations as those reported in the initial study. Overall, this close replication study obtained mixed results in comparison to the initial study. For example, the Twilex Group found few meaningful differences between the RWL and RO reading conditions, regardless of the amount of exposure, whereas the initial study found RWL to be more effective than RO after two exposures. Between-study similarities were found for working memory, though, showing that higher working memory capacities resulted in greater learning gains. In terms of the main objective of the replication, to assess what role L1

background played, the results indicated important interactions among L1 background, WM, and reading condition (RWL, RO). For example, the demand of WM was greater for L1 Chinese learners in RWL conditions. Taken together, this replication study sheds important light on the effectiveness of bimodal input on L2 learning outcomes and the potentially mediating role that L1 background can play.

Revisiting the role of transfer in third language (L3) acquisition, Parrish (2024) reports an approximate replication of Rothman (2011), a theoretically and empirically important study in the field examining how the perceived typological similarity of previously learned languages shaped L3 learning. Indeed, the Typological Primacy Model, an important theory in L3 research, is based on Rothman (2011). In addition, Parrish (2024) notes limitations with the initial study's small sample size and the statistical tests used. In the approximate replication study, two major changes were implemented. First, new materials were created because the ones used in the initial study were lost and thus not available. These materials were recreated based on the descriptions and examples provided in the initial study. Second, new speaker groups were created. In the initial study, speakers were (a) L1 English–L2–Spanish–L3 Brazilian Portuguese and (b) L1 Italian–L2 English–L3 Spanish. In the replication study, mirror image groups of Spanish–English bilinguals were recruited: (a) L1 English–L2 Spanish–L3 Brazilian Portuguese and (b) L1 Spanish–L2 English–L3 Brazilian Portuguese. The sample size was also increased. Overall, Parrish's (2024) analysis showed that some aspects of the initial study's findings replicated while others did not. For example, whereas the initial study showed that both learner groups showed evidence of transfer by Spanish in their L3, these results were only partially replicated by Parrish (2024). The importance of adequate sample sizes is noted as a possible explanation for these findings, along with the use of different data collection materials. All in all, this replication calls for future research to consider the role of sample size in decisions about what to replicate going forward.

Revisiting a central question in bilingualism research about the ways in which L2 speakers access and activate words in their known languages, Pelzl and Van Hell (2024) conducted an approximate replication of Van Hell and Dijkstra (2002), an influential and frequently cited study in the bilingual word processing literature. In addition, Van Hell and Dijkstra (2002) focused on effects from L2 to L1, providing a particularly strong test of language nonselectivity, which serves as an empirical foundation for many theoretical models of bilingual processing in the field (e.g., Dijkstra, Wahl, Buytenhuijs, Van Halem, Al-Jibouri, De Korte, & Rekké, 2019; Shook & Marian, 2013). For these reasons, determining the replicability of Van Hell and Dijkstra's (2002) findings is important for continued theorization and growth in the field. In this approximate replication study, Pelzl and Van Hell (2024) made two major modifications to the initial study. First, the target population was modified from Dutch–English–French trilinguals to Dutch–English bilinguals. This change was implemented to test whether the initial study's findings held when there was no L3 present, thus strengthening and refining the claims made in the initial study. Second, the study was conducted online rather than in-person in a lab context to recruit a more diverse and larger group of participants, thus addressing small sample size concerns in the initial study. All other aspects of the initial study's design and methods were not intentionally modified, with any other changes considered to be minor in scope. Also, since Van Hell was involved in both the initial study and this replication, additional information about the study and access to the materials was available (e.g., providing stimuli not provided in the initial study), thus providing a critical resource for the replication study. In terms of determining the replicability of the findings, Pelzl and

Van Hell's (2024) replication used two indices: the direction and statistical significance of effects. Overall, Pelzl and Van Hell (2024) found the same patterns of results as those reported in the initial study, confirming that L2 knowledge can influence L1 performance in important and detectable ways. Furthermore, by recruiting a large sample of participants, the replication study showed stronger effects than the initial study, some of which did not reach statistical significance. For example, the initial study observed a 10-ms facilitation effect, a result that was not statistically significant, but a larger sample of participants resulted in a facilitation effect that was 19 ms, which did reach statistical significance. Taken together, this replication study not only confirmed the foundational claims made in the initial study but also strengthened them in many cases, thus providing a strong validation of the initial study claims.

In line with Pelzl and Van Hell (2024), the close replication study conducted by Ryan et al. (2024) also included a researcher from the initial study. The study selected for replication was Tavakoli and Foster (2008), a theoretically and empirically important study in the field of SLA that is also frequently cited in the literature. That study investigated the ways in which L2 oral performance was influenced by narrative task design, showing that narratives with both foreground and background information elicited significantly greater syntactic complexity than those with only foreground information. In this close replication study, Ryan et al. (2024) included one intentional modification only by adding L1 literacy. This was achieved by recruiting adult refugees to New Zealand with low L1 literacy. This variable modification is important for several reasons, not only because most SLA research is based on highly literate populations (see Andringa & Godfroid, 2020), but also because L1 literacy can strongly influence language processing and metalinguistic knowledge. Indeed, advanced levels of L1 literacy have been associated with more accurate L2 performance (Havron & Arnon, 2017). Thus, this close replication study set out to examine to what extent L2 oral performance can also be influenced by low levels of L1 literacy, thereby impacting many theories of L2 learning concerning explicit teaching, metalinguistics knowledge, the provision of corrective feedback, and key aspects of language instruction that depend on metalinguistic knowledge and its processing. All other aspects of the initial study's design and methods were not intentionally modified, with all other differences considered minor in scope. Ryan et al. (2024) used indices of direction and statistical significance to determine the replicability of the initial study's findings. Overall, the replication study did not confirm the initial study's findings about interactions between narrative task design and L2 oral performance because background information in the narrative tasks had no impact on the syntactic complexity, lexical diversity, or fluency of learners' performance. For example, participants in the replication recounted narrative tasks with and without background events in comparable ways, whereas the initial study found that background events led to performance with longer and more complex syntactic patterns. Thus, Ryan et al.'s (2024) findings indicated that learners let background events in the narrative pass unremarked, instead focusing on foreground events. Informed by these findings, Ryan et al. (2024) suggest that schooling experience and levels of visual literacy may explain these findings better than levels of low literacy. This is because, the authors argue, the replication did not require literacy-supported cognitive tools (e.g., awareness of word boundaries) to follow sequences of events as presented in picture-based narratives. Thus, as an active area for future SLA research, Ryan et al. (2024) encourage replication of key studies in task-based learning and teaching, which tend to presume experience and familiarity with visual prompts. Such study designs would appear not to work as well with all learner populations.

The last study in this special issue that included a researcher from the initial study is Wong and Prange (2024), a close replication of Wong, Zhao, and MacWhinney (2018), an important study in instructed SLA about the specific contribution of instruction couched in cognitive linguistics when compared with “traditional” instruction. In the initial study, Wong et al. (2018) developed an experimentalized computer-assisted language learning (eCALL) system called the English Preposition Tutor, to investigate the effectiveness of a cognitive linguistics approach to learning English prepositions, which indicated promising findings for this instructional approach. However, as noted by Wong and Prange (2024), the initial study included no delayed posttest, and the statistical analyses (analysis of variance, ANOVA) focused on analyzing separate effects only, leading to a limited understanding of the initial study’s effectiveness in supporting L2 development. Thus, in this close replication study, Wong and Prange (2024) integrated two intentional modifications to the initial design. First, rather than using ANOVAs to analyze separate effects, Bayesian mixed-effects models were used. Importantly, the initial study’s findings were used to inform the study’s priors (see Garcia, 2021; Norouzian, de Miranda, & Plonsky, 2018). Second, a three-week delayed posttest was added to examine the potential for extended effects of learning in the replication. In line with recommended practice, the sample sizes in the instructional groups were also increased. All other aspects were not intentionally modified. Overall, Wong and Prange’s results are broadly consistent with those reported in the initial study. For example, all groups improved significantly following the instruction, as also reported in the initial study. The incorporation of the delayed posttest was able to show that learning gains documented immediately after the instruction were broadly maintained after the instruction. Thus, in sum, Wong and Prange’s replication study was able to broadly confirm the findings reported in the initial study.

In an approximate replication study, Wang and Sun (2024) revisited the Second Language Motivational Self System (Dörnyei, 2005, 2009), an influential theoretical framework in programs of empirical research and theory-building about language learner’s motivation and its role in second language learning. However, recent commentary has expressed concerns with this framework, highlighting issues such as “the fantasy problem,” the “ought-to L2 self problem,” and the “context problem” (Al-Hoorie, Hiver, & In’nam, 2024). Papi *et al.* (2019) responded to these concerns about the Second Language Motivational Self System with the 2×2 model, designed to offer a more nuanced account of motivational dynamics. Indeed, subsequent research across educational and cultural contexts has indicated that the 2×2 model appears to more effectively predict language learning behaviors and outcomes (e.g., Zhou & Papi, 2023). Indeed, Papi and Khajavy (2021) showed how this model can account for language learning outcomes in an Iranian context, thus providing useful information about the applicability of the 2×2 model in a novel context within this line of research. In the current approximate replication, Wang and Sun (2024) sought to revisit and extend this line of research in a Chinese English as a foreign language (EFL) context given the relatively different educational and cultural contexts in Iran and China. For example, students receive English language instruction at a younger age in China, and China’s promotion of English aligns with its broader strategy for global competitiveness, factors that could impact motivational dynamics toward language learning outcomes. In their approximate replication study, Wang and Sun (2024) intentionally modified three variables: learner background, the language used in the questionnaire, and the method of measuring L2 achievement. The replication team used the same questionnaire as used in the initial study (obtained from IRIS) and increased the sample size from 324 in the initial study to 855. Using this design, Wang and Sun (2024) examined to what

extent Papi and Khajavy's (2021) findings hold in a Chinese EFL context. Overall, Wang and Sun (2024) findings broadly replicated those reported in the initial study, especially in terms of how regulatory focus plays a critical role in language learning. One notable difference, however, was that the replication reported stronger impacts of ought-to-selves, anxiety, and enjoyment on language learning outcomes. The authors discuss these different patterns of results in light of the educational and contextual differences between Iran and China.

Acknowledgements. This special issue would not have been possible without the support and expert guidance of many people, especially Susan Gass, Luke Plonsky, Graeme Porte, and Lizz Huntley. I am grateful to all the authors who submitted proposals and carried out the studies, to the reviewers who carefully evaluated the studies reported in the special issue, and to SSLA Associate Editor Kazuya Saito for handling one of the submissions. I appreciate the feedback received by Amanda Huensch, Luke Plonsky, and Graeme Porte on an earlier version of this introduction.

Competing interest. The author declares none.

References

- Alajami, A. (2020). Beyond originality in scientific research: Considering relations among originality, novelty, and ecological thinking. *Thinking Skills and Creativity*, 38, 100723. <https://doi.org/10.1016/j.tsc.2020.100723>
- Alexopoulou, T., Geertzen, J., Korhonen, A., & Meurers, D. (2015). Exploring big educational learner corpora for SLA research: Perspectives on relative clauses. *International Journal of Learner Corpus Research*, 1(1), 96–129. <https://doi.org/10.1075/ijlcr.1.1.04ale>
- Al-Hoorie, A. H., Hiver, P., & In'nam, Y. (2024). The validation crisis in the L2 motivational self system tradition. *Studies in Second Language Acquisition*, 46(2), 307–329. <https://doi.org/10.1017/S0272263123000487>
- Andringa, S., & Godfroid, A. (2020). Sampling Bias and the Problem of Generalizability in Applied Linguistics. *Annual Review of Applied Linguistics*, 40, 134–142. <https://doi.org/10.1017/S0267190520000033>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Barlow, S., Beardsley, G., Bsharah, Z., Crofts, R., De La Rosa, C., Gutierrez, A., ... & Hayes-Harb, R. (2024). The effects of exposure and explicit stereotypes on veracity judgments of Polish-accented English speech: A preregistered close replication and extension of Boduch-Grabka & Lev-Ari (2021). *Studies in Second Language Acquisition*, 1–17. <https://doi.org/10.1017/S0272263124000123>
- BNC Consortium. (2007). *The British National Corpus*, XML Edition. Oxford Text Archive. <http://hdl.handle.net/20.500.14106/2554>.
- Boduch-Grabka, K., & Lev-Ari, S. (2021). Exposing individuals to foreign accent increases their trust in what nonnative speakers say. *Cognitive Science*, 45(11), e13064. <https://doi.org/10.1111/cogs.13064>
- Boers, F., Bryfonski, L., Faez, F., & McKay, T. (2021). A call for cautious interpretation of meta-analytic reviews. *Studies in Second Language Acquisition*, 43(1), 2–24. <https://doi.org/10.1017/S0272263120000327>
- Brandt, M. J., Ijzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, ... & van 't Veer, A. (2014). The Replication Recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217–224. <https://doi.org/10.1016/j.jesp.2013.10.005>
- Brown, J. D. (2012). Writing up a replication report. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 173–197). Cambridge University Press.
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, 20(2), 136–163.
- Darcy, I., Mora, J. C., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing: Inhibitory control and L2 phonology. *Language Learning*, 66(4), 741–773. <https://doi.org/10.1111/lang.12161>

- Dean, D., Hender, J., Rodgers, T., & Santanen, E. (2006). Identifying quality, novel, and creative ideas: Constructs and scales for idea evaluation. *Journal of the Association for Information Systems*, 7(10), 646–699. <https://doi.org/10.17705/1jais.00106>
- Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, 22(4), 657–679. <https://doi.org/10.1017/S1366728918000287>
- Dirk, L. (1999). A measure of originality: The elements of science. *Social Studies of Science*, 29(5), 765–776. <https://doi.org/10.1177/030631299029005004>
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge.
- Dörnyei, Z. (2009). The L2 motivational self system. In *Motivation, language identity and the L2 self* (pp. 9–42). Multilingual Matters.
- Duff, P. A., & Byrnes, H. (2019). SLA across disciplinary borders: Introduction to the special issue. *The Modern Language Journal*, 103, 3–5. <https://doi.org/10.1111/modl.12537>
- Ellis, N. C. (2019). Essentials of a theory of language cognition. *The Modern Language Journal*, 103, 39–60. <https://doi.org/10.1111/modl.12532>
- Ellis, N. C., Hafeez, K., Martin, K. I., Chen, L., Boland, J., & Sagarra, N. (2014). An eye-tracking study of learned attention in second language acquisition. *Applied Psycholinguistics*, 35(3), 547–579. <https://doi.org/10.1017/S0142716412000501>
- Errington, T. M., Mathur, M., Soderberg, C. K., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Investigating the replicability of preclinical cancer biology. *eLife*, 10, e71601. <https://doi.org/10.7554/eLife.71601>
- Garcia, G. D. (2021). *Data visualization and analysis in second language research* (1st ed.). Routledge. <https://doi.org/10.4324/9781003032243>
- Gass, S. M., Loewen, S., & Plonsky, L. (2021). Coming of age: The past, present, and future of quantitative SLA research. *Language Teaching*, 54(2), 245–258. <https://doi.org/10.1017/S0261444819000430>
- Godfroid, A., & Andringa, S. (2023). Uncovering sampling biases, advancing inclusivity, and rethinking theoretical accounts in second language acquisition: Introduction to the special issue *SLA for all?* *Language Learning*, 73(4), 981–1002. <https://doi.org/10.1111/lang.12620>
- Gordon, M. (2007). What makes interdisciplinary research original? Integrative scholarship reconsidered. *Oxford Review of Education*, 33(2), 195–209. <https://doi.org/10.1080/03054980701259642>
- Gudmestad, A., Edmonds, A., Henderson, C., & Lindqvist, C. (2024). The interpretation of verbal moods in Spanish: A close replication of Kanwit and Geeslin (2014). *Studies in Second Language Acquisition*, 1–18. <https://doi.org/10.1017/S027226312400010X>
- Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is originality in the humanities and the social sciences? *American Sociological Review*, 69(2), 190–212. <https://doi.org/10.1177/000312240406900203>
- Hamada, A., Shimizu, H., Hoshino, Y., Takaki, S., & Ushiro, Y. (2024). Robust evidence for the simple view of second language reading: Secondary meta-analysis of Jeon and Yamashita (2022). *Studies in Second Language Acquisition*, 1–18. <https://doi.org/10.1017/S0272263124000226>
- Hanzlíková, D., & Skarnitzl, R. (2017). Credibility of native and non-native speakers of English revisited: Do non-native listeners feel the same? *Research in Language*, 15(3), 285–298. <https://doi.org/10.1515/rela-2017-0016>
- Havron, N., & Arnon, I. (2017). Reading between the words: The effect of literacy on second language lexical segmentation. *Applied Psycholinguistics*, 38(1), 127–153. <https://doi.org/10.1017/S0142716416000138>
- Huensch, A. (2024). Clarifying the role of inhibitory control in L2 phonological processing: A preregistered, close replication of Darcy et al. (2016). *Studies in Second Language Acquisition*, 1–21. <https://doi.org/10.1017/S0272263124000238>
- Isager, P. M., Van Aert, R. C. M., Bahnik, Š., Brandt, M. J., DeSoto, K. A., Giner-Sorolla, ... & Lakens, D. (2023). Deciding what to replicate: A decision model for replication study selection under resource and knowledge constraints. *Psychological Methods*, 28(2), 438–451. <https://doi.org/10.1037/met0000438>
- Jeon, E. H., & Yamashita, J. (2022). L2 reading comprehension and its correlates: An updated meta-analysis. In E. H. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (Vol. 13, pp. 29–86). John Benjamins. <https://doi.org/10.1075/bpa.13.03jeo>
- Kahn, K. B. (2018). Understanding innovation. *Business Horizons*, 61(3), 453–460. <https://doi.org/10.1016/j.bushor.2018.01.011>

- Kanwit, M., & Geeslin, K. L. (2014). The interpretation of Spanish subjunctive and indicative forms in adverbial clauses. A cross-sectional study. *Studies in Second Language Acquisition*, 36(3), 487–533. <https://doi.org/10.1017/S0272263114000126>
- King, K. A., & Mackey, A. (2016). Research methodology in second language studies: Trends, concerns, and new directions. *The Modern Language Journal*, 100(S1), 209–227. <https://doi.org/10.1111/modl.12309>
- Language Teaching Review Panel. (2008). Replication studies in language learning and teaching: Questions and answers. *Language Teaching*, 41(1), 1–14. <https://doi.org/10.1017/S0261444807004727>
- Lantolf, J. P., Poehner, M. E., and Swain, M. (Eds.). (2018). *The Routledge handbook of sociocultural theory and second language development*. Routledge.
- LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, 1(3), 389–402. <https://doi.org/10.1177/2515245918787489>
- Ledgerwood, A. (2014). Introduction to the special section on moving toward a cumulative science: Maximizing what our research can tell us. *Perspectives on Psychological Science*, 9(6), 610–611. <https://doi.org/10.1177/1745691614553989>
- Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6), 1093–1096. <https://doi.org/10.1016/j.jesp.2010.05.025>
- Liu, Y., & Lu, X. (2024). Development of verb argument constructions in L2 English learners: A close replication of research question 3 in Römer and Berger (2019). *Studies in Second Language Acquisition*, 1–19. <https://doi.org/10.1017/S027226312400024X>
- Long, M. H. (1993). Assessment strategies for second language acquisition theories. *Applied Linguistics*, 14(3), 225–249. <https://doi.org/10.1093/applin/14.3.225>
- Mackey, A. (2012). Why (or why not), when, and how to replicate research. In G. Porte (Ed.), *Replication research in applied linguistics*. Cambridge University Press.
- Mackey, A., & Gass, S. M. (2022). *Second language research: Methodology and design* (Third edition). Routledge.
- Malone, J. (2018). Incidental vocabulary learning in SLA. Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*, 40(3), 651–675. <https://doi.org/10.1017/S0272263117000341>
- Marsden, E., Mackey, A., & Plonsky, L. (2016). The IRIS Repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS Repository of Instruments for Research into Second Languages* (pp. 1–21). Routledge.
- Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning*, 68(2), 321–391. <https://doi.org/10.1111/lang.12286>
- Martin, G. N., & Clarke, R. M. (2017). Are psychology journals anti-replication? A snapshot of editorial practices. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.00523>
- McManus, K. (2022a). Are replication studies infrequent because of negative attitudes? Insights from a survey of attitudes and practices in second language research. *Studies in Second Language Acquisition*, 44(5), 1410–1423. <https://doi.org/10.1017/S0272263121000838>
- McManus, K. (2022b). Replication research in instructed SLA. In L. Gurzynski-Weiss & Y. Kim (Eds.), *Instructed second language acquisition research methods* (pp. 103–122). John Benjamins.
- McManus, K. (2023). How and why to conduct a replication study. In A. Mackey & S. M. Gass (Eds.), *Current Approaches in Second Language Acquisition Research* (pp. 334–351). Wiley.
- McManus, K. (2024a). The future of replication in applied linguistics: Toward a standard for replication studies. *Annual Review of Applied Linguistics*, 1–17. <https://doi.org/10.1017/S0267190524000011>
- McManus, K. (Ed.). (2024b). *Usage in second language acquisition. Critical reflections for theory and research*. Routledge.
- McManus, K., & Marsden, E. (2018). Online and offline effects of L1 practice in L2 grammar learning: A partial replication. *Studies in Second Language Acquisition*, 40(2), 459–475. <https://doi.org/10.1017/S0272263117000171>
- Morgan-Short, K., Marsden, E., Heil, J., Issa Ii, B. I., Leow, R. P., Mikhaylova, A., ... & Szudarski, P. (2018). Multisite replication in second language acquisition research: Attention to form during listening and

- reading comprehension. *Language Learning: A Journal of Research in Language Studies*, 68(2), 392–437. <https://doi.org/10.1111/lang.12292>
- National Academies of Sciences, Engineering, and Medicine (with National Academies of Sciences, Engineering, and Medicine). (2019). *Reproducibility and replicability in science: A consensus study report*. The National Academies Press.
- Nind, M., Wiles, R., Bengry-Howell, A., & Crow, G. (2013). Methodological innovation and research ethics: Forces in tension or forces in harmony? *Qualitative Research*, 13(6), 650–667. <https://doi.org/10.1177/1468794112455042>
- Norouzian, R., de Miranda, M., & Plonsky, L. (2018). The Bayesian revolution in second language research: An applied approach. *Language Learning*, 68(4), 1032–1075.
- Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLOS Biology*, 18(3), e3000691. <https://doi.org/10.1371/journal.pbio.3000691>
- Nosek, B. A., & Lakens, D. (2013). Special issue of *Social Psychology* on “Replications of important results in social psychology.” *Social Psychology*, 44(1), 59–60. <https://doi.org/10.1027/1864-9335/a000143>
- Papi, M., Bondarenko, A. V., Mansouri, S., Feng, L., & Jiang, C. (2019). Rethinking L2 motivation research: The 2 × 2 model of self-guides. *Studies in Second Language Acquisition*, 41(2), 337–361. <https://doi.org/10.1017/S0272263118000153>
- Papi, M., & Khajavy, G. H. (2021). Motivational mechanisms underlying second language achievement: A regulatory focus perspective. *Language Learning*, 71(2), 537–572. <https://doi.org/10.1111/lang.12443>
- Parrish, K. (2024). Statistical Insignificance is not wholesale transfer in L3 Acquisition: An approximate replication of Rothman (2011). *Studies in Second Language Acquisition*, 1–18. <https://doi.org/10.1017/S0272263124000342>
- Pelzl, E., Jonczyk, R., & van Hell, J. G. (2024). Second language knowledge can influence native language performance in exclusively native contexts: An approximate replication of Van Hell & Dijkstra (2002). *Studies in Second Language Acquisition*, 1–18. <https://doi.org/10.1017/S0272263124000457>
- Peterson, D., & Panofsky, A. (2021). Self-correction in science: The diagnostic and integrative motives for replication. *Social Studies of Science*, 51(4), 583–605. <https://doi.org/10.1177/03063127211005551>
- Plonsky, L. (2012). Replication, meta-analysis, and generalizability. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 116–132). Cambridge University Press.
- Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, 35(4), 655–687. <https://doi.org/10.1017/S0272263113000399>
- Plonsky, L. (2024). Study quality as an intellectual and ethical imperative: A proposed framework. *Annual Review of Applied Linguistics*, 1–15. <https://doi.org/10.1017/S0267190524000059>
- Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist*, 56(2), 90–100. <https://doi.org/10.1080/00461520.2021.1895796>
- Polio, C. (2012). Replication in published applied linguistics research. A historical perspective. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 47–91). Cambridge University Press.
- Polio, C., & Gass, S. (1997). Replication and reporting: A commentary. *Studies in Second Language Acquisition*, 19(4), 499–508.
- Porte, G. (Ed.). (2012). *Replication research in applied linguistics*. Cambridge University Press.
- Porte, G. (2013). Who needs replication? *CALICO Journal*, 30(1), 10–n/a. <https://doi.org/10.11139/cj.30.1.10-15>
- Porte, G., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.
- Riazi, A. M., & Farsani, M. A. (2023). Mixed-methods research in applied linguistics: Charting the progress through the second decade of the twenty-first century. *Language Teaching*, 1–40. <https://doi.org/10.1017/S0261444823000332>
- Römer, U., & Berger, C. M. (2019). Observing the emergence of constructional knowledge. Verb patterns in German and Spanish learners of English at different proficiency levels. *Studies in Second Language Acquisition*, 41(5), 1089–1110. <https://doi.org/10.1017/S0272263119000202>
- Rose, H., McKinley, J., & Baffoe-Djan, J. B. (2020). *Data collection research methods in applied linguistics*. Bloomsbury Academic.
- Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research*, 27(1), 107–127. <https://doi.org/10.1177/0267658310386439>

- Ryan, J., Foster, P., Wang, Y., Fester, A., & Yap, J. R. (2024). Task design, L1 literacy, and second language oracy: A close replication of Tavakoli and Foster (2008). *Studies in Second Language Acquisition*, 1–18. <https://doi.org/10.1017/S0272263124000445>
- Santos, T. (1989). Replication in applied linguistics research. *TESOL Quarterly*, 23(4), 699–702. ERIC.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90–100. <https://doi.org/10.1037/a0015108>
- Shibayama, S., & Wang, J. (2020). Measuring originality in science. *Scientometrics*, 122(1), 409–427. <https://doi.org/10.1007/s11192-019-03263-0>
- Shook, A., & Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism: Language and Cognition*, 16(2), 304–324. <https://doi.org/10.1017/S1366728912000466>
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9(1), 59–71. <https://doi.org/10.1177/1745691613514450>
- Sudina, E. (2021). Study and scale quality in second language survey research, 2009–2019: The case of anxiety and motivation. *Language Learning*, 71(4), 1149–1193. <https://doi.org/10.1111/lang.12468>
- Tarone, E., Swain, M., & Fathman, A. (1976). Some limitations to the classroom applications of current second language acquisition research. *TESOL Quarterly*, 10(1), 19. <https://doi.org/10.2307/3585936>
- Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 58(2), 439–473. <https://doi.org/10.1111/j.1467-9922.2008.00446.x>
- The TwiLex Group. (2024). First language effects on incidental vocabulary learning through bimodal input: A multisite, preregistered, and close replication of Malone (2018). *Studies in Second Language Acquisition*, 1–26. <https://doi.org/10.1017/S0272263124000275>
- Tschichold, C. (2023). Replication in CALL. *ReCALL*, 35(2), 139–142. <https://doi.org/10.1017/S0958344023000083>
- Valdman, A. (1993). Replication study. *Studies in Second Language Acquisition*, 15, 505.
- Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, 9(4), 780–789. <https://doi.org/10.3758/BF03196335>
- Vu, D. V., & Peters, E. (2022). Learning vocabulary from reading-only, reading-while-listening, and reading with textual input enhancement: Insights from Vietnamese EFL learners. *RELC Journal*, 53(1), 85–100. <https://doi.org/10.1177/0033688220911485>
- Wiles, R., Crow, G., & Pain, H. (2011). Innovation in qualitative research methods: A narrative review. *Qualitative Research*, 11(5), 587–604. <https://doi.org/10.1177/1468794111413227>
- Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, 179, 213–220. <https://doi.org/10.1016/j.cognition.2018.05.008>
- Wang, L., & Sun, X. (2024). Reexamining the role of regulatory focus in second language achievement: An approximate replication of Papi and Khajavy (2021). *Studies in Second Language Acquisition*, 1–22. <https://doi.org/10.1017/S0272263124000512>
- Wong, M. H. I., & Prange, J. (2024). A Bayesian approach to (re)examining learning effects of cognitive linguistics-inspired instruction: A close replication of Wong, Zhao, and MacWhinney (2018). *Studies in Second Language Acquisition*, 1–21. <https://doi.org/10.1017/S0272263124000603>
- Wong, M. H. I., Zhao, H., & MacWhinney, B. (2018). A cognitive linguistics application for second language pedagogy: The English preposition tutor. *Language Learning*, 68(2), 438–468. <https://doi.org/10.1111/lang.12278>
- Yamashita, T., & Neiriz, R. (2024). Why replicate? Systematic review of calls for replication in Language Teaching. *Research Methods in Applied Linguistics*, 3(1), 100091. <https://doi.org/10.1016/j.rmal.2023.100091>
- Zhou, Y., & Papi, M. (2023). The role of future L2 selves in L2 speech development: A longitudinal study in an instructional setting. *System*, 119, 103156. <https://doi.org/10.1016/j.system.2023.103156>
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41, e120. <https://doi.org/10.1017/S0140525X17001972>

Cite this article: McManus, K. (2024). Replication studies in second language acquisition research: Definitions, issues, resources, and future directions: Introduction to the special issue. *Studies in Second Language Acquisition*, 46: 1299–1319. <https://doi.org/10.1017/S0272263124000652>