

Article

Sampling and Generalizability in Lx Research: A Second-Order Synthesis

Luke Plonsky Northern Arizona University, S San Francisco St, Flagstaff, AZ 86011, USA; lukeplonsky@gmail.com

Abstract: As in many other social sciences, second/additional language (Lx) researchers are often interested in generalizing their findings beyond the samples they collect data from. However, very little is known about the range of learner backgrounds and settings found in Lx research. Moreover, the few papers that have addressed the range of settings and demographics sampled in Lx research paint a disappointing picture). The current study examines the extent to which concerns expressed over this issue are merited and worthy of further attention. Toward that end, sample-related features such as L1, Lx/target language, age, proficiency, and educational setting (or lack thereof) were extracted from a sample of 308 systematic reviews of Lx research. The data from this “meta-synthetic” sample are then used to estimate the extent to which Lx research has sampled—and might or might not be able to generalize to—different populations and contexts including those pertinent to migrant populations, the focus of this special issue. The results reveal an incredibly disproportionate interest in participants with English as a first or target language and as well as university students in a narrow range of countries. The findings are used to call out the applied linguistics community on this gross oversight while also seeking to inform future research and contribute to the ongoing methodological reform movement in applied linguistics.

Keywords: sampling; demographics; second language research; research synthesis; meta-synthesis; second-order review; generalizability

**Citation:** Plonsky, Luke. 2023.Sampling and Generalizability in Lx Research: A Second-Order Synthesis. *Languages* 8: 75. <https://doi.org/10.3390/languages8010075>

Academic Editors: Fanny Forsberg Lundell and Suzie Beaulieu

Received: 3 August 2022

Revised: 23 February 2023

Accepted: 1 March 2023

Published: 6 March 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Applied linguistics as a discipline has made significant substantive and methodological progress since its inception in the second half of the 20th century (see historical treatments in Ortega 2009, chap. 1; Pica 1997). However, until recently, relatively little scrutiny has been paid to the state or development of the field’s quantitative methods, an unfortunate circumstance given the empirical rigor needed to reliably and accurately advance second language (L2) theory and practice (see Gass et al. 2021).

The introduction of research synthetic and meta-analytic techniques has raised researchers’ awareness of methodological issues as examined within and across different domains of L2 research. “Methodological syntheses” have addressed topics such as the psychometric properties of our instruments, transparency in data reporting, replication research, the ethics of data handling and reporting, and the appropriacy of statistical techniques being employed, among other issues (e.g., Crowther et al. 2021; Gass et al. 2021; Isbell et al. 2022; Marsden et al. 2018; Plonsky 2013; Sudina 2021). Stemming from the literature on synthetic methods in this and related fields (e.g., Aguinis et al. 2023; Ioannidis et al. 2015; Plonsky and Gass 2011), this line of inquiry systematically examines different facets of study quality and other methodological features as a means to provide empirically grounded recommendations meant to lead to higher validity evidence (see Plonsky 2022).

The present study seeks to further this agenda by considering sampling practices in a subset of L2 research. The focus here is not on statistical power, defined as the probability of a study detecting a statistically significant result given an expected effect and sample size (Cohen 1988), which can affect the internal validity of a study’s findings (see recent

discussion in [Nicklin and Vitta 2021](#)). Rather than consider the *quantity* of samples in this field, the present study focuses on their qualities and, more specifically, the extent to which research findings at the field-wide level might hold beyond the participants and contexts sampled (i.e., generalizability or external validity).

Specifically, this brief report seeks to examine the extent to which L2 research has drawn samples from learners of different ages, educational levels, nationalities, first languages, and so forth as reported in a sample of meta-analyses and other types of synthetic research in applied linguistics.

2. Literature Review

To date, there is very little empirical evidence to support the notion that the field has—or has not—conducted research across the broad range of contexts and learner demographics that it seeks to generalize to worldwide. Taking a similar approach to that of the present study, [Andringa and Godfroid \(2020\)](#) examined several sampling characteristics in a selection of 17 meta-analyses of L2 research. Among other findings, the authors showed that L2 research samples consist overwhelmingly of adults in university contexts. In their own words, “A staggering 88% of all adult samples are university student samples” (p. 138). Staggering indeed.

[Collins and Muñoz \(2016\)](#) coded 97 primary studies of classroom-based “foreign language” learning published in the *Modern Language Journal* in the 21st century for demographics and other features. Almost half (45%) of the studies were carried out in the US, followed by 10% in Japan and 5% in Canada. Also represented were 21 additional countries, mostly in western or central Europe or East Asia. A total of 75% of the samples were comprised of adults, most of whom spoke as their L1 either English (52%) or a small handful of other languages with origins, again, mostly in Europe or East Asia. The target languages were similarly limited (e.g., English = 35%, Spanish = 20%, French = 14%, German = 13%).

[Guo et al. \(2021\)](#) conducted a similar review of target languages as found in the journal *System*. The authors were particularly interested in the extent to which primary studies had focused on languages other than English (LOTE). Their examination of 1974 articles found that only 10.5% had addressed the learning and teaching of LOTEs, most commonly Spanish, Chinese, French, German, Japanese, and Arabic.

The findings of the three reviews described here align well with some of the more anecdotal evidence and associated calls for greater sampling in AL research beyond the so-called WEIRD demographics (i.e., Western, Educated, Industrialized, Rich, Democratic; [Henrich et al. 2010](#)). Indeed, a growing chorus of scholars have criticized the demographic limitations in L2 research, noting especially a lack of research with naturalistic or non-instructed learners, children, heritage learners, and adults that vary in socioeconomic status, educational level, and literacy, and with LOTEs (e.g., [Blum 2017](#); [Cox 2019](#); [Gironzetti and Belpoliti 2018](#); [Mackey and Sachs 2012](#); [Mathews-Aydinli 2008](#); [Ortega 2009, 2019](#); [Plonsky 2017](#); [Vinogradov 2016](#); [Young-Scholten 2013](#)). If verified empirically, the natural consequence of such limited sampling is a lack of generalizability ([Tarone and Bigelow 2007](#)); for pedagogy, the consequence is that practice cannot be accurately informed and thus these populations of learners cannot be best served ([Ortega 2005](#)).

The concern here actually goes beyond generalizability and practicality. A number of widely believed findings may or may not hold up in different settings and with different populations, thus pushing the field to reconsider the theories and models that led to those results. [Huettig and Mishra \(2014\)](#), for example, present compelling evidence that one’s literacy may affect the way they perceive the world and the way they learn in ways that are not limited to the linguistic domain. Such findings are particularly relevant to the positions and findings presented in the current issue, focused on migrant populations. The present study seeks to build on the evidence provided here in previous reviews—both anecdotal and empirical—to help us understand the extent to which different populations and demographics have been examined in Lx research to date.

This issue is not unlike the observation often made about psychological research that oversamples English-speaking university students yet often seeks to generalize to much broader populations (e.g., [Shen et al. 2011](#)). Similar biases toward certain demographics can also be found in biomedical research, where [Oh et al. \(2015\)](#) warn that practitioners “are informed by research extrapolated from a largely homogeneous population, usually white and male” (p. 1).

There is little consolation in knowing that our field is not unique in this respect. Although the vast majority of language learning occurs outside of tertiary institutions in North America, much of the published L2 research appears to be conducted within a small range of institutional settings. Additionally, despite the status of English as likely one of, if not *the* most, commonly learned (additional) language in the world, there are, of course, multitudes of multilingual communities where English is not a main additional language (e.g., [Sridhar 1994](#)). The lack of attention to these contexts and learners, if found, introduces a serious threat to the generalizability of L2 research (see [Bachman 2006](#)). To be sure, a comprehensive theory for language learning and teaching needs to be able to account for learning that takes place in a variety of settings and contexts and with learners of many different backgrounds. This goal cannot be met with such limited sampling in the empirical domain.

The present study seeks to build on the existing evidence of limited sampling practices in applied linguistics. The study is also motivated by the potential ramifications that such restricted sampling may have for theory, research, and for the field’s ability to make a meaningful contribution to society at large ([Ortega 2019](#)). In response to these concerns, a second-order synthesis was conducted to address the following research question: To what extent does applied linguistics research sample from different proficiency levels, ages, educational levels, learning settings, countries, L1s, and target languages?

3. Materials and Methods

3.1. Sample

The sample for the present study consists of 308 reports of synthetic applied linguistics research. The 308 syntheses comprise 15,019 primary studies and 2,519,933 individual participants¹. This sample was drawn from an ongoing collection of synthetic research in applied linguistics. Various forms of synthesis are included in the sample including meta-analyses, scoping reviews, and methodological syntheses (for a breakdown of different types of synthetic research, see [Chong and Plonsky, forthcoming](#)). However, all reports in the sample share core principles of research synthesis such as systematicity both in collecting primary studies as well as in coding them for substantive and/or methodological features of interest ([Plonsky et al., forthcoming](#)). Although the sample covers a wide range of topics within applied linguistics, it may not be representative of the field as a whole due to the fact that some topics and methods are more immediately amenable to synthetic techniques than others. A number of syntheses have been published since the coding was performed on this sample. Therefore, the present study should not be seen as exhaustive but it is believed to be representative of the domains that have been subject to systematic reviews in applied linguistics. It is also worth noting that some of the syntheses addressed a similar scope and included some of the same individual/primary studies, which may thus be slightly over-represented in the current sample.

3.2. Data Collection

A coding scheme was developed following best practices in research synthesis. In addition to basic bibliographic information (e.g., author, publication outlet), the items in the coding scheme pertained to sampling and demographic features reported in the sample of systematic reviews. These included: (a) sample size (number of primary studies and total number of participants), (b) proficiency levels, (c) participant ages, (d) types of educational institutions, (e) the country where the research was conducted, (f) research setting, (g) participants’ L1, and (h) target language.

Extracted from each report in the sample was the percentage of primary studies in their samples that pertained to each of the demographic/sample-related items. As we might expect, the reporting of this information throughout the sample was irregular and inconsistent. All coding was carried out by two doctoral students with extensive training in research synthesis and meta-analysis.

3.3. Analysis

The research question driving this study asked about the extent to which applied linguistics research has sampled across different proficiency levels, ages, educational levels, learning settings, countries, L1s, and target languages. To address this question, the percentages of each feature were simply averaged across all syntheses that had reported them. Only those averages based on three or more syntheses are reported here. In some of the results presented below, the number of secondary studies contributing to the different percentages is relatively small compared to the entire sample of 308. This is because the results are, by nature, limited to what secondary authors reported. Additionally, some syntheses' inclusion criteria necessarily limited their samples to learners of English (e.g., [Hall et al. 2017](#)). In such cases, data pertaining to the percentages of different target languages were excluded from the aggregated results.

4. Results

4.1. Proficiency Levels

Looking across the sample of secondary research, as shown in Figure 1, intermediate learners comprise the largest group at 32%. Almost 23% of the participants in synthesized applied linguistics studies have been labeled as beginners and only 14% as possessing advanced proficiency. An insufficient number of syntheses focused on or reported the presence of heritage learners which were, therefore, not included here.

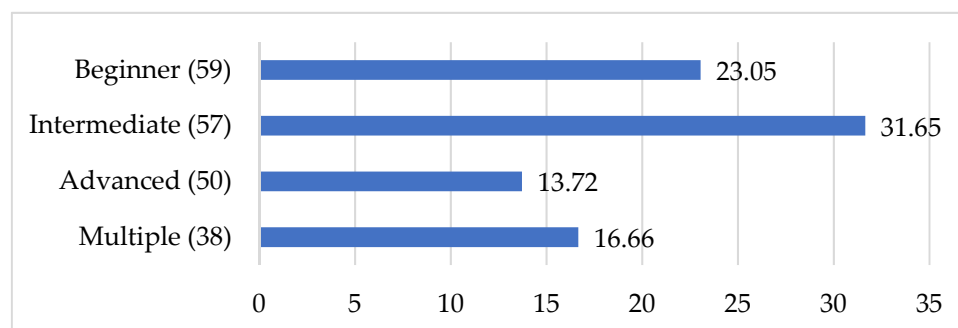


Figure 1. Percentage of participants in different proficiency levels. Note for this and all subsequent figures: the number of syntheses that contributed to each value is found in parentheses next to the data label. For example, “Beginner (59)” here in Figure 1 indicates that 59 syntheses (out of the total 308) reported the percentage of primary studies in their samples that sampled beginner-level participants. These numbers are not equal because syntheses do not uniformly report breakdowns of different primary-sample demographics. As we can see from the results for proficiency, most syntheses in the current sample did not report the percentages of primary studies in their samples across different proficiency levels.

4.2. Age

Another important demographic variable to consider is age. The rate of inclusion of adults (18–55; labeled “Adult-1” in Figure 2) is nearly twice that of children (0–12). Very few secondary studies in the sample included teens in their sample and only 1% focused on older adults (>55; “Adult-2”).

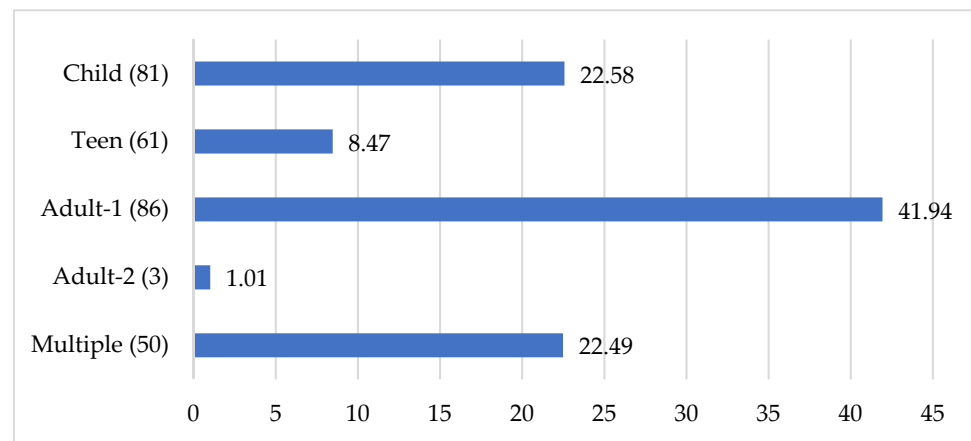


Figure 2. Percentage of participants in different age groups.

4.3. Type of Educational Institution

As displayed in Figure 3, the proportion of samples drawn from university settings is approximately twice that of elementary and middle schools and four times that of high schools. Data collection with participants in language institutes is rare (2%) and even less frequent than samples drawn outside of educational settings (6%).

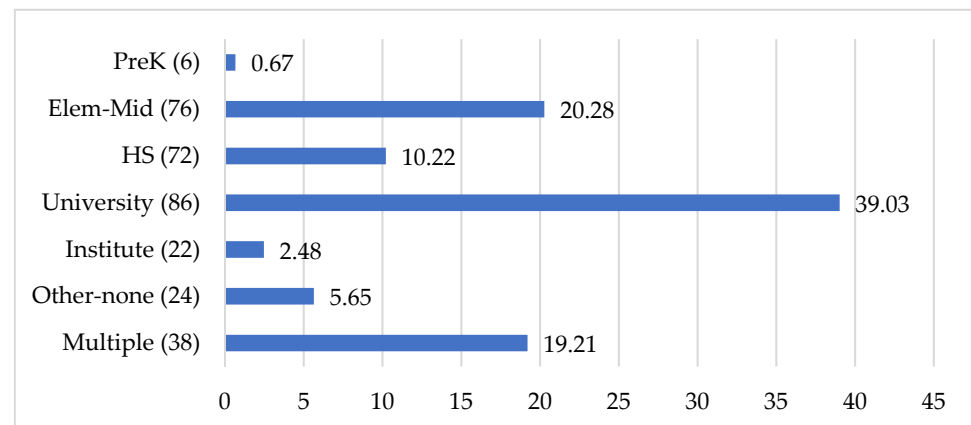


Figure 3. Percentage of participants in different educational contexts.

4.4. Setting

Primary studies included in research syntheses in applied linguistics are carried out fairly evenly in classroom and laboratory settings (see Figure 4). Approximately one in five studies are carried out in both classrooms and labs. Of note as well is the fact that no syntheses in the sample separately reported the percentage of primary studies that focused on migrant populations.

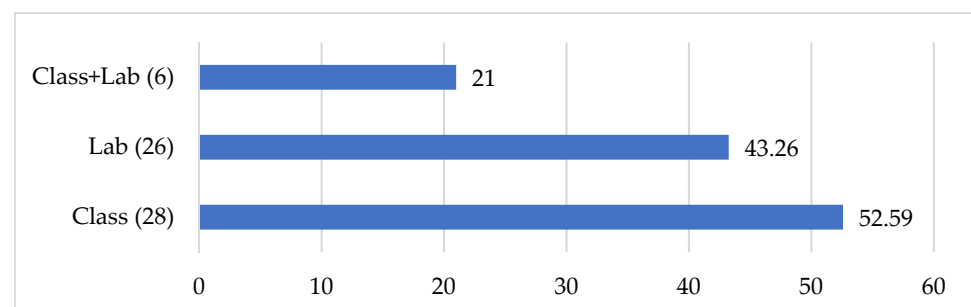


Figure 4. Percentage of participants in different research settings.

4.5. Country Where Research Was Conducted

The present sample of syntheses shows that primary applied linguistics research is carried out in many different regions (see Figure 5). There are a number of particularly strong concentrations that are worth noting, however. These ‘hubs’ include Central and Western Europe (e.g., Spain), the Middle East (Saudi Arabia, Iran), and East Asia (e.g., Japan, Taiwan). The US is arguably an outlier here and even more so if we consider it along with Canada as comprising a single North American region. Together they represent the setting for approximately 60% of all primary studies in the sample of research syntheses.

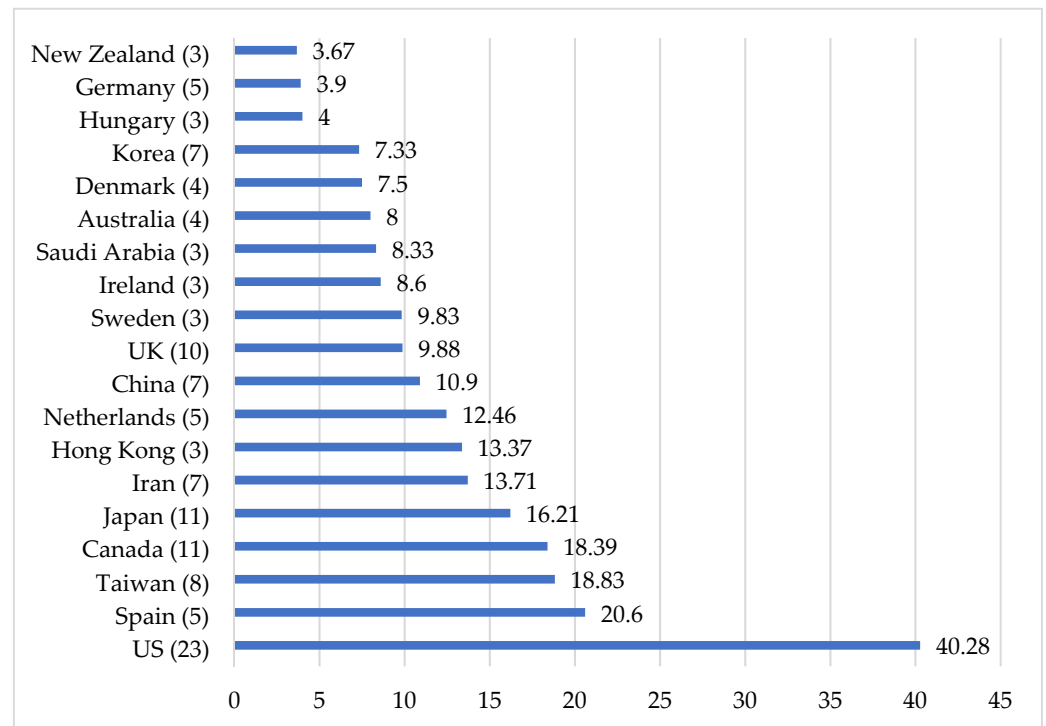


Figure 5. Percentage of primary studies conducted in different countries.

4.6. First Language

A range of first languages is represented in the primary studies that made up the current sample of research syntheses (see Figure 6). Somewhat reflective of the findings for the countries where applied linguistics research has been carried out, there is a substantial presence of Asian and European languages and especially English (>40%). There is little to no evidence of sampling among L1 speakers of African, Indigenous/First Peoples, or Aboriginal languages, for example.

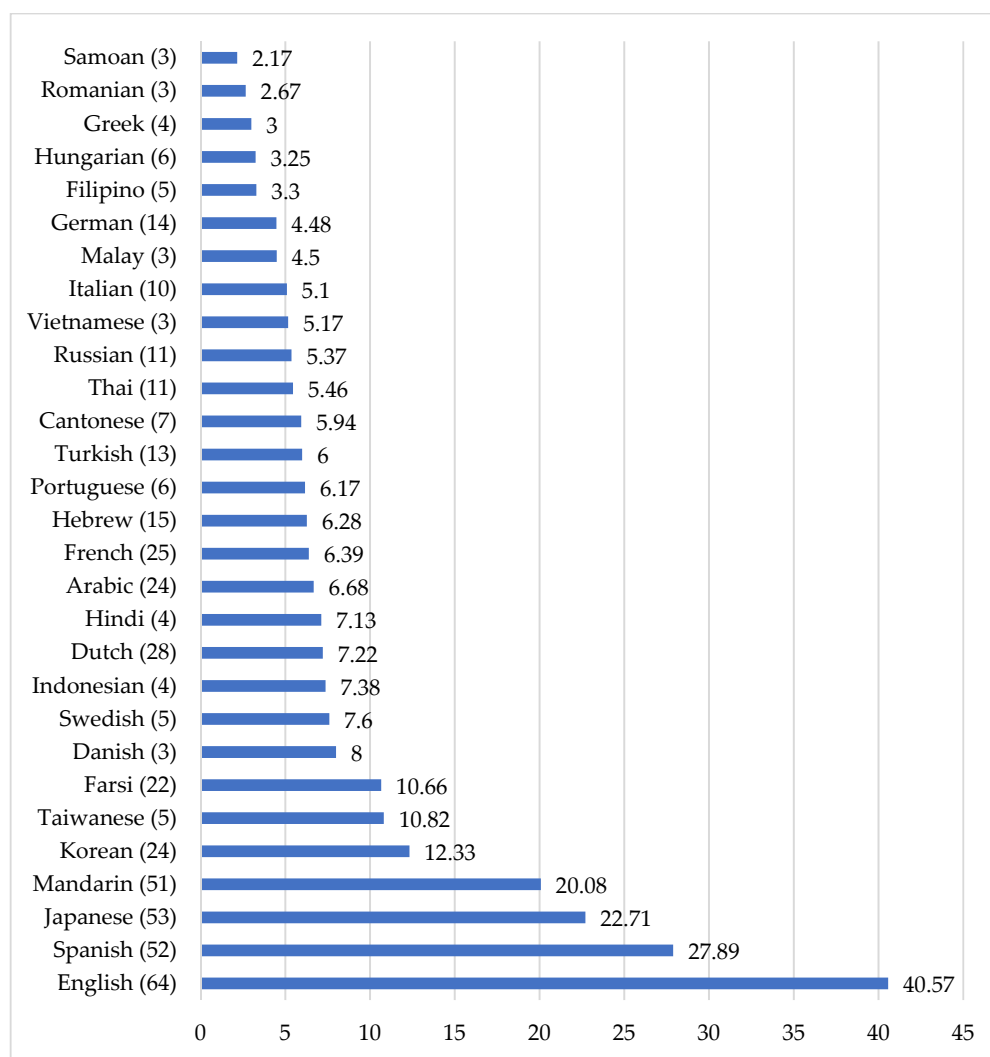


Figure 6. Percentage of first languages across the sample.

4.7. Target Language

The final category that was coded for in this study is the target language of the participants in the primary studies. The range of target languages is actually somewhat smaller than that of the first languages (see Figure 7). Furthermore, the dominance of English as the most commonly studied target language in applied linguistics research is even more pronounced.

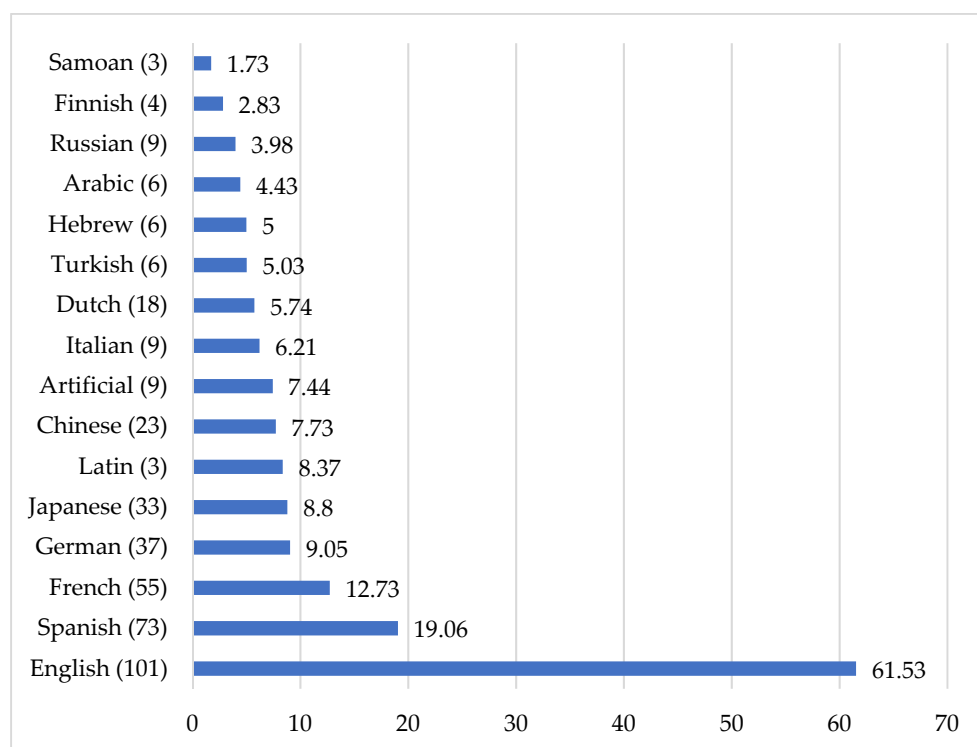


Figure 7. Percentage of target languages across the sample.

5. Discussion

The present study was carried out in response to concerns expressed over the years—and a growing body of empirical evidence—regarding the potentially limited sampling practices within applied linguistics (e.g., [Andringa and Godfroid 2020](#); [Bigelow and Watson 2012](#); [Mackey and Sachs 2012](#); [Ortega 2005](#); [Plonsky 2017](#)). More specifically, a number of scholars have warned that we may not be sampling in a manner that allows us to serve and generalize findings to the diverse populations of interest to the field. With these concerns in mind, a second-order synthesis was conducted (see [Plonsky and Ziegler 2016](#)) in which a sample of 308 syntheses was coded for a set of sample-related and demographic variables found in their samples of primary studies such as age, research setting, proficiency, L1, and target language.

The results presented above illustrate a fairly clear profile of the typical participant in applied linguistics who is a younger adult with intermediate² proficiency in a classroom-based study of English at a North American university. This demographic is certainly important and can contribute to many very useful insights into the nature of language development, language teaching, language usage, language assessment, and so forth. At the same time, the results also confirm what many have long since suspected about severely limited sampling practices found in applied linguistics. They also confirm to some extent [DeKeyser et al.’s \(2010\)](#) claim that “... almost every sample has been one of convenience” (p. 416).

Our colleagues in the psychological sciences are often criticized for relying very heavily on participant pools comprised of undergraduate university students who differ in many respects (e.g., cognitively, socially, educationally, racially, and economically) from the population of adult humans that psychologists often intend to generalize to. (A set of statistical techniques have even been developed by psychologists to account for the attenuation in observed effects that occurs when, as is so common in psychometric studies including in applied linguistics, researchers are working with a restricted range of values in a given scale (see [McKay and Plonsky 2021](#)).) We are no better and we have no empirical, intellectual, or moral high ground to stand on.

The problem here, of course, goes well beyond lacking a moral or intellectual high ground compared to our colleagues in other social sciences. The consequences of such limited samples are felt at both theoretical and practical levels. Theoretically, a lack of research involving a more diverse range of demographics and populations necessarily yields a restricted base of knowledge. We as applied linguists are interested in advancing our understanding of all language (learning, teaching, usage, etc.) among humans, not solely among the severely truncated populations we tend to sample from. From a practical standpoint, we are simply unable to speak to and inform the many matters of an applied nature that pertain to understudied contexts and demographics when our research does not sample them. As [Ortega \(2019\)](#) put it, “If the goal of linguistic–cognitive SLA communities is to contribute to knowledge building about the human capacity for language—a goal that rests on ideas of post-positivist logic, quantitative rigor, and the golden rule of generalizability—then clearly part of the disciplinary mandate is to produce knowledge about all types, all shades and grades, of multilingualism” (p. 34). In my experience as a researcher, editor, reviewer, teacher trainer, and so forth, it is this brand of applied linguistics that Ortega refers to, however, that is perhaps least likely to seek out samples in less studied settings.

The findings of this study demonstrate that the deficits in our knowledge are particularly grave among children, older adults, non-institutional/classroom learners, and individuals with L1s and target languages other than those typically spoken and studied outside a narrow range of countries generally found in North America, Europe, and East Asia. If anyone was looking for yet another instance of systematic bias and disfavor toward already-marginalized or indigenous populations or the global south, look no further (see [Pennycook and Makoni 2020](#)). All this is to say nothing of the many forms of bias and racism experienced by scholars in less privileged settings in applied linguistics. That we as a field do not sample in a wider range of contexts and with a wider range of populations, though clearly detrimental, is perhaps too simple and comfortable of an explanation. Whether for financial, practical, and/or cultural reasons, researchers in applied linguistics tend to focus on populations in their vicinity (e.g., [Grenoble and Osipov 2023](#); [Lloyd-Smith and Kupisch 2023](#)). Although outside the scope of the present study, some might argue that researchers’ preferences also reflect epistemological racism and a bias in favor of white Eurocentric hegemonic knowledge ([Flores and Rosa, forthcoming](#); [Kubota 2020](#)).

Recommendations for Future Research

The findings of the present study carry with them several immediate implications for future research that include but also go beyond the need for broader sampling. Yes, of course, sampling in a wider range of settings and with a wider range of demographics and languages is needed. The need is perhaps most acute among non-instructed learners, non-literate and low-literate learners, children, older adults, learners of very low and very high proficiency, and L1–TL combinations that do not include English.

One perhaps-obvious step to take, which aligns with ongoing movements in the field toward open science and synthetic-mindedness ([Marsden and Plonsky 2018](#); [Norris and Ortega 2006](#)), is to engage in replication research with samples that differ from that of the initial study. I hasten to add at this point, though, that simply altering the participant demographic is not in and of itself necessarily sufficient justification for a replication (see [Porte and McManus 2019](#)). One logistically attractive approach to broaden our sampling may come through samples obtained virtually (i.e., through the internet; see relevant discussion and examples in [Shepperd 2022](#)). However, even with the use of such technologies, samples that are comparable (e.g., literacy-wise) may not be possible. In such cases, researchers might aim for conceptual (not “close”) replication by employing alternative measures and procedures (see [Ranta and Zavialova 2022](#), this issue, for further recommendations and considerations).

There is also an urgent need for research that is informed by and that subsequently seeks to inform on the practical issues pertinent to marginalized settings, countries, and

regions. To state the obvious, working in such contexts can introduce logistical, cultural, administrative, and technological hurdles. As I see it, however, this work is our ethical duty and responsibility to each other, to the field, and to the communities who stand to benefit from participating in such research. This is not a call for researchers to ride in on the proverbial white horse of (our own brand of) empiricism. This is a call to develop sincere and hopefully, long-lasting professional relationships wherein all parties may benefit intellectually, practically, and professionally. (As an example, see Rhonda Oliver's work over the last 20+ years in Australian Aboriginal communities such as [Oliver and Forrest 2021](#), and [Shay et al. 2022](#); in the context of migrant language learners, see work by Bigelow, Tarone, and colleagues.) The next steps needed for the field are not as simple as engaging with such populations in research, however. There may also be a need for the field to consider alternative or novel theoretical models and to consider a range of logistical, cultural, and/or ethical issues (see further guidance on these matters in Hamel-Michaud et al. this issue).

At a more conceptual level, this is also a call for epistemological open-mindedness. If we want to move away from the structural biases that are baked into the paths by which knowledge advances in applied linguistics, we have to put aside some of the conventions we have been trained to adhere to regarding the 'right' ways of knowing. I personally do not know how to do this or exactly what this looks like, but I want to learn. One point of departure on the path toward epistemological open-mindedness might be to simply listen by, for example, attending talks by scholars from traditions that differ from our own without assuming their standards for knowledge construction will be the same as ours. When doing so, it can be helpful to then question our own ways of arriving at new knowledge. A third step toward epistemological open-mindedness would be to collaborate with scholars from a tradition other than our own and, in doing so, to try to embrace their approach.

One potential threat to the validity of the present study concerns whether the syntheses in my sample conducted adequately comprehensive searches. In other words, it may be that primary applied linguistics research does actually sample from a wider range of settings and demographics but that the synthesists that carried out the studies in my sample did not search enough to locate those studies. [Plonsky and Brown \(2015\)](#), in fact, found the searches in L2 meta-analyses to be far from comprehensive. Whether or not this is the case, I urge synthesists to engage in comprehensive searches and, in particular, to seek to identify primary studies that are unpublished or that are not necessarily published in journals hosted by large, western-based, revenue-seeking publishers.

6. Conclusions

It would be fairly obvious to mention to a colleague the overreliance in applied linguistics on university learners, on WEIRD samples, and on younger so-called college-aged adults. Nevertheless, this study allows us to go beyond such casual—if also largely accurate—observations, to gauge the extent of the problem at hand, which is quite substantial. It is a problem not only on both theoretical and practical grounds; it is also an ethical beach in my view that we as applied linguists have failed so grossly at the field-wide level to contribute to the vast majority of the language-related settings and populations in the world. The near-exclusive representation in mainstream journals both of learners in wealthy countries and of L1 or TL users of English is particularly concerning. However, we can improve and expand. We will improve, I believe, and I hope that the evidence and recommendations provided here take us one step closer to doing so.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data from this study will be made available upon request on the IRIS database.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- ¹ Not all syntheses reported this information. A total of 293 and 161 of the syntheses reported the number of primary studies and individual participants in their samples, respectively. The true number of samples and participants may actually be significantly higher than what is reported here. However, also worth noting is the potential overlap among the samples of primary studies included in these syntheses which would reduce the total number of unique primary studies and individual participants.
- ² In considering findings related to proficiency, it must be recognized that proficiency is notoriously difficult to define and measure. It almost goes without saying that the labeling and descriptions that led to these percentages varied substantially across authors at both the primary and secondary levels. For reviews of operationalizations and reporting of proficiency-related information in primary research, see [Thomas \(2006\)](#) and [Park et al. \(2022\)](#).

References

- Aguinis, Herman, Ravi S. Ramani, and Nawaf Alabduljader. 2023. Best-practice recommendations for producers, evaluators, and users of methodological literature reviews. *Organizational Research Methods* 26: 46–76. [\[CrossRef\]](#)
- Andringa, Sible, and Aline Godfroid. 2020. Sampling bias and the problem of generalizability in applied linguistics. *Annual Review of Applied Linguistics* 40: 134–42. [\[CrossRef\]](#)
- Bachman, Lyle. 2006. Generalizability: A journey into the nature of empirical research in applied linguistics. In *Inference and Generalizability in Applied Linguistics: Multiple Perspectives*. Edited by Micheline Chalhoub-Deville, Carol Ann Chapelle and Patsy A. Duff. Amsterdam: John Benjamins, pp. 165–207.
- Bigelow, Martha, and Jill Watson. 2012. The role of educational level, literacy, and orality in L2 learning. In *Handbook of Second Language Acquisition*. Edited by Susan Gass and Alison Mackey. London: Routledge, pp. 461–75.
- Blum, Susan D. 2017. Unseen WEIRD Assumptions: The so-called language gap discourse and ideologies of language, childhood, and learning. *International Multilingual Research Journal* 11: 23–38. [\[CrossRef\]](#)
- Chong, Sin Wang, and Luke Plonsky. forthcoming. A typology of secondary research in applied linguistics. *Applied Linguistics Review*.
- Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Mahwah: Lawrence Erlbaum Associates.
- Collins, Laura, and Carmen Muñoz. 2016. The foreign language classroom: Current perspectives and future considerations. *Modern Language Journal* 100: 133–47. [\[CrossRef\]](#)
- Cox, Jessica G. 2019. Multilingualism in older age: A research agenda from the cognitive perspective. *Language Teaching* 52: 360–73. [\[CrossRef\]](#)
- Crowther, Dustin, Susie Kim, Jonogbong Lee, Jungmin Lim, and Shawn Loewen. 2021. Methodological synthesis of cluster analysis in second language research. *Language Learning* 71: 99–130. [\[CrossRef\]](#)
- DeKeyser, Robert, Iris Alfa-Shabtay, and Dorit Ravid. 2010. Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics* 31: 413–38. [\[CrossRef\]](#)
- Flores, Nelson, and Jonathan Rosa. forthcoming. Undoing competence: Coloniality, homogeneity, and the overrepresentation of whiteness in applied linguistics. *Language Learning*. [\[CrossRef\]](#)
- Gass, Susan, Shawn Loewen, and Luke Plonsky. 2021. Coming of age: The past, present, and future of quantitative SLA research. *Language Teaching* 54: 245–58. [\[CrossRef\]](#)
- Gironzetti, Elisa, and Flavia Belpoliti. 2018. Investigación y pedagogía en la enseñanza del español como lengua de herencia (ELH): Una metasíntesis cualitativa. *Journal of Spanish Language Teaching* 5: 16–34. d.
- Grenoble, Lenore, and Boris Osipov. 2023. The dynamics of bilingualism in language shift ecologies: Bilingualism and language shift. *Linguistic Approaches to Bilingualism* 13: 1–39. [\[CrossRef\]](#)
- Guo, Qianjiang, Xiaozhou Zhou, and Xuesong Gao. 2021. Research on learning and teaching of languages other than English in *System*. *System* 100: 102541. [\[CrossRef\]](#)
- Hall, Colby, Garret J. Roberts, Eunsoo Cho, Lisa V. McCulley, Megan Carroll, and Sharon Vaughn. 2017. Reading instruction for English learners in the middle grades: A meta-analysis. *Educational Psychology Review* 29: 763–94. [\[CrossRef\]](#)
- Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33: 61–83. [\[CrossRef\]](#) [\[PubMed\]](#)
- Huettig, Falk, and Ramesh K. Mishra. 2014. How literacy acquisition affects the illiterate mind—A critical examination of theories and evidence. *Language and Linguistics Compass* 8: 401–27. [\[CrossRef\]](#)
- Ioannidis, John S. P., Daniele Fanelli, Debbie D. Dunne, and Steven N. Goodman. 2015. Meta-research: Evaluation and improvement of research methods and practices. *PLoS Biology* 13: e1002264. [\[CrossRef\]](#)
- Isbell, Daniel, Daniel Brown, Meishan Chen, Deirdre Derrick, Romy Ghanem, Maria Nelly Gutiérrez Arvizu, Erin Schnur, Meixiu Zhang, and Luke Plonsky. 2022. Misconduct and questionable research practices: The ethics of quantitative data handling and reporting in applied linguistics. *Modern Language Journal* 106: 172–95. [\[CrossRef\]](#)

- Kubota, Ryuko. 2020. Confronting epistemological racism, decolonizing scholarly knowledge: Race and gender in applied linguistics. *Applied Linguistics* 41: 712–32.
- Lloyd-Smith, Anika, and Tanja Kupisch. 2023. Methodological challenges in working with Indigenous communities. *Linguistic Approaches to Bilingualism* 13: 65–69. [\[CrossRef\]](#)
- Mackey, Alison, and Rebecca Sachs. 2012. Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning* 62: 704–40. [\[CrossRef\]](#)
- Marsden, Emma, and Luke Plonsky. 2018. Data, open science, and methodological reform in second language acquisition research. In *Critical Reflections on Data in Second Language Acquisition*. Edited by Aarnes Gudmestad and Amanda Edmonds. Amsterdam: John Benjamins, pp. 219–28.
- Marsden, Emma, Kara Morgan-Short, Sophie Thompson, and David Abugaber. 2018. Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning* 68: 321–91. [\[CrossRef\]](#)
- Mathews-Aydinli, Julie. 2008. Overlooked and understudied? A survey of current trends in research on adult English language learners. *Adult Education Quarterly* 58: 198–213. [\[CrossRef\]](#)
- McKay, Todd, and Luke Plonsky. 2021. Reliability analyses: Estimating error in L2 research. In *The Routledge Handbook of Second Language Acquisition and Language Testing*. Edited by Paula Winke and Tineke Brunfaut. New York: Routledge, pp. 468–82.
- Nicklin, Christopher, and Joseph P. Vitta. 2021. Effect-driven sample sizes in second language instructed vocabulary acquisition research. *Modern Language Journal* 105: 218–36. [\[CrossRef\]](#)
- Norris, John. M., and Lourdes Ortega, eds. 2006. The value and practice of research synthesis for language learning and teaching. In *Synthesizing Research on Language Learning and Teaching*. Amsterdam: John Benjamins, pp. 3–50.
- Oh, Sam S., Joshua Galanter, Neeta Thakur, Maria Pino-Yanes, Nicolas E. Barcelo, Marquitta J. White, Danielle M. de Bruin, Ruth M. Greenblatt, Kirsten Bibbins-Domingo, Alan H. B. Wu, and et al. 2015. Diversity in clinical and biomedical research: A promise yet to be fulfilled. *PLoS Medicine* 12: e1001918. [\[CrossRef\]](#)
- Oliver, Rhonda, and Simon Forrest. 2021. Supporting the diverse language background of aboriginal and torres strait Islander students. In *Indigenous Education in Australia: Learning and Teaching for Deadly Futures*. Edited by Marnee Shay and Rhonda Oliver. London: Routledge, pp. 97–111.
- Ortega, Lourdes. 2005. For what and for whom is our research? The ethical as transformative lens in instructed SLA. *Modern Language Journal* 89: 427–43. [\[CrossRef\]](#)
- Ortega, Lourdes. 2009. *Understanding Second Language Acquisition*. London: Routledge.
- Ortega, Lourdes. 2019. SLA and the study of equitable multilingualism. *Modern Language Journal* 103: 23–38. [\[CrossRef\]](#)
- Park, Hae In, Megan Solon, Marzieh Dehghan-Chaleshtori, and Hessameddin Ghanbar. 2022. Proficiency reporting practices in SLA research: Have we made any progress? *Language Learning* 72: 198–236. [\[CrossRef\]](#)
- Pennycook, Alistair, and Sinfrey Makoni. 2020. *Innovations and Challenges in Applied Linguistics from the Global South*. London: Routledge.
- Pica, Teresa. 1997. Second language teaching and research relationships: A North American View. *Language Teaching Research* 1: 48–72. [\[CrossRef\]](#)
- Plonsky, Luke. 2013. Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition* 35: 655–87. [\[CrossRef\]](#)
- Plonsky, Luke. 2017. Quantitative research methods. In *The Routledge Handbook of Instructed Second Language Acquisition*. Edited by Shawn Loewen and Masatoshi Sato. New York: Routledge, pp. 505–21.
- Plonsky, Luke. 2022. Quantitative research methods and the reform movement in applied linguistics. In *Research Questions in Language Education and Applied Linguistics*. Edited by Hassan Mohebbi and Christine Coombe. Berlin and Heidelberg: Springer.
- Plonsky, Luke, and Daniel Brown. 2015. Domain definition and search techniques in meta-analyses of L2 research (Or why 18 meta-analyses of feedback have different results). *Second Language Research* 31: 267–78. [\[CrossRef\]](#)
- Plonsky, Luke, and Nicole Ziegler. 2016. The CALL-SLA interface: Insights from a second-order synthesis. *Language Learning and Technology* 20: 17–37.
- Plonsky, Luke, and Susan Gass. 2011. Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning* 61: 325–66. [\[CrossRef\]](#)
- Plonsky, Luke, Yuhang Hu, Ekaterina Sudina, and Frederick L. Oswald. forthcoming. Advancing meta-analytic methods in L2 research. In *Current Approaches in Second Language Acquisition Research*. Edited by Alison Mackey and Susan Gass. Hoboken: Wiley Blackwell.
- Porte, Graeme, and Kevin McManus. 2019. *Doing Replication Research in Applied Linguistics*. London: Routledge.
- Ranta, Lelia, and Alisa Zavialova. 2022. The challenges of conducting research in diverse classrooms: Reflections on a pragmatics teaching experiment. *Languages* 7: 223. [\[CrossRef\]](#)
- Shay, Marnee, Rhonda Oliver, Helen C. D. McCarthy, Tatiana Bogachenko, and Boori Monty Pryor. 2022. Developing culturally relevant and collaborative research approaches: A case study of working with remote and regional Aboriginal students to prepare them for life beyond school. *The Australian Educational Researcher* 49: 657–74. [\[CrossRef\]](#)
- Shen, Winny, Thomtas B. Kiger, Stacey E. Davies, Rena L. Rasch, Kara M. Simon, and Deniz. S. Ones. 2011. Samples in applied psychology: Over a decade of research in review. *Journal of Applied Psychology* 96: 1055–64. [\[CrossRef\]](#) [\[PubMed\]](#)

- Shepperd, Louise. 2022. Including underrepresented language learners in SLA research: A case study and considerations for internet-based methods. *Research Methods in Applied Linguistics* 1: 100031. [\[CrossRef\]](#)
- Sridhar, Shikaripur N. 1994. A reality check for SLA theories. *TESOL Quarterly* 28: 800–5. [\[CrossRef\]](#)
- Sudina, Ekaterina. 2021. Study and scale quality in second language survey research, 2009–2019: The case of anxiety and motivation. *Language Learning* 71: 1149–93. [\[CrossRef\]](#)
- Tarone, Elaine, and Martha Bigelow. 2007. Alphabetic print literacy and processing of oral corrective feedback in the L2. In *Interaction and Second Language Acquisition*. Edited by Alison Mackey. Oxford: Oxford University Press, pp. 101–21.
- Thomas, Margaret. 2006. Research synthesis and historiography: The case of assessment of second language proficiency. In *Synthesizing Research on Language Learning and Teaching*. Edited by Lourdes Ortega and John M. Norris. Philadelphia: John Benjamins, pp. 279–98.
- Vinogradov, Igor. 2016. Linguistic corpora of understudied languages: Do they make sense? *Káñina* 40: 116–30. [\[CrossRef\]](#)
- Young-Scholten, Martha. 2013. Low-educated immigrants and the social relevance of second language acquisition research. *Second Language Research* 29: 441–54. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.