CAMBRIDGE
UNIVERSITY PRESS

PLENARY SPEECH

Reducing 'avoidable research waste' in applied linguistics research: Insights from healthcare research

Talia Isaacs[1]* (iD) and Hamish Chalmers[2] (iD)

[1]University College London, London, UK and [2]University of Oxford, Oxford, UK
*Corresponding author. Email: talia.isaacs@ucl.ac.uk

Abstract

This paper explores Chalmers and Glasziou's (2009) notion of 'research waste' from healthcare research to examine what it can offer the field of applied linguistics. Drawing on examples from both disciplines, we unpack Macleod et al.'s (2014) five research waste categories: (1) asking the wrong research questions, (2) failing to situate new research in the context of existing research, (3) inefficient research regulation/management, (4) failing to disseminate findings, and (5) poor research reporting practices. We advance this typology to help applied linguists identify and reduce avoidable research waste and improve the relevance, quality, and impact of their research.

1. Introduction

The nature of our scholarly collaborations has allowed both authors the privilege of working across the disciplines of applied linguistics and healthcare research. While applied linguistics is our natural academic home, our collaborations have enabled us to learn from healthcare researchers, exposing us to thinking from this field and providing a vantage point from which to examine and assess the relevance of work in healthcare in relation to our own. One such aspect of that work is avoidable research waste. In 2009, *The Lancet*, which had the highest impact factor of all medical journals in 2022 (Clarivate, 2023), published a highly influential paper exploring the concept of avoidable research waste (Chalmers & Glasziou, 2009). This paper precipitated a series of follow-up articles (Macleod et al., 2014; Robinson et al., 2021), including a taxonomy of avoidable waste sources that, if acknowledged and adequately addressed, would serve to improve healthcare research. In this paper, we unpack the work of Chalmers and Glasziou (2009) and those who took up their theme, and assess its relevance to practically-grounded applied linguistics research. We would note that, in line with *Language Teaching*'s Plenary Speeches section description, readers may find this article 'at times to be **provocative and spontaneous**' (original emphasis). Our hope is to advance the research waste typology as a mechanism for taking stock of where we are and where we need to go as a field to continue to improve the transparency, efficiency, accessibility, relevance, and quality of applied linguistics research and catalyse further discussion.

We recognise that some commentators consider healthcare research and social sciences research to be fundamentally different, such that lessons from one field are not applicable to the other (e.g.,

---

Furedi, 2013; Thomas, 2013, february 3). In particular, criticisms of cross-disciplinary work of the sort we present here tend to focus on the relative complexity of social sciences contexts, arguing that healthcare research is more straightforward, both in conduct and interpretation. We disagree. Depending on the specific nature of the study, the variables at play in healthcare research can involve both psychocognitive and sociocultural/interactional dimensions, just as in social sciences research. For example, in behaviour change research on mask wearing as a population-level intervention to reduce COVID-19 transmission, a sole focus on whether a cloth barrier prevents respiratory aerosols from travelling far enough from the mask-wearer to pose a risk to others – although seeming straight-forward and with objective outcome measures – ignores complex human and contextual factors. This includes variation in people's mask-wearing attitudes within and across contexts, responses to mask-promoting educational interventions, identification of 'spent' masks, their disposal availability and having the economic means to replace them, culturally-mediated attitudes towards authority, and so forth – all of which could have a bearing on outcomes. Our understanding of healthcare research includes manifold examples of research that must account for the complexity of human beings, the places they inhabit, and the variety and competition in what constitutes meaningful outcomes for research conducted with and for them. Although we acknowledge fieldwide differences between health research and education (e.g., genomic differences are more prominent in the former), we maintain the importance of conducting rigorous, well-designed, well-executed, and well-warranted research regard-less of the field. We are also of the opinion that one field can learn from another and expand upon this here through illustrative examples.

From a purely academic perspective, it is important to acknowledge and reduce potential sources of research waste and improve efficiency, such that research positively contributes to our collective under-standing of the world. But, more fundamentally, like readers of this paper, we are indirect funders and consumers of research. Through taxation, our money is used to support research. For example, UK Research and Innovation (UKRI), comprised of seven government-funded research councils to the tune of £6 billion a year (GOV.UK, n.d.), is the largest funder of medical and social sciences research in the UK. As members of society, we all stand to benefit from the results of publicly funded research. We all, therefore, have vested interests in understanding insights relating to research waste from any field and in considering the relevance of the research to our lives. In this paper, we focus on the objective of reducing research waste at different phases of the research process. We relate our understanding of what this means in healthcare research to generate a discussion on what avoidable research waste means for applied linguistics research. Our objective is to influence the ways in which applied linguistics researchers conduct their research in the hope that they will concur that there are measures we should take to maximise its potential.

To clarify, we do not seek to put healthcare research on a pedestal through this contribution nor make any value judgments about the relative quality of healthcare research and applied linguistics research, which would be an unfair comparison. In fact, healthcare researchers themselves have emphasised that, in the years after the initial publication of the research waste typology (Chalmers & Glasziou, 2009), poor practice persists. This is evidenced by Glasziou and Chalmers's (2018) follow-up article, entitled 'Research waste is still a scandal' and is echoed in Pirosca et al.'s (2022) 'Tolerating bad health research: The continuing scandal.' In line with these critiques, we highlight areas where applied linguists could improve efficiencies and avoid wasteful practice. Our overall inten-tion is to examine and ask readers to consider what the research waste typology from healthcare research can offer the field of applied linguistics.

2. Unpacking research waste for applied linguistics

Chalmers and Glasziou (2009) and Macleod et al. (2014) characterise five sources of avoidable research waste in producing and reporting research evidence: (1) relevance of the research questions, (2) neces-sity of the research and appropriateness of its design, (3) efficiency of research regulation and man-agement, (4) extent to which the findings are published, and (5) quality of research reporting.
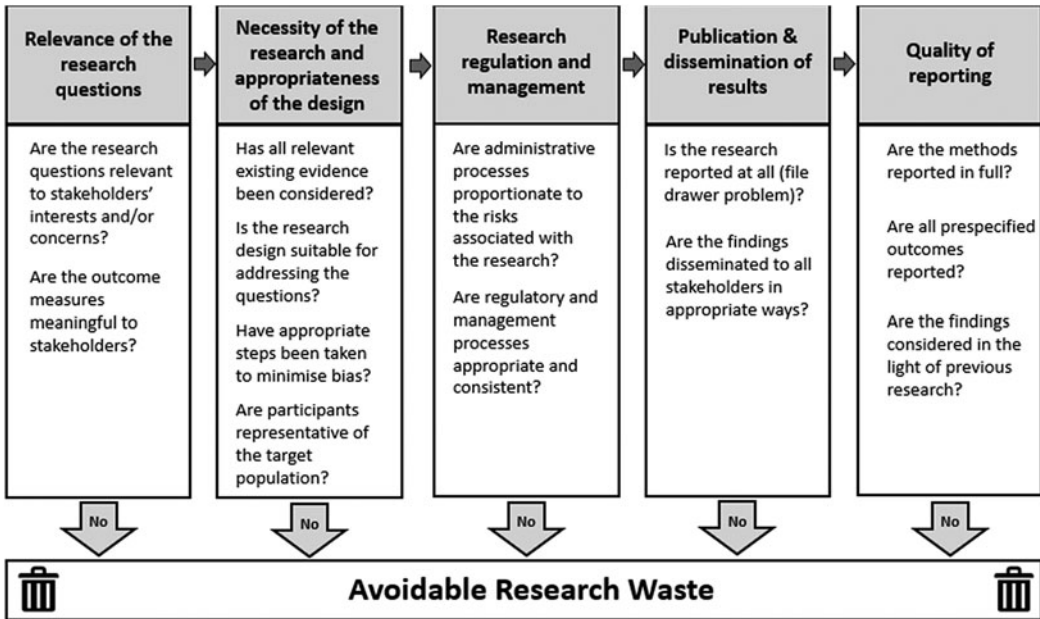
Figure 1. Questions to ask to evaluate potential sources of research waste or explore inefficiencies in applied linguistics research

These waste sources are expanded upon in Figure 1 and adapted from Macleod et al. (2014) for an applied linguistics audience.

2.1 Asking the wrong research questions

The first source of research waste results from 'choosing the wrong questions for research' (Chalmers & Glasziou, 2009, p. 86). There are many ways in which poorly framed or conceptualised research questions and the methods used to address them could be wasteful. In meta-analyses, for example, asking 'to what extent and in what contexts' research questions about the effectiveness of an intervention might be more informative (and therefore less wasteful) than a binary 'yes/no question.' However, Chalmers and Glasziou are concerned with a more fundamental point: the relevance of research questions to people whose lives and practices the research is intended to inform or influence. In healthcare settings, these are patients, clinicians, and policymakers. In educational settings, these are learners, educators, and policymakers. If applied researchers address questions that are of no practical or theoretical relevance to stakeholder groups, they are being wasteful. We start with classroom-based research examples to illustrate recent thinking in applied linguistics and then describe initiatives in healthcare research to draw parallels and extend the points raised.

Chong's (2022) methodological seminar on synthesis methods in applied linguistics – the event at which we presented the thinking represented in this paper – was subtitled, 'Facilitating research-pedagogy dialogue' (p. 142). This subtitle partially captures issues that can lead to the first source of research waste. Several second language (L2) researchers have expressed concern about educational practitioners' limited uptake of research findings. In Sato and Loewen's (2018) article, 'Do teachers care about research? The research-pedagogy dialogue', for example, they note that instructed second language acquisition (SLA) researchers have generated evidence-based findings that teachers should be able to successfully incorporate into their classrooms. However, they express disappointment about the extent of teachers' uptake of research. To address this, they contend that classroom-based SLA researchers should have the intention to share their findings with teachers and that teachers should be open, in principle, to integrating the research results into their classroom practice. Naturally, for

teachers to be open to doing so, that research must be responsive to their needs and interests. As Ellis (2010) observes, 'it is always the teacher who ultimately determines the relevance of SLA constructs and findings' (p. 197).

If researchers fail to investigate what teachers want from research, they risk wasting time, energy, and resources addressing questions about which nobody cares except them. A solution to this is embedding teachers' views into the research process at the start of the research cycle, including deciding what to research. Ellis (2010), Sato and Loewen (2018, 2022), and Sato et al. (2022) propose ways of fostering a researcher–teacher collaborative mindset, as both parties stand to benefit from working together. However, Sato and Loewen (2022) ascribe researchers the overriding responsibility for fostering such a collaboration. McKinley (2019) concurs, calling on applied linguists 'to collaborate with teachers to ensure research questions are driven by practice-based problems' (p. 876). By involving teachers in defining the focus of new research, the ensuing research is more likely to respond to classroom realities, inform practice, and drive the field forward. McKinley cites his recommendations in the context of an action research paradigm, which, by nature, is often specific to an individual teacher or school. But the principle of engaging teachers in articulating what sort of research would be meaningful and useful to them can be applied more broadly.

Related to this, some research traditions in applied linguistics invite teachers not only to give an opinion, but also to share in the conception of the research, identify relevant research questions, and actively participate in the co-production or co-creation of data and the analysis and dissemination of those data. This includes different versions of practitioner and participatory approaches to research, including action research linked to reflective practice, with teachers adopting the dual role of investigator and participant in their own classrooms (see e.g., Burns, 2010; Mann & Walsh, 2017), and Exploratory Practice, which also emphasises a role for learners alongside teachers as co-researchers (e.g., Allwright, 2005; Allwright & Hanks, 2009). These approaches position teachers in the role of research agents rather than simply as consumers or objects of research, driving their own research and educational agendas (Hanks, 2019).

The second author initiated a project aiming to include educational stakeholders in the research enterprise through a priority setting partnership involving educators and parents invested in the education of multilingual learners – a group more commonly referred to in the UK as learners of English as an Additional Language (EAL) (Chalmers et al., 2021). Modelled on a consensus-building approach used in healthcare research to establish patient- and clinician-informed priorities for further research (i.e., Delphi panel, James Lind Alliance, 2022), a steering committee consisting of representatives from the main stakeholder groups was convened to oversee the project. The committee informed the design of a survey eliciting other EAL stakeholders' most pressing unanswered questions about educating EAL learners. A total of 199 respondents (EAL teachers, mainstream or subject teachers, EAL learners, parents, school governors, headteachers, ethnic minority achievement services managers, and bilingual learning assistants) submitted a total of 767 'unanswered questions.' The researchers collapsed questions that addressed similar themes, yielding 81 unique research questions. Stakeholders then ranked these questions in order of priority. Stakeholder group representatives subsequently discussed and debated each of the highest ranked questions in a workshop, working democratically and collaboratively to generate a Top 10 list of research priorities for EAL. The researchers then publicised the resulting list to funders, in the hope that funders would take into account the demonstrated needs of EAL stakeholders when commissioning and evaluating new research.

The value of this approach to informing research priorities and directions and its relationship to reducing avoidable research waste can be seen by comparing Chalmers et al.'s (2021) findings with the results of a similar exercise among 'experts' in the same general field (Duarte et al., 2023). While some respondents to Duarte et al.'s Delphi exercise were reportedly educational practitioners, three-quarters were academics. None of the ranked priorities for new research that these experts identified aligned with Chalmers et al.'s (2021) participants' (mainly teachers and other educators') priorities. While we defend a researcher's right to investigate research problems at the cutting-edge of their field, at least some of their attention could be oriented to addressing issues of known interest to the

intended beneficiaries of their research. The use of Delphi panels to undertake priority-setting exercises such as the one described above, are, as yet, uncommon in applied linguistics research (see Sterling et al., 2023). Our research community could better serve its stakeholders and reduce the risk of engaging in wasteful research if it paid more attention to what stakeholders consider meaningful and useful.

Research for All (2022) is an open access education journal dedicated to social sciences publications that take stakeholder engagement seriously. The journal's aims state that:

> . . . engagement *with* research goes further than participation *in* it (original emphasis). Engaged individuals and communities initiate research, advise, challenge or collaborate with researchers. Their involvement is always active and they have a crucial influence on the conduct of the research – on its design or methods, products, dissemination or use.

Established in 2017, *Research for All* has brought together a research community conducting consultative work for stakeholder-informed research – a promising development.

Stakeholder input embedded in the research cycle is well-established in UK health research, with health bodies of the four UK nations issuing 'UK standards for public involvement' (NIHR et al., 2019). The National Institute for Health and Care Research (NIHR, 2021), the UK's largest funder of clinical healthcare research, requires researchers to demonstrate patient and public involvement when applying for grants, including, for example, by having patients with first-hand experience of the condition or phenomenon being researched serving on the funding panel or steering committee. The NIHR has a training infrastructure to support researchers' engagement efforts with patients and health practitioners, including for priority setting. Overall, NIHR and other healthcare funders have driven change in putting patients at the heart of the research process and ensuring, through top-down funding application requirements, that patients' perspectives are built into grant applications. There is evidence that patient and public involvement can improve research quality, enhance end-user involvement and experience, and promote participant enrolment in clinical trials (e.g., Price et al., 2018). This underscores the importance of ensuring that stakeholder involvement is not simply a tokenistic, box-ticking exercise, but that there is genuine engagement.

An innovative example of public involvement is 'the People's Trial' (Finucane et al., 2021), an Irish-led web-based project publicised through social media that involved volunteers from the general public in trial design decisions. This was designed to improve the public's understanding of randomised controlled trials and their ability to think critically when evaluating health claims. The study is unique in that the research questions and variables of interest (i.e., interventions, control/comparator, outcome measures) were not known at the outset of the study and were defined through panelling members of the public's ideas and preferences. In a similar process to the survey component of Chalmers et al.'s (2021) priority setting partnership, participants generated and then evaluated unanswered questions that compared the effectiveness of health-related interventions. This enabled the researchers to rank those ideas in order of priority. The highest ranked question asked whether or not reading a book at bedtime improved sleep quality. To investigate this, participants were randomly allocated to either a read-in-bed condition (15–30 min before going to sleep), or a no-reading condition. Both groups adhered to otherwise identical bedtime routines and followed the same caffeine consumption restrictions. At the end of the weeklong trial, 42% of reading group participants said that their sleep had improved compared with 28% in the non-reading group. The researchers concluded that there was some perceived benefit of reading in bed before sleeping compared with not reading. There is potential in applied linguistics for studies where participants define the research questions along these lines, with researchers acting as facilitators rather than as unilateral agenda-setters.

In sum, asking the right questions can play a crucial role in mitigating the first source of research waste. Equally important is that the research captures outcome measures that are important and meaningful to stakeholders. In 1993, the American College of Rheumatology (ACR) administered a survey to rheumatologists to develop a set of outcome measures recommended for investigation in trials

involving patients with rheumatoid arthritis (Felson et al., 1993). The outcome measures that these experts prioritised were assessments of patients' pain, physical function, and swelling and tenderness. Tellingly, people who were actually suffering from rheumatoid arthritis were not consulted. When patients were finally asked which outcomes were important to them (Hewlett et al., 2005), they identified fatigue as a priority. The outcomes that the ACR prioritised for attention were well-meaning but misdirected. This example highlights the importance of listening to stakeholder voices to ensure that related research addresses what matters to them.

Educational funders of experimental and quasi-experimental studies that focus solely on academic attainment as outcome measures, to the exclusion of others, may be too restrictive (see e.g., Lortie-Forgues & Inglis', 2019 meta-analysis of 141 large-scale randomised controlled trials in education). Restricting outcome measures to test scores and course grades implies that these are the only valid outcomes of education research. This view must be challenged. Other potential outcomes, and most especially those that have been demonstrated to matter to stakeholders, must also be considered to help ensure that the research is maximally meaningful, and, therefore, less wasteful.

2.2 Conducting unnecessary and/or poorly-designed research

2.2.1 Unnecessary research

The second source of research waste is 'doing studies that are unnecessary, or poorly designed.' Chalmers and Glasziou elaborate, 'new research should not be done unless, at the time it is initiated, the questions it proposes to address cannot be answered satisfactorily with existing evidence' (2009, p. 87). Conducting new research that addresses a question for which an answer is already known wastes time and resources. For intervention studies, failing to take into account results from previous studies risks unnecessarily exposing participants to ineffective, inefficient, or possibly detrimental treatments. To minimise the chances of waste of this sort, we must take into account what is already known about a topic before embarking on new related research.

Systematic reviews are considered by some to occupy the top of the research evidence hierarchy, offering comprehensive authority on a topic, if done well (Murad et al., 2016). For the uninitiated, a systematic review is an approach to evidence synthesis aiming to consider the totality of relevant evidence relating to a topic in a systematic, transparent, and replicable way (Gough et al., 2012). By systematically considering everything that has already been done to address a given research question, we protect ourselves from being misled by biases (e.g., citation bias, hot stuff bias, conscious/unconscious bias), and are, thus, more likely to reach conclusions that genuinely reflect what is already known. Referring to published systematic reviews before embarking on new research, or undertaking them ourselves if no trustworthy reviews exist, is, therefore, crucial in helping ensure that new research builds on what we already know rather than unnecessarily duplicating it.

Macleod et al. (2014) note that over 50% of health intervention studies are designed without reference to systematic reviews. Not paying systematic attention to the existing evidence can have unintended negative consequences, which can contribute to avoidable research waste. For example, at the height of the COVID-19 pandemic, the first author reviewed an NIHR grant application broadly on remote healthcare delivery (telehealth) for ethnic minorities. Because emerging data suggested that some ethnic groups were suffering more acutely from the effects of the pandemic than others (Webb Hooper et al., 2020), some planned projects integrated an ethnicity component. This was one such project. The NIHR (2022b) expects bids to 'demonstrate an awareness and understanding of previous relevant research,' often explicitly requiring applicants to explain 'the need for the proposed line of research … drawing particularly from systematic reviews and other relevant literature' (NIHR, 2022a). These researchers did not do that. As it turns out, systematic reviews about telehealth consultations among ethnic minority groups with the potential to inform their planned research were available (see Isaacs et al., 2016). Such reviews drew conclusions pre-pandemic that would have been important for researchers preparing new primary research to know (e.g., technological accessibility, language barriers, etc.). Addressing the challenges for ethnic minority groups in 2020, when that

grant application was submitted, was urgent. Reference to existing systematic reviews would have allowed immediate action to be taken to start addressing the challenges and would have informed decisions about the research questions, outcome measures, and designs of new primary research. Patients would have been better served and resources more usefully deployed.

Despite the fact that a social scientist coined the term 'meta-analysis' (Glass, 1976), colleagues in healthcare now more routinely conduct and use the evidence that they generate. For example, the Cochrane Collaboration, established in 1993, has conducted and published over 7,500 systematic reviews in healthcare (Cochrane, 2023). Its closest cousin in the social sciences, the Campbell Collaboration, established seven years thereafter, has published a comparatively small 224 systematic reviews (2023). This notwithstanding, there is evidence of growing momentum for synthesis research in applied linguistics – for example, through professional associations (e.g., BAAL Research Synthesis in Applied Linguistics SIG), journal special issues (Chong et al., 2023), and dedicated article types for evidence syntheses with expanded word count allowances in some journals (e.g., Harding & Winke, 2022). Applied linguists can work towards addressing the related category of avoidable research waste by consulting evidence syntheses when available, regardless of the funder's or editor's requirements, or conducting their own syntheses when they are not.

While potential negative outcomes of failing to consider existing evidence properly in applied linguistics research are perhaps less dramatic than failures to do so in healthcare research (people rarely die because of wasteful applied linguistics research), properly taking into account existing evidence avoids other wasteful practices. When a body of evidence investigating the same topic exists and a robust understanding of a phenomenon can be articulated, it is questionable whether new primary research advances knowledge in meaningful ways. If knowledge is not advanced by newly addressing old questions, resources ought to be directed elsewhere. An example of potentially wasteful practices can be found in bilingualism research. Melby-Lervåg and Lervåg's (2011) meta-analysis exploring crosslinguistic transfer in bilinguals, for example, produced numerous meta-correlations of evidence relating to first language (L1) and L2 proficiency. Figure 2, reproduced from the original publication, illustrates one of these.

In over a decade of study, research in different contexts with different participants and different L1s and L2s have coalesced around the same general finding: a positive relationship between L1 and L2 phonological awareness. Collectively, they suggest a meta-correlation of about $r = .6$. So, how many more grant applications should be written, participants recruited, materials created, tax contributions spent, and person-hours dedicated to addressing this question anew? Nonetheless, this seemingly settled position continues to attract new (and, therefore, potentially wasteful) research (see the systematic reviews by Míguez-Álvarez et al., 2022 and Yang et al., 2017, both of which include studies addressing the relationship between phonological awareness in L1 and L2 published after Melby-Lervåg and Lervåg (2011), and, therefore, after this question had been settled). We see this repeated elsewhere in our field. For example, there are at least six systematic reviews that conclude that bilingual education is more beneficial than monolingual education for linguistic minority children (Krashen & McField, 2005; McField, 2002; Reljić et al., 2015; Rolstad et al., 2005; Slavin & Cheung, 2005; Willig, 1985). Twelve systematic reviews synthesising evidence from 455 primary studies report that Mobile Assisted Language Learning is effective (Darmi & Albion, 2014; Elaish & Shuib, 2019; Huang, 2020; Lee et al., 2014; Li, 2022a; Lin & Lin, 2019; Mahdi, 2018; Peng et al., 2021; Persson & Nouri, 2018; Sung et al., 2015; Taj et al., 2016; Toto & Limone, 2019). Three meta-analyses, synthesising evidence from 202 studies, reveal negative correlations between self-reported foreign language anxiety and L2 performance (Li, 2022b; Teimouri et al., 2019; Zhang, 2019).

We recognise the importance, indeed necessity, of replication, both at the primary and meta-analytic levels. That said, we must be alert to the possibility that necessary replication can morph into unnecessary duplication if the totality of the existing available evidence is not taken into account when embarking on new research. The overabundance of primary research addressing the same or very similar questions that has allowed the syntheses we have identified above to be conducted in the first place suggests the possibility that some of it is merely reaffirming our understanding of a topic rather than expanding it.
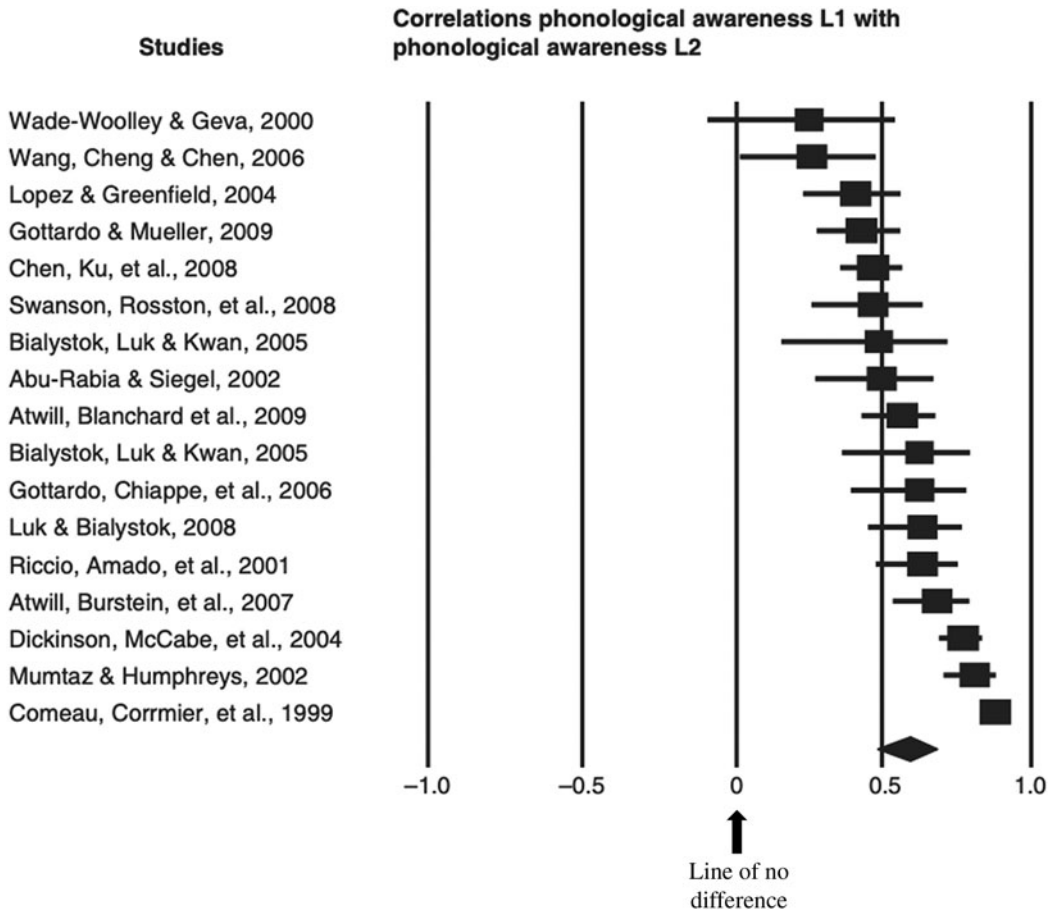
Figure 2. How many more studies do we need to be convinced that there are cross-linguistic relationships between L1 and L2 phonological awareness?
Source: Melby-Lervåg and Lervåg (2011, p. 126), reprinted with permission. Note that the line at '0' on the x-axis (i.e., the middle line in the figure) is the line of no difference. Studies to the right of this line show a positive association between L1 and L2 phonological awareness. Studies to the left reveal a negative association.

There are ways to address this source of research waste. A large group of clinical trial methodologists, for example, developed guidance to help researchers assess whether new planned research is worthy of further investigation (Treweek et al., 2020). They advise healthcare researchers to consider the findings from existing relevant systematic reviews, including the extent of uncertainty around estimated effects, the similarity of the contexts in which the contributing evidence was generated, the extent of clarity around the benefits and drawbacks to participants, and so forth. Another way that helps researchers think carefully about their contribution to knowledge is to require authors to explicitly state 'What is already known on this topic' and 'What this study adds.' The prestigious medical journal, *BMJ*, requires this for abstracts (e.g., Crocker et al., 2018). The point here is to highlight that decision-aids and reporting practices can help researchers assess whether their suggested topic is a good candidate for further research.

The International Database of Education Systematic Reviews (IDESR, n.d.) is an online database of published systematic reviews, a registry of review protocols in all areas in education, and the only dedicated database for published systematic reviews in language education. Researchers should consult it before undertaking new language education research to establish what is already known, update that

knowledge if necessary, and only then decide whether their planned primary or secondary research is warranted.

2.2.2 Poorly or inappropriately designed research

Assuming that a new primary study is well-motivated, another source of research waste concerns methodological soundness. In an article that inspired our exploration of this theme in applied linguistics, Macleod et al. (2014) reported that half of health intervention studies do not take the necessary methodological measures to reduce biases that could mislead us in interpretating the findings. For example, to reduce the possibility that systematic differences between groups being compared in an intervention study could confound the effects of the interventions being evaluated, participants should be randomly allocated to those groups. To reduce the possibility of allocation bias, information about participants and nature of the intervention to which they are being allocated should be concealed from the people doing the allocating. To reduce unconscious bias in interpretating data, data analysts should not know the group allocation (i.e., experimental or control) of the data they are analysing. To avoid statistical imprecision, the number of participants should be large enough to detect an unbiased estimate of an effect. This appears only to have worsened over time. A recent audit of published healthcare research estimated that 62% of clinical trials had a high risk of bias and only 8% an unequivocally low risk of bias (Pirosca et al., 2022). This failure to take appropriate steps to minimise bias in healthcare research, where these design considerations are generally well-understood, is concerning. In applied linguistics research, it is still all too common to find research that has not taken appropriate steps to minimise the potential for biases to mislead (see, for example, the risk of bias appraisals in Chalmers, 2019; Huang & Chalmers, 2023). Of course, researchers working within any research paradigm should do what they can to ensure that their work is rigorous and fair. When measures to support rigorous and fair research are not taken, our trust in the resulting findings is reduced. Untrustworthy research is wasteful research.

In healthcare research, several tools exist that can help assess the trustworthiness of research. One of these is the Cochrane risk of bias tool (Cochrane Methods, 2022). To evaluate methodological quality, the tool can be used to audit published reports for evidence of six sources of bias (selection, performance, detection, attrition, reporting, and 'other' sources of bias). It can help to judge the extent to which each source of bias has been mitigated and generate ratings of low, high, or unclear risk of bias for each study. When used in evidence syntheses, the results of this audit feed into assessments of the overall trustworthiness of included studies (see Isaacs et al., 2016, for a systematic review on language and health that adopts these Cochrane conventions). When planning and conducting new research, the principles laid out in the tool can inform the design and conduct of those studies.

In applied linguistics, there is currently no established tool for assessing methodological quality. However, tools used in other fields to assess the quality of different research designs may be instructive. The Mixed Methods Appraisal Tool (Hong et al., 2018), Quality in Qualitative Evaluation tool (Spencer et al., 2003), Eight 'Big Tent' criteria for Excellent Qualitative Research (Tracey, 2010), Newcastle Ottawa Scale for observational studies (Wells et al., 2021), and AXIS for cross-sectional studies (Downes et al., 2016) could help researchers assess the quality and trustworthiness of applied linguistics research of all stripes.

In an example of where failing to account for methodological quality could mislead us about our understanding of a topic, we turn to an influential systematic review on the effects of different types of instructional interventions on L2 performance. Norris and Ortega (2000) are rightly celebrated for pioneering work promoting the use of evidence syntheses in applied linguistics research. They located 77 relevant reports of experiments and quasi-experiments (i.e., higher and lower quality designs for causal inferences), synthesising the findings of 49 such studies in their meta-analysis to conclude that 'explicit types of instruction are more effective than implicit types, and that Focus on Form and Focus on Forms interventions result in equivalent and large effects' (p. 417). However, they failed to identify which of the included studies adopted experimental versus quasi-experimental designs (more and less trustworthy in terms of internal validity, respectively). Norris and Ortega reported

these details in aggregate but not at individual study level. They also did not assess other aspects of the methodological quality of the studies. Had the authors provided separate syntheses for the experiments and quasi-experiments or used a tool like Cochrane's risk of bias to discriminate between higher and lower quality evidence (Cochrane Methods, 2022), the bottom-line conclusion might have looked quite different.

Some more recent syntheses in applied linguistics have revealed more serious shortcomings. For example, Bryfonski and McKay's (2019) meta-analysis on the effectiveness of task-based language teaching (TBLT) has been subject to robust criticism. Xuan et al. (2022) highlighted that the authors' use of 'loose inclusion criteria' misaligned with their stated research aims while also raising technical concerns (e.g., effect size calculations). Linked to the former point, and among other criticisms, Harris and Leeming (2022) underscored Bryfonski and McKay's (2019) inclusion of studies that: (1) lacked a control group and/or lacked or used different pre- and post-tests, (2) had different instructional orientations (e.g., grammar-translation and communicative) without accounting for these differences, and (3) published in predatory journals, four of which – on close inspection – they claimed contained plagiarised content. These flaws undermine the review's credibility and any conclusions that can be drawn from it – a wasteful endeavour with potentially misleading results.

2.3 Inefficiencies in the regulation and management of research

The third source of research waste centers on 'inefficiencies in the regulation and management of research' (Macleod et al., 2014 p. 2). This refers to inconsistently applied or unduly burdensome bureaucratic processes that researchers are often confronted with. For example, ethical principles and processes protecting human dignity are essential for all research involving human participants (e.g., British Educational Research Association, 2018; World Medical Association, 2013). However, a researcher proposing to conduct interviews on language needs and use in healthcare settings, for example, should not be subject to the same bureaucratic burdens, legal safeguards, risk assessments, and researcher background checks as a researcher conducting invasive health intervention research with potentially harmful side-effects. When organisations with oversight of this process apply a one-size-fits-all approach to ethics review, they risk erecting unnecessary roadblocks (Haggerty, 2004; Snooks et al., 2023), potentially wasting valuable time, resources, and opportunities.

Bureaucratic overreach can clearly cause avoidable research waste. So too can bureaucratic inaction. We have already described the importance of understanding existing evidence before undertaking new research. Occasionally in applied linguistics research, we find that existing evidence is faulty, plagiarised, or fraudulent (Isbell et al., 2022). In such circumstances, it is important that mechanisms exist and are followed to address this. As pages of testimony on the website *Retraction Watch* (https://retractionwatch.com) illustrate, this is by no means straightforward. On discovering academic misconduct in a published study, if the academic record is not corrected, with explanations accompanying corrections or retractions, we cannot recalibrate our understanding to take account of the change in the evidence landscape (Xu & Hu, 2023).

It is not clear to us that robust systems are in place in our field to deal with waste stemming from bureaucratic inefficiencies rigorously, methodically, consistently, and fairly. No doubt, money, prestige, and vested interests in addition to human fallibility can influence these bureaucratic processes, but we believe that more can be done to minimise these and related sources of research waste.

2.4 Comprehensive dissemination and mitigating publication bias

The fourth source of avoidable research waste is 'failure to publish relevant research promptly, or at all' (Chalmers & Glasziou, 2009, p. 87). A major contributor to this is publication bias – the tendency to write up, submit, and have accepted for publication studies that demonstrate significant effects more readily than studies that do not demonstrate an effect. When this happens, a biased picture of the evidence results. Evidence suggesting that an intervention works is overrepresented, whereas

contradictory or unexciting evidence is underrepresented (Sun et al., 2018). This biases evidence syntheses and means that decisions about policy and practice are based on an incomplete picture (with all the financial and logistical waste that this implies).

For instance, it is orthodox to consider that being bilingual confers specific cognitive benefits. However, a careful examination of publishing habits in the field reveals an apparent reticence to publish studies that go against that orthodoxy. de Bruin et al. (2014) examined conference abstracts over a 13-year period reporting on studies assessing the validity of the claim that bilinguals have an executive functioning advantage over monolinguals. They then followed up to examine the extent to which these works-in-progress were formally published. They found that 68% of studies supporting bilingual advantage theory had been published compared with only 29% of studies challenging it. If a sizeable proportion of the evidence is hidden from view, how can we make well-informed, unbiased assessments? Some aspects of this waste source can be addressed by researchers – for example, protecting against the file drawer problem by, regardless of its outcomes, promptly writing up and submitting every study they conduct and making preprints available. However, many aspects of this waste category are out of researchers' hands. As above, even if research is timely, well conducted, and well reported, nonorthodox results tend to be rejected by journals more frequently than orthodox ones (Plonsky, 2013). Again, we can learn from other fields. Applied linguistics editors and reviewers should be encouraged to see the value of publishing null findings in cases of well-conducted studies, and author guidelines should encourage submission of robust studies regardless of their findings. This can be argued on ethical and academic grounds. There is an ethical obligation to ensure that the goodwill of participants is not wasted by failing to publish the knowledge that they helped generate through their participation. Many will have signed consent agreements stating that contribution to our understanding is a benefit of participation. There is also an ethical obligation to ensure that the money used to conduct research, often from the public purse, is not wasted by failing to report the research that it funded.

In accordance with the open science movement, *Language Learning* and *Language Testing* now have a 'Registered Reports' (i.e., research blueprint) manuscript submission category, which contributes to addressing this waste source (Isaacs & Winke, in press; Marsden et al., 2018). On acceptance of the research protocol, the journal agrees in principle to publish the completed study, irrespective of its findings.

Prospective registration of research protocols is, as yet, uncommon in applied linguistics. But this should not stop researchers who are serious about their scholarship from making their protocols publicly available prior to publication. Using platforms like Open Science Framework (OSF, Center for Open Science, 2011-2023), researchers can prospectively publish protocols as well as research instruments, datasets, and statistical code. When research protocols are published, the ensuing research is less likely to be lost from view, and, if a promised report fails to materialise, the authors can be asked why.

This principle holds for evidence synthesis. In addition to providing a free online library of systematic reviews, IDESR, described above, provides a protocol registry for planned and ongoing systematic reviews in language education. Modelled on PROSPERO, which has been publishing systematic review protocols in healthcare for over a decade (Booth et al., 2012), IDESR accepts submissions of review protocols, which detail the background literature and methods used. Any protocol amendments that the authors make while undertaking the review can be documented with justification. As well as encouraging good scientific practice, registered reports and prospectively published protocols help ensure that studies do not get lost from view, with the knowledge they generate less likely to go to waste.

2.5 Untransparent, inaccessible, irreplicable, biased reporting

The fifth source of research waste is 'biased or unusable reports of research' (Chalmers & Glasziou, 2009, p. 87). We explicate a component of this waste category using a historical example. Hill was

a British statistician and epidemiologist well-known for his 1948 streptomycin trial, often (erroneously) considered the first randomised trial in healthcare research. Another methodological contribution was his work on clarity and completeness in research reporting, articulating four simple questions that researchers must address: 'Why did you start, what did you do, what answer did you get, and what does it mean anyway?' (1965, p. 870).

Without clear reporting of primary research and with omission of relevant methodological details, a study cannot be properly understood and is a poor candidate for replication (Porte & McManus, 2019). Similarly, a systematic reviewer cannot reliably or comprehensively extract key details from a study nor assess its methodological quality when it is poorly reported. A recent audit of 307 systematic reviews in language education (Chalmers et al., 2023) and 120 meta-analyses of L2 experimental and quasi-experimental studies (Vuogan & Li, 2023) have revealed notable omissions in research reporting (e.g., eligibility criteria, primary study/participant sample characteristics, articulation of the research questions, risk of bias/trustworthiness, statistical model selection, etc.). When important information that allows readers to interpret a review and its findings are absent from supposedly authoritative syntheses of a field of research, readers are denied the opportunity to accurately assess the state of knowledge and strength of evidence – a wasted opportunity with repercussions for subsequent research that relies on it for information. In addition to errors of omission, there may be 'errors' of commission in the way reports are written (e.g., to obfuscate or gloss over unflattering results). Another outcome of de Bruin et al.'s (2014) investigation into publication bias in bilingualism research was their observation of an apparent artefact in the literature that they called the decline effect. They observed that the fewer the outcome measures (dependent variables) reported, the more likely a study was to find an overall bilingual advantage (de Bruin & Della Sala, 2019). The authors speculated that selective outcome reporting (i.e., authors cherry picking and only reporting outcome measures confirming the orthodoxy) explains this finding. Plonsky (2013) identified similar trends in his review of 606 SLA studies (1990–2010), whereby authors were much less likely to fully report statistically nonsignificant results. Compounding the effects of publication bias, selective outcome reporting further devalues research and works against conveying a representative picture of the state of knowledge. As much as the prospective registration processes described above could help address publication bias, they can also mitigate dishonest reporting practices. A prospectively published protocol allows peer-reviewers and critical consumers of research to examine the alignment of completed research outputs with the research plan and to query substantial deviations.

In healthcare research, there are additional tools used by convention to improve the quality of primary and secondary reporting, and, thus, mitigate the waste associated with biased or unusable research outputs. A reporting guideline is a 'checklist, flow diagram, or structured text to guide authors in reporting a specific type of research, developed using explicit methodology' (Equator Network, n.d.). This is essentially reified and more detailed guidance built on Hill's succinct message. The Equator Network, an organisation that creates, curates, and disseminates reporting guidelines for healthcare research, describes them as 'a minimum list of information needed to ensure a manuscript [which] can be, for example:

- Understood by a reader,
- Replicated by a researcher,
- Used by a doctor to make a clinical decision, and
- Included in a systematic review.'

In healthcare research, teams of experienced researchers typically create reporting guidelines using explicit methods to transparently document the tool's development and then periodically update it to reflect advances in the field. Using appropriate guidelines can improve the transparency and comprehensiveness of research reports. For instance, the Consolidated Standards of Reporting Trials (CONSORT) guidelines are used for reporting randomised trials (Schulz et al., 2010). The items for inclusion are presented by section of the final report. Under each section, recommended items are

elaborated. For example, under 'Title and Abstract,' Item 1 requires authors to identify the study as a randomised trial. Under 'Methods,' Item 5 asks authors to describe the interventions used in each group in enough detail to enable replication. Item 8 asks them to describe the allocation method. Under 'Results,' Item 13a recommends showing the numbers of participants assessed for eligibility, recruited, randomly allocated, and who dropped out in a flow diagram. In short, the 37 checklist items encourage authors to fully and transparently report their research to make it understandable to readers, amenable to replication, usable in systematic reviews, and able to inform practice. CONSORT has been shown to improve research reporting quality in healthcare in synthesis research (e.g., Turner et al., 2012).

There are many other similar guidelines that can be used or adapted for applied linguistics research. The Equator Network has 521 freely available reporting guidelines for varied study designs (some, as the number implies, extremely niche). Of the more generally applicable guidelines, those that seem most relevant to applied linguistics research in addition to CONSORT are PRISMA for systematic reviews (Page et al., 2021), STROBE for observational studies (von Elm et al., 2007), SRQR for qualitative research (O'Brien et al., 2014), and STROCSS for cohort, cross-sectional, and case-control studies (Mathew et al., 2021). The American Psychological Association (APA) published updated reporting 'standards' for quantitative journal articles in Appelbaum et al. (2018). Some applied linguistics journals, in turn, have published reporting 'guidelines' (e.g., Norris et al., 2015, for quantitative studies in *Language Learning*; Mahboob et al., 2016, for notes and examples for different study types, including qualitative traditions, in *TESOL Quarterly*). As in healthcare research, applied linguistics journals insisting on the adoption of appropriate standards/guidelines for authors have the potential to drive progress in our field, promoting positive washback effects on study design and more rigorous and systematic peer-review.

3. Taking stock and recommendations

In this article, we have explored the issue of avoidable research waste, using conceptual thinking from healthcare research to interrogate the applicability of related concepts in applied linguistics research. Using illustrative examples, we have demonstrated that research waste is often not discipline-specific. The five sources of avoidable research waste that Chalmers and Glasziou (2009) and Macleod et al. (2014) identified – asking the wrong questions, failing to situate new research in the context of existing research, inefficient research regulation and management, failing to disseminate findings, and poor reporting practices – are as relevant to the quality of applied linguistics research as they are to healthcare research.

We are encouraged by movements in our field beginning to address these issues. For example, stakeholder involvement in research has been demonstrated to be possible and valuable for addressing the first waste source (e.g., Elliott & Hodgson, 2021). Publication categories for registered reports, coupled with the growing prevalence of evidence syntheses in our field and use of online repositories and supplemental appendices to ensure the completeness and availability of findings, help address the second source of waste. Applied linguists, and particularly those in leadership positions, can and do work with institutions to remove unnecessary bureaucracy that can impede research progress. While it is relatively rare for journals to mandate adherence to reporting guidelines, we propose that authors should nonetheless adopt reporting guidelines appropriate to their research design to help to address the fifth source of waste. While the beginnings of a movement might be detectable here, we are a long way from establishing these as norms. As a field, applied linguistics can do better. We hope that the thinking presented in this paper encourages fellow applied linguists (including funders and publishers) to see that we can improve research relevance and quality by addressing these waste sources.

References

Allwright, D. (2005). Developing principles for practitioner research: The case of exploratory practice. *Modern Language Journal*, *89*(3), 353–366. doi:10.1111/j.1540-4781.2005.00310.x

Allwright, D., & Hanks, J. (2009). *The developing language learner: An introduction to exploratory practice*. Palgrave Macmillan.

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. doi:10.1037/amp0000389

Booth, A., Clarke, M., Dooley, G., Ghersi, D., Moher, D., Petticrew, M., & Stewart, L. (2012). The nuts and bolts of PROSPERO: An international prospective register of systematic reviews. *Systematic Reviews*, *1*(2), 2–10. doi:10.1186/2046-4053-1-2

British Educational Research Association. (2018). *Ethical guidelines for educational research* (4th ed.). British Educational Research Association (BERA). doi: https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2018.

Bryfonski, L., & McKay, T. H. (2019). TBLT implementation and evaluation: A meta-analysis. *Language Teaching Research*, *23*(5), 603–632. doi:10.1177/1362168817744389

Burns, A. (2010). *Doing action research in English language teaching. A guide for practitioners*. Routledge.

Campbell Collaboration. (2023). *Research evidence*. Retrieved March 1, 2023, from https://www.campbellcollaboration.org/better-evidence.html

Center for Open Science. (2011-2023). OSF home. https://osf.io/

Chalmers, H. (2019). Leveraging the L1: The role of EAL learners' first language in their acquisition of English vocabulary. [Ph.D thesis, Oxford Brookes University, UK]. doi:10.24384/fhr0-jr5

Chalmers, H., Brown, J., & Koryakina, A. (2023). Topics, publication patterns, and reporting quality in systematic reviews in language education. Lessons from the international database of education systematic reviews (IDESR). *Applied Linguistics Review*. doi:10.1515/applirev-2022-0190

Chalmers, H., Faitaki, F., & Murphy, V. A. (2021). *Setting research priorities for English as an Additional Language: What do stakeholders want from EAL research?* Report submitted to the British Association for Applied Linguistics (BAAL). https://ealpsp.files.wordpress.com/2021/09/eal-psp-report.pdf

Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *The Lancet*, *374*(9683), 86–89. doi:10.1016/S0140-6736(09)60329-9

Chong, S. W. (2022). Research synthesis in applied linguistics: Facilitating research-pedagogy dialogue. *Language Teaching*, *55*(1), 142–144. doi:10.1017/S0261444821000343.

Chong, S. W., Bond, M., & Chalmers, H. (2023). Opening the methodological black box of research synthesis in language education: Where are we now and where are we heading? *Applied Linguistics Review*, (1). https://doi.org/doi:10.1515/applirev-2022-0193

Clarivate. (2023). 2022 Journal Impact Factor. *Journal Citation Reports*.

Cochrane. (2023). *About us*. Retrieved July 12, 2023, from https://www.cochrane.org/about-us

Cochrane Methods. (2022). RoB 2: A revised Cochrane risk-of-bias tool for randomized trials. https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials

Crocker, J. C., Ricci-Cabello, I., Parker, A., Hirst, J. A., Chant, A., Petit-Zeman, S., Evans, D., & Rees, S. (2018). Impact of patient and public involvement on enrolment and retention in clinical trials: Systematic review and meta-analysis. *BMJ*, *363*, k4738. doi:10.1136/bmj.k4738

Darmi, R., & Albion, P. (2014). A review of integrating mobile phones for language learning. In A. Sanchez & P. Isaias, *Proceedings of the 10th international conference Mobile Learning* (pp. 93–100). IADIS.

de Bruin, A., & Della Sala, S. (2019). The bilingual advantage debate: Publication biases and the decline effect. In J. W. Schwieter (Ed.), *The handbook of the neuroscience of multilingualism* (pp. 736–753). Wiley. doi:10.1002/9781119387725.ch35

de Bruin, A., Treccani, B., & Della Sala, S. (2014). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, *26*(1), 99–107. doi:10.1177/0956797614557866

Downes, M. J., Brennan, M. L., Williams, H. C., & Dean, R. S. (2016). Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open*, *6*(12). doi:10.1136/bmjopen-2016-011458

Duarte, J., García-Jimenez, E., McMonagle, S., Hansen, A., Gross, B., Szelei, N., & Pinho, A. S. (2023). Research priorities in the field of multilingualism and language education: A cross-national examination. *Journal of Multilingual and Multicultural Development*, *44*(1), 50–64. doi:10.1080/01434632.2020.1792475

Elaish, M. M., & Shuib, L. (2019). Mobile English language learning (MELL): A literature review. *Educational Review*, *71*(2), 257–276. doi:10.1080/00131911.2017.1382445

Elliott, V., & Hodgson, J. (2021). Setting an agenda for English education research. *English in Education*, *55*(4), 369–374. doi:10.1080/04250494.2021.1978737

Ellis, R. (2010). Second language acquisition, teacher education and language pedagogy. *Language Teaching*, *43*(2), 182–201. doi:10.1017/S0261444809990139

Equator Network. (n.d.). *What is a reporting guideline?* Retrieved July 12, 2023, from https://www.equator-network.org/about-us/what-is-a-reporting-guideline/

Felson, D. T., Anderson, J. J., Boers, M., Bombardier, C., Chernoff, M., Fried, B., Furst, D., Goldsmith, C., Kieszak, S., & Lightfoot, R. (1993). The American college of rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. *Arthritis and Rheumatism*, *36*(6), 729–740. doi:10.1002/art.1780360601

Finucane, E., O'Brien, A., Treweek, S., Newell, J., Das, K., Chapman, S., Wicks, P., Galvin, S., Healy, P., Biesty, L., Gillies, K., Noel-Storr, A., Gardner, H., O'Reilly, M. F., & Devane, D. (2021). Does reading a book in bed make a difference to sleep in comparison to not reading a book in bed? The people's trial—an online, pragmatic, randomised trial. *Trials*, *22*(1), 873. doi:10.1186/s13063-021-05831-3

Furedi, F. (2013, September 9). Don't import the scourge of scientism into schools. Spiked Online. http://www.spiked-online.com/newsite/article/dont_import_the_scourge_of_scientism_into_schools/14005

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3–8. doi:10.2307/1174772

Glasziou, P., & Chalmers, I. (2018). Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers. *BMJ*, *363*, k4645. doi: 10.1136/bmj.k4645.

Gough, D., Oliver, S., & Thomas, J. (2012). *An introduction to systematic reviews*. SAGE.

GOV.UK. (n.d.). *UK research and innovation*. Retrieved March 25, 2023, from https://www.ukri.org

Haggerty, K. D. (2004). Ethics creep: Governing social science research in the name of ethics. *Qualitative Sociology*, *27*(4), 391–414. doi:10.1023/B:QUAS.0000049239.15922.a3

Hanks, J. (2019). From research-as-practice to exploratory practice-as-research in language teaching and beyond. *Language Teaching*, *52*(2), 143–187. doi:10.1017/S0261444819000016

Harding, L., & Winke, P. (2022). Innovation and expansion in language testing for changing times. *Language Testing*, *39*(1), 3–6. doi:10.1177/0265532221105321

Harris, J., & Leeming, P. (2022). The impact of teaching approach on growth in L2 proficiency and self-efficacy: A longitudinal classroom-based study of TBLT and PPP. *Journal of Second Language Studies*, *5*(1), 114–143. doi:10.1075/jsls.20014.har

Hewlett, S., Cockshott, Z., Byron, M., Kitchen, K., Tipler, S., Pope, D., & Hehir, M. (2005). Patients' perceptions of fatigue in rheumatoid arthritis: Overwhelming, uncontrollable, ignored. *Arthritis Care & Research*, *53*(5), 697–702. doi:10.1002/art.21450

Hill, A. B. (1965). The reasons for writing. *British Medical Journal*, *1965*, 870–872. doi:10.1136/bmj.2.5466.870

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Wedel, I. (2018). *Mixed methods appraisal tool (MMAT) version 2018 user guide*. http://mixedmethodsappraisaltoolpublic.pbworks.com/w/n/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf

Huang, X., & Chalmers, H. (2023). Implementation and effects of pedagogical translanguaging in EFL classrooms: A systematic review. *Languages*, *8*(3), 194. doi:10.3390/languages8030194

Huang, Z. (2020). *Thirteen years since the first iPhone: A systematic review on the effectiveness of language learning apps on smart devices* [Master's dissertation, University of Oxford].

IDESR. (n.d.). *International database of education systematic reviews*. https://idesr.org/

Isaacs, T., Hunt, D., Ward, D., Rooshenas, L., & Edwards, L. (2016). The inclusion of ethnic minority patients and the role of language in telehealth trials for type 2 diabetes: A systematic review. *JMIR*, *18*(9), e256. doi:10.2196/jmir.6374

Isaacs, T., & Winke, P. M. (in press). Purposeful turns for more equitable and transparent publishing in language testing and assessment. *Language Testing*, *41*(1). doi:10.1177/02655322231203234

Isbell, D. R., Brown, D., Chen, M., Derrick, D. J., Ghanem, R., Arvizu, M. N. G., Schnur, E., Zhang, M., & Plonsky, L. (2022). Misconduct and questionable research practices: The ethics of quantitative data handling and reporting in applied linguistics. *Modern Language Journal*, *106*(1), 172–195. doi:10.1111/modl.12760

James Lind Alliance. (2022). *About priority setting partnerships*. Retrieved July 12, 2023, from https://www.jla.nihr.ac.uk/about-the-james-lind-alliance/about-psps.htm

Krashen, S., & McField, G. (2005). What works? Reviewing the latest evidence on bilingual education. *Language Learner*, *1*(2), 7–10.

Lee, Y.-S., Sung, Y.-T., Chang, K.-E., Liu, T.-C., & Chen, W.-C. (2014). A meta-analysis of the effects of learning languages with mobile devices. In Y. Cao, T. Väljataga, J. K. T. Tang, H. Leung, & M. Laanpere (Eds.), *New horizons in web based learning: ICWL 2014* (pp. 106–114). Springer. doi:10.1007/978-3-319-13296-9_12).

Li, R. (2022a). Effects of mobile-assisted language learning on EFL/ESL reading comprehension. *Educational Technology & Society*, 25(3), 15–29. doi:10.30191/ETS.202304_26(2).0003.

Li, R. (2022b). Foreign language reading anxiety and its correlates: A meta-analysis. *Reading and Writing*, 35, 995–1018. doi:10.1007/s11145-021-10213-x

Lin, J.-J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning*, 32(8), 878–919. doi:10.1080/09588221.2018.1541359

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher*, 48(3), 158–166. doi:10.3102/0013189X198328

Macleod, M. R., Michie, S., Roberts, I., Dirnagl, U., Chalmers, I., Ioannidis, J. P., Al-Shahi Salman, R., Chan, A. W., & Glasziou, P. (2014). Biomedical research: Increasing value, reducing waste. *Lancet*, 383(9912), 101–104. doi:10.1016/S0140-6736(13)62329-6

Mahboob, A., Paltridge, B., Phakiti, A., Wagner, E., Starfield, S., Burns, A., Jones, R. H., & De Costa, P. I. (2016). TESOL quarterly research guidelines. *TESOL Quarterly*, 50(1), 42–65. doi:10.1002/tesq.288

Mahdi, H. S. (2018). Effectiveness of mobile devices on vocabulary learning: A meta-analysis. *Journal of Educational Computing Research*, 56(1), 134–154. doi:10.1177/0735633117698826

Mann, S., & Walsh, S. (2017). *Reflective practice in English language teaching: Research-based principles and practices*. Routledge.

Marsden, E., Morgan-Short, K., Trofimovich, P., & Ellis, N. C. (2018). Introducing registered reports at *Language Learning*: Promoting transparency, replication, and a synthetic ethic in the language sciences. *Language Learning*, 68(2), 309–320. doi: 10.1111/lang.12284.

Mathew, G., Agha, R., Albrecht, J., Goel, P., Mukherjee, I., Pai, P., D'Cruz, A. K., Nixon, I. J., Roberto, K., Enam, S. A., Basu, S., Muensterer, O. J., Giordano, S., Pagano, D., Machado-Aranda, D., Bradley, P. J., Bashashati, M., Thoma, A., Afifi, R. Y., Johnston, M., Challacombe, B., Chi-Yong Ngu, J., Chalkoo, M., Raveendran, K., Hoffman, J. R., Kirshtein, B., Lau, W. Y., Thorat, M.A., Miguel, D., Beamish, A. J., Roy, G., Healy, D., Ather, H. M., Raja, S. G., Mei, Z., Manning, T. G., Kasivisvanathan, V., Rivas, J. G., Coppola, R., Ekser, B., Karanth, V. L., Kadioglu, H., Valmasoni, M., Noureldin, A., & STROCSS Group (2021). STROCSS 2021: Strengthening the reporting of cohort, cross-sectional and case-control studies in surgery. *International Journal of Surgery*, 96, 106165. doi:10.1016/j.ijsu.2021.106165

McField, G. (2002). Does program quality matter? A meta-analysis of select bilingual education studies. [PhD thesis, University of Southern California, CA]. http://digitallibrary.usc.edu/cdm/ref/collection/p15799coll16/id/255011

McKinley, J. (2019). Evolving the TESOL teaching–research nexus. *TESOL Quarterly*, 53(3), 875–884. doi:10.1002/tesq.509

Melby-Lervåg, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading*, 34(1), 114–135. doi:10.1111/j.1467-9817.2010.01477.x

Míguez-Álvarez, C., Cuevas-Alonso, M., & Saavedra, A. (2022). Relationships between phonological awareness and reading in Spanish: A meta-analysis. *Language Learning*, 72(1), 113–157. doi:10.1111/lang.12471

Murad, M. H., Asi, N., Alsawas, M., & Alahdab, F. (2016). New evidence pyramid. *BMJ Evidence-Based Medicine*, 21(4), 25–127. doi:10.1136/ebmed-2016-110401

NIHR. (2021). *Patient and public involvement and engagement*. Retrieved June 1, 2021, from https://www.nihr.ac.uk/about-us/our-contribution-to-research/how-we-involve-patients-carers-and-the-public.htm

NIHR. (2022a). *Policy Research Programme: Guidance for stage 1 applications*. Retrieved May 19, 2022, from https://www.nihr.ac.uk/documents/policy-research-programme-guidance-for-stage-1-applications-updated/26398

NIHR. (2022b). *Policy Research Programme: Standard information for applicants*. Retrieved July 12, 2023, from https://www.nihr.ac.uk/documents/policy-research-programme-standard-information-for-applicants/27427

NIHR, Chief Scientist Office, Health and Care Research Wales, & HSC Public Health Agency. (2019). *UK standards for public involvement*. https://sites.google.com/nihr.ac.uk/pi-standards/standards?authuser=0

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417–528. doi:10.1111/0023-8333.00136

Norris, J. M., Plonsky, L., Ross, S. J., & Schoonen, R. (2015). Guidelines for reporting quantitative methods and results in primary research. *Language Learning*, 65(2), 470–476. doi:10.1111/lang.12104

O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., & Cook, D. A. (2014). Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine: Journal of the Association of American Medical Colleges*, 89(9), 1245–1251. doi:10.1097/ACM.0000000000000388

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. doi:10.1136/bmj.n71

Peng, H., Lowie, W., & Jager, S. (2021). Narrative review and meta-analysis of MALL research on L2 skills. *ReCALL Journal*, 33(3), 278–295. doi:10.1017/S0958344020000221

Persson, V., & Nouri, J. (2018). A systematic review of second language learning with mobile technologies. *International Journal of Emerging Technologies in Learning*, 13(2), 188–210. doi:10.3991/ijet.v13i02.8094

Pirosca, S., Shiely, F., Clarke, M., & Treweek, S. (2022). Tolerating bad health research: The continuing scandal. *Trials*, 23 (458). doi:10.1186/s13063-022-06415-5

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. doi:10.1017/S0272263113000399

Porte, G. K., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.

Price, A., Albarqouni, L., Kirkpatrick, J., Clarke, M., Liew, S. M., Roberts, N., & Burls, A. (2018). Patient and public involvement in the design of clinical trials: An overview of systematic reviews. *Journal of Evaluation in Clinical Practice*, 24(1), 240–253. doi:10.1111/jep.12805

Reljić, G., Ferring, D., & Martin, R. (2015). A meta-analysis on the effectiveness of bilingual programs in Europe. *Review of Educational Research*, 85(1), 92–128. doi:10.3102/0034654314548514

Research for All. (2022). *Research for All: Aims and scope*. Retrieved July 12, 2023, from https://www.uclpress.co.uk/pages/research-for-all

Robinson, K. A., Brunnhuber, K., Ciliska, D., Juhl, C. B., Christensen, R., & Lund, H. (2021). Evidence-based research series-paper 1: What evidence-based research is and why it is important? *Journal of Clinical Epidemiology*, 129 (January), 151–157. doi:10.1016/j.jclinepi.2020.07.020

Rolstad, K., Mahoney, K., & Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy*, 19(4), 572–594. doi:10.1177/0895904805278067

Sato, M., & Loewen, S. (2018). Do teachers care about research? The research–pedagogy dialogue. *ELT Journal*, 73(1), 1–10. doi:10.1093/elt/ccy048

Sato, M., & Loewen, S. (2022). The research–practice dialogue in second language learning and teaching: Past, present, and future. *Modern Language Journal*, 106(3), 509–527. doi:10.1111/modl.12791

Sato, M., Loewen, S., & Pastushenkov, D. (2022). 'Who is my research for?': Researcher perceptions of the research–practice relationship. *Applied Linguistics*, 43(4), 625–652. doi:10.1093/applin/amab079

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*, 340, c332. doi:10.1136/bmj.c332

Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research*, 75(2), 247–284. doi:10.3102/00346543075002247

Snooks, H., Khanom, A., Ballo, R., Bower, P., Checkland, K., Ellins, J., Ford, G. A., Locock, L., & Walshe, K. (2023). Is bureaucracy being busted in research ethics and governance for health services research in the UK? Experiences and perspectives reported by stakeholders through an online survey. *BMC Public Health*, 23(1), 1119. doi:10.1186/s12889-023-16013-y

Spencer, L., Ritchie, J., Lewis, J., & Dillon, L. (2003). *Quality in qualitative evaluation: A framework for assessing research evidence*. Government Chief Social Researcher's Office. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/498321/Quality-in-qualitative-evaulation_tcm6-38739.pdf

Sterling, S., Plonsky, L., Larsson, T., Kytö, M., & Yaw, K. (2023). Introducing and illustrating the Delphi method for applied linguistics research. *Research Methods in Applied Linguistics*, 2(1), 100040. doi:10.1016/j.rmal.2022.100040

Sun, J., Freeman, B. D., & Natanson, C. (2018). Meta-analysis of clinical trials. In J. I. Gallin, F. P. Ognibene, & L. L. Johnson (Eds.), *Principles and practice of clinical research* (4th ed., pp. 317–327). Academic Press. doi:10.7326/0003-4819-107-2-224

Sung, Y.-T., Chang, K.-E., & Yang, J.-M. (2015). How effective are mobile devices for language learning? A meta-analysis. *Educational Research Review*, 16, 68–84. doi:10.1016/j.edurev.2015.09.001

Taj, I.-H., Sulan, N. B., Sipra, M. A., & Ahmed, W. (2016). Impact of mobile assisted language learning (MALL) on EFL: A meta-analysis. *Advances in Language and Literary Studies*, 7(2), 77–83. doi:10.7575/aiac.alls.v.7n.2p.76

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, 41(2), 363–387. doi:10.1017/S0272263118000311

Thomas, G. (2013, February 3). No one can control for a sense of when 4-3-3 might turn the game. *Times Higher Education*. https://www.timeshighereducation.com/comment/opinion/no-one-can-control-for-a-sense-of-when-4-3-3-might-turn-the-game/2001371.article

Toto, G. A., & Limone, P. (2019). Contemporary trends in studies on mobile learning of foreign languages: A meta-analysis. *International Journal of Engineering Education*, 1(2), 85–90. doi:10.14710/ijee.1.2.85-90

Tracey, S. J. (2010). Qualitative quality: Eight 'big-tent' criteria for excellent qualitative research. *Qualitative Inquiry*, 16(10), 837–851. doi:10.1177/107780041038312

Treweek, S., Bevan, S., Bower, P., Briel, M., Campbell, M., Christie, J., Collett, C., Cotton, S., Devane, D., El Feky, A., Galvin, S., Gardner, H., Gillies, K., Hood, K., Jansen, J., Littleford, R., Parker, A., Ramsay, C., Restrup, L., … & Clarke, M. (2020). Trial forge guidance 2: How to decide if a further Study Within A Trial (SWAT) is needed. *Trials*, 21(1), 33. doi:10.1186/s13063-019-3980-5

Turner, L., Shamseer, L., Altman, D. G., Schulz, K. F., & Moher, D. (2012). Does use of the CONSORT statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Systematic Reviews*, *1*, 60. doi:10.1186/2046-4053-1-60

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P., & STROBE Initiative (2007). The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Annals of Internal Medicine*, *147*(8), 573–577. doi:10.7326/0003-4819-147-8-200710160-00010

Vuogan, A., & Li, S. (2023). A systematic review of meta-analyses in second language research: Current practices, issues, and recommendations. *Applied Linguistics Review*, doi:10.1515/applirev-2022-0192

Webb Hooper, M., Nápoles, A. M., & Pérez-Stable, E. J. (2020). COVID-19 and racial/ethnic disparities. *JAMA*, *323*(24), 2466–2467. doi: 10.1001/jama.2020.8598

Wells, G. A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2021). *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses*. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Willig, A. C. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, *55*(3), 269–317. doi:10.2307/1170389

World Medical Association. (2013). World medical association declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191–2194. doi: 10.1001/jama.2013.281053

Xu, S. B., & Hu, G. (2023). What to communicate in retraction notices? *Learned Publishing*, *36*(3), 463–467. doi:10.1002/leap.1548

Xuan, Q., Cheung, A., & Liu, J. (2022). How effective is task-based language teaching to enhance second language learning? A technical comment on Bryfonski and McKay (2019). *Language Teaching Research*, doi:10.1177/13621688221131127

Yang, M., Cooc, N., & Sheng, L. (2017). An investigation of cross-linguistic transfer between Chinese and English: A meta-analysis. *Asian-Pacific Journal of Second and Foreign Language Education*, *2*(15), doi:10.1186/s40862-017-0036-9

Zhang, X. (2019). Foreign language anxiety and foreign language performance: A meta-analysis. *Modern Language Journal*, *103*(4), 763–781. doi:10.1111/modl.12590

Talia Isaacs is Associate Professor of Applied Linguistics and TESOL at IOE–UCL's Faculty of Education and Society, University College London and Co-Editor of *Language Testing*. Her research centres on language assessment, second language speaking, language teaching, validity, and health communication. She is an open science proponent, having won a 2023 UCL Open Science and Scholarship Award for the category Open resources, Publishing, and Textbooks. Alongside her language-focused activities, she brings applied linguistics expertise to the clinical trials community, having previously co-led the Communications theme of the MRC-NIHR (Medical Research Council-National Institute for Health and Care Research) Trials for Methodology Research Partnership Trial Conduct Working Group.

Hamish Chalmers is Lecturer in Applied Linguistics and Second Language Acquisition at the University of Oxford. A former primary school teacher in UK state schools and international schools overseas, his primary research interests centre on evaluations of pedagogical approaches to teaching children who use English as an Additional Language (EAL). His methodological interests include randomised trials, systematic reviews, and user engagement in research. He is co-director of the University of Oxford Education Deanery, an organisation dedicated to empowering educators worldwide to understand, use, and co-produce high-quality research evidence in education, and editor of the International Database of Education Systematic Reviews (IDESR.org).