

A Primer on Qualitative Research Synthesis in TESOL

SIN WANG CHONG 

*Queen's University Belfast
Belfast, UK*

LUKE PLONSKY

*Northern Arizona University
Flagstaff, USA*

Abstract

Secondary research in the form of literature reviews facilitates consolidation and transfer of knowledge. In the field of TESOL, the majority of secondary research is conducted in the form of narrative reviews, which rely on the researchers' selection and interpretation of primary studies and findings. Systematic reviews, which can be broadly categorized into meta-analysis (focusing on quantitative data) and qualitative research synthesis (focusing on qualitative data), are gaining popularity (see Plonsky, 2017) but are still less common. In particular, qualitative data collected from language classrooms, which are often criticized because of their lack of generalizability, are seldom synthesized in a systematic fashion. Against this backdrop, this article first attempts to make a case for conducting qualitative research synthesis in the field of TESOL. Second, this article provides a methodological framework and an example of how qualitative research synthesis can be conducted. The article closes with recommendations to promote qualitative research synthesis in the field of TESOL.

doi: 10.1002/tesq.3030

THE PLACE AND VALUE OF (QUALITATIVE) RESEARCH SYNTHESIS IN TESOL

Research synthesis comprises a set of well-developed techniques that allow researchers to overcome many of the pitfalls inherent to traditional, "stand-alone" literature reviews (see Ellis, 2015 for a compar-

Correspondence Email: S.Chong@qub.ac.uk

ison between research syntheses and traditional literature reviews). Foremost among its distinguishing features is systematicity. This feature manifests itself throughout the synthetic process, from (a) the exhaustive search for primary studies to (b) the application of a principled set of eligibility criteria and (c) the coding scheme applied to the final sample as a means to extract data relevant to the question. By thoroughly describing the steps and decision points along the way, readers can evaluate the choices being made and can more readily identify the potential issues and problems related to study findings (Norris & Ortega, 2007; Oswald & Plonsky, 2010). This approach reduces researcher bias and fosters greater confidence in the outcomes of the review; it also allows for replication at the secondary level (i.e., meta-analytic replication; see Plonsky, 2012).

Given these and other significant benefits, it is perhaps not surprising that researchers in TESOL, applied linguistics, and closely related fields (e.g., CALL, instructed second language acquisition; see Plonsky, 2017; Plonsky & Ziegler, 2016) have turned increasingly in recent years to research synthesis as a preferred means of taking stock of different domains. Figure 1 shows this growth of synthetic research in the realm of TESOL since 2000. The vast majority of these works have taken the form of meta-analysis. Meta-analysis can be a very powerful tool when the primary research of interest is quantitative in nature and provides a relevant effect size index (e.g., Cohen’s *d*, Pearson’s *r*) or sufficient data for the meta-analyst to calculate one. However, it is by no means the only type of synthetic research nor is it applicable to all domains and data types. Others include, for example, scoping review, historical

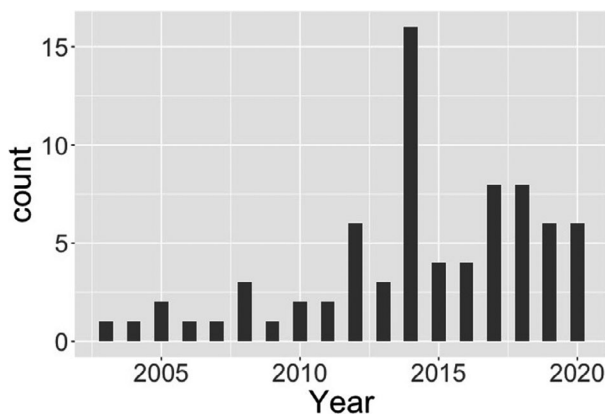


FIGURE 1. Research syntheses in TESOL (2000–present)Notes. Values based on citation data in Plonsky’s (n.d.) bibliography of research synthesis and meta-analysis in applied linguistics. The high number of studies in 2014 is due in part to a special issue of *English Teaching and Learning* that focused on meta-analysis.

review, and methodological synthesis (e.g., Gurzynski-Weiss & Plonsky, 2017; Marsden, Thompson, & Plonsky, 2018; Müller, Howard, Wilson, Gibson, & Katsos, in press). This article introduces to researchers in TESOL an additional and, in our view, underutilized approach to synthesizing previous research in a given domain: qualitative research synthesis (QRS).

QRS is a type of research synthesis widely used in medical research to report evidence-based practices, with a particular emphasis on the experiences of practitioners and patients. QRS can be applied to primary studies that use both qualitative and mixed research methods. QRS can be combined with meta-analysis for the latter, resulting in a mixed review (e.g., Jackson & Suethanapornkul, 2013; Tullock & Ortega, 2017), which refers to a single manuscript, reporting syntheses of both qualitative and quantitative findings. An alternative approach, as demonstrated in a recent QRS example (Chong & Reinders, 2020), is to focus only on the qualitative data of the mixed-methods studies involved.

In recent years, the use of QRS in the field of education has begun to grow. In TESOL, there are only a handful of QRSs. However, we believe that QRS is a useful method to aggregate qualitative findings of naturalistic, classroom-based studies, which are often criticized because of their lack of generalizability. Particularly, QRS in TESOL has more potential to systematically summarize qualitative findings in small-scale studies conducted by practitioners (e.g., action research, exploratory research, appreciative inquiry) to promote reflective teaching and evidence-based innovations. By synthesizing perceptions, beliefs, and experiences of teachers and learners in various educational milieus on a common topic, QRS offers a more comprehensive view on how a particular pedagogical intervention is implemented and experienced. It is especially valuable to practitioners and policy-makers who are looking for evidence-based pedagogical ideas to address new challenges (e.g., COVID-19). Complementing meta-analyses that often focus somewhat narrowly on the effectiveness of pedagogical interventions, QRS unveils a more holistic view of the factors associated with the instructional effectiveness. Because visualization techniques are usually used to report the synthesized qualitative data in a reader-friendly way (see, e.g., Çiftçi et al., 2018), there is also untapped potential for QRS to reach audiences beyond academia (e.g., teachers), and thus, facilitate research-pedagogy dialogue (Chong, 2020).

A METHODOLOGICAL FRAMEWORK

As a type of systematic review, QRS is conducted in a structured manner. That is, each step is undertaken systematically, usually

involving multiple reviewers, and laid out transparently to reduce bias and to maximally inform readers (Macaro, 2020). This section illustrates the methodological steps of conducting QRS in TESOL. By way of illustration, we refer to a recent synthesis by the first author on 16 technology-mediated task-based language teaching (TBLT) studies (Chong & Reinders, 2020) (Figure 2). Our discussion throughout will focus on the distinguishing features of QRS as a means to synthesize qualitative data in primary studies.

Formulate research questions. Because QRS is a type of secondary research, like any research, its design and process are guided by carefully conceived research questions. Qualitative synthesists usually set questions related to features of classroom practices, or students' and teachers' perceptions of pedagogical interventions. For instance, in Chong and Reinders (2020), the synthesis is guided by the following research questions:

1. What are the characteristics of technology-mediated tasks in the primary studies?
2. What are the affordances and limitations of technology-mediated tasks reported in these studies?
3. What are other emergent themes resulting from the grounded theory analysis?

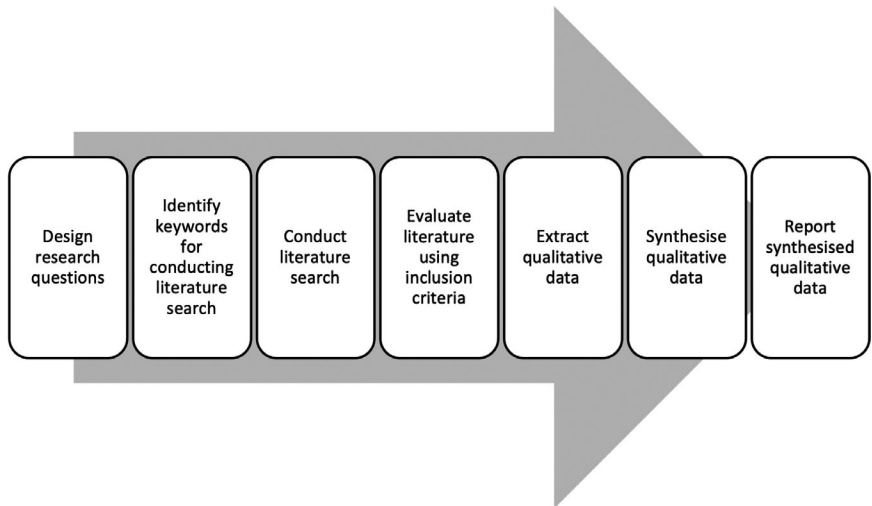


FIGURE 2. A methodological framework for conducting qualitative research synthesis in TESOL (Chong & Reinders, 2020)

Reminiscent of other qualitative studies which follow an interpretivist paradigm of research, QRS is designed to answer questions that aim to unravel complexities of phenomena in naturalistic classroom settings (e.g., the features of technology-mediated tasks) and offer rich descriptions of viewpoints and beliefs of different stakeholders (e.g., students' and teachers' beliefs toward technology-mediated TBLT).

Identify keywords. The second step is to devise an appropriate literature search strategy. At this stage, the research team have to agree on the search terms to be used and the databases to conduct the search. A number of search terms are developed, which can be searched alone (e.g., “task-based language teaching”) or in combination using Boolean Operators (e.g., “task-based language teaching” AND “technology”). Occasionally, interchangeable terms can be included (e.g., “task-based instruction”). Developing an appropriate set of keywords can be much more difficult than anticipated; the synthesist must balance substantive and methodological inclusiveness with the practical constraints of a potentially very large number of search “hits.” Some review teams involve librarians or experts in library science in the process of developing a list of search terms to maximize search results (Swinkels, Briddon, & Hall, 2006).

Conduct literature search. Three common avenues for conducting a search for primary literature include digital databases (e.g., Web of Science), journal websites (e.g., *TESOL Quarterly*), and the World Wide Web (e.g., Google Scholar). We also recommend searching the reference list of the included studies (snowballing) and contacting authors of the included studies for suggestions (outsourcing). The decision on where to search for literature is contingent on a number of factors, namely time frame, language, types of publication (e.g., whether you want to include less accessible sources such as conference proceedings, technical reports, and theses; i.e., gray literature). A more inclusive approach is generally preferred to limit bias and gain a more comprehensive view of the domain (see Plonsky & Brown, 2015). In Chong and Reinders (2020), because we wanted the synthesis to be as representative as possible, we searched on digital databases, journal websites, and Google Scholar. The literature search yielded a total of 99 publications.

Evaluate literature using inclusion criteria. One stage that distinguishes QRS from traditional narrative review is the formal appraisal of candidate studies. To arrive at a comprehensive sample, the reviewers must screen the studies obtained through the various search techniques using a common set of inclusion criteria. Then, they meet to

resolve disagreements. Two appraisal mechanisms are usually in place. First, appraisal focuses on the studies' compatibility with the nature, the topic, and the scope of the synthesis. The five inclusion criteria in our example study were as follows (Chong & Reinders, 2020, p. 5-6):

1. The articles report primary research (commentaries and reviews were not included).
2. The articles were published between 1997 and 2017.
3. The articles include at least one type of technology and adopt a well-defined conceptual or theoretical framework of TBLT.
4. The studies were conducted in second/foreign language classrooms.
5. The articles adopted either a qualitative or mixed-methods research design with a significant qualitative component to the research.

Second, reviewers appraise the research rigor of the primary studies. For our synthesis on technology-mediated TBLT, one of the inclusion criteria was that "The qualitative analysis of the articles follows... the *Qualitative Research Guidelines of Journals of Language Learning and Technology* and *TESOL Quarterly*. In particular, there should be inclusion of some raw data (e.g., transcribed verbatim of student interviews) when authors describe and discuss qualitative findings" (p. 6). After the appraisal of the 99 studies, 16 studies were included. Sometimes, it is challenging to evaluate the rigor of qualitative studies because guidelines are more open to interpretation than for quantitative research; therefore, it is important for the review team to reach a consensus regarding the standards they use for benchmarking (e.g., the *TESOL Quarterly* guidelines). To document the process of appraisal of studies and to uphold the principle of transparency, reviewers are recommended to adopt the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram (Moher et al., 2009).

Extract qualitative data. To handle a large amount of qualitative data, data extraction is often performed using reference management software (e.g., EndNote, RefWorks), research software (e.g., NVivo), and/or review managers (e.g., Covidence, Microsoft Access, RevMan). This stage differs from quantitative forms of research synthesis, which usually employ spreadsheet-based coding schemes. In Chong and Reinders (2020), NVivo was used to extract qualitative findings in the included studies and to ensure effective data management. Similar to screening of studies, data extraction is usually first performed independently by reviewers using a form or checklist before reaching an

agreement. This form or checklist includes items that can be related to PICO (population, intervention, comparator group, outcome), theoretical framework, and findings, or it could be designed to suit the purpose of the synthesis.

One major decision when extracting qualitative data is whether to extract only “raw data” or “interpreted findings,” or both. “Raw data” in qualitative studies refer to verbatim transcripts of interviewees and artifacts used by research participants, as opposed to “interpreted findings” (i.e., researcher discussions and interpretations without the support of raw data). Although there is not a consensus regarding the preferred type of qualitative data (or whether “interpreted findings” constitute “data”) to be extracted for synthesis, qualitative synthesists are advised to adopt a consistent approach and to report and justify their decision (Major & Savin-Baden, 2010). For our QRS on technology-mediated TBLT, we extracted both raw data and interpreted findings because of the relatively small number of studies included in our synthesis.

Synthesize qualitative data. This is the most crucial methodological stage in conducting a QRS. Some methods of synthesis used in QRS include grounded theory, thematic analysis, narrative synthesis, meta-ethnography, meta-summary (for a comprehensive discussion on these methods, see Dixon-Woods, Booth, & Sutton, 2007; Booth et al., 2016). These methods adopt either an inductive or deductive approach to data analysis. For example, an inductive approach was adopted in Chong and Reinders (2020) using grounded theory. Using a constant comparison method, descriptive and conceptual categories were generated through initial coding, focused coding, and axial coding (Charmaz, 2014). Ultimately, a meta-theory was proposed based on the emergent categories (for another example of using grounded theory to conduct QRS, see Chen (2016)). Alternatively, a deductive approach can be used. In this case, qualitative data are coded with reference to a conceptual or theoretical framework (e.g., metacognition).

Using Chong and Reinders (2020) as an example, we will illustrate how grounded theory can be used to synthesize qualitative data (Figure 3).

Grounded theorists follow three stages of data analysis to generate emergent themes: *initial coding*, *focused coding*, and *axial coding*¹. It is worth mentioning that the three stages of coding are iterative, and a constant comparison method is employed, made possible by using memo writing techniques (the researcher’s personal notes on the analytical process). Here, too, the process of QRS differs substantially

¹ For a detailed description of these coding stages, see Charmaz (2014).

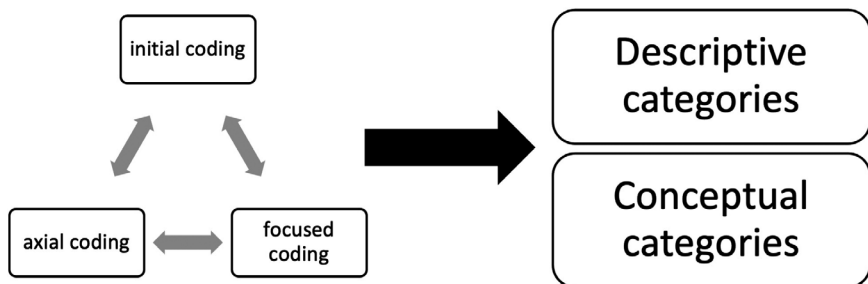


FIGURE 3. Using grounded theory to synthesize qualitative data

from that of other synthetic approaches such as meta-analysis, which are much more quantitatively oriented and less iterative. An example of the qualitative data synthesis process using grounded theory is provided as an online supplementary material (Appendix A).

Report synthesized qualitative data. Considerations on appropriate ways to report synthesized data are as important as the synthesis process itself. On the one hand, qualitative synthesists need to present synthesized findings in a reader-friendly and accessible manner, which usually suggests the adoption of a *thematic-narrative approach* to reporting. This involves categorizing synthesized findings thematically to respond to the research questions. When discussing the aggregated findings, a narrative approach is often preferred, which excludes quotations from primary studies.

This method of reporting synthesized qualitative data is not without problems. First, a thematic-narrative approach to reporting may be challenged by journal editors and readers regarding the trustworthiness of the data reported. Therefore, it is recommended that an *evidence-based approach* to reporting synthesized qualitative data is used, which substantiates claims without disrupting the flow of the narrative. An evidence-based approach to reporting synthesized qualitative data refers to the presentation of the documented data synthesis process. This can be achieved in a number of ways. First, a detailed coding scheme including conceptual categories, descriptive categories, the number of studies endorsing the categories, and an example code from the primary studies can be included (Appendix A). This elaborate coding scheme not only provides information, which enhances credibility of the synthesized data, but also includes new insights into the gravity of each category. For instance, by including the number of studies endorsing a category, it gives readers an impression of how prevalent a category is in the synthesized literature. It also enables synthesists to rank the categories according to importance. Second, a data

synthesis map can be included to illustrate the relationships between the coded categories (for an excellent example, see Figure 4 in Çiftçi et al. (2018)). The advantages of including a data synthesis map are twofold: (1) it summarizes the synthesized findings in a diagrammatic manner, which helps readers distil the information and (2) it gives readers confidence that the synthesists have put effort into collating qualitative data in a structural and systematic manner.

The way forward. The purpose of this article is to make a case for an underutilized type of systematic review, QRS, in the field of TESOL. Toward that end, we have described the process, presenting an example along the way to showcase how QRS can be conducted. At the outset, we argued that there is potential for QRS to play a vital role in efforts to synthesize qualitative evidence to inform policy and practice of language education. Echoing Norris and Ortega's (2007) call to develop a "synthetic mindset" (p. 812), research synthesis, especially QRS, is worthy of greater consideration and application in the field of TESOL. In fact, we are seeing some exciting developments that address methodological issues in synthetic studies. For instance, *Language Learning* now accepts methodological review and systematic review submissions. More can be done, for example, through the publication of QRS protocols by leading journals of our field. Moreover, interdisciplinary dialogues with research fields that possess a more well-established tradition of conducting research synthesis (e.g., medicine, healthcare) would benefit researchers in TESOL as well.

THE AUTHOR

Sin Wang Chong (PhD) is Lecturer (Assistant Professor) in TESOL at Queen's University Belfast in the United Kingdom. His research interests include language assessment, educational assessment, educational technology, and research synthesis methodologies. He is Associate Editor of *Innovation in Language Learning and Teaching* and *Higher Education Research and Development*.

Luke Plonsky (PhD Michigan State) is Associate Professor of Applied Linguistics at Northern Arizona University. His work, focusing primarily on SLA and research methods, has appeared in more than 90 articles, book chapters, and books. Luke is Associate Editor of *SSLA*, Managing Editor of *Foreign Language Annals*, and Co-Director of IRIS.

ACKNOWLEDGMENTS

We would like to thank the editor and the two anonymous reviewers who took the time to read our manuscript and offered very constructive feedback.

CONTRIBUTION & FUNDING STATEMENT

This submission is not supported by any funding.

REFERENCES

- Booth, A., Noyes, J., Flemming, K., Gerhaidus, A., Wahlster, P., van der Wilt, G., ... Rehfuss, E. (2016). *Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions*. Retrieved from <http://www.integrate-hta.eu/downloads/>
- Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). Thousand Oaks, CA: SAGE.
- Chen, T. (2016). Technology-supported peer feedback in S/FL writing classes: A research synthesis. *Computer Assisted Language Learning*, 29(2), 365–397. <https://doi.org/10.1080/09588221.2014.960942>
- Chong, S. W. (2020). The role of research synthesis in promoting research-pedagogy dialogue. *ELT Journal*, 74(4), 484–487. <https://doi.org/10.1093/elt/ccaa046>
- Chong, S. W., & Reinders, H. (2020). Technology-mediated task-based language teaching: A qualitative research synthesis. *Language Learning & Technology*, 24(3), 70–86. <http://hdl.handle.net/10125/44739>
- Çiftçi, E. Y., & Savaş, P. (2018). The role of telecollaboration in language and intercultural learning: A synthesis of studies published between 2010 and 2015. *ReCALL*, 30(3), 278–298. <https://doi.org/10.1017/s0958344017000313>
- Dixon-Woods, M., Booth, A., & Sutton, A. J. (2007). Synthesizing qualitative research: A review of published reports. *Qualitative Research*, 7(3), 375–422. <https://doi.org/10.1177/1468794107078517>
- Ellis, R. (2015). Introduction: Complementarity in research syntheses. *Applied Linguistics*, 36(3), 285–289. <https://doi.org/10.1093/applin/amv015>
- Gurzynski-Weiss, L., & Plonsky, L. (2017). Look who's interacting: A scoping review of research involving non-teacher/non-peer interlocutors. In L. Gurzynski-Weiss (Ed.), *Expanding individual difference research in the interaction approach: Investigating learners, instructors, and other interlocutors* (pp. 305–324). Philadelphia, PA: John Benjamins.
- Jackson, D. O., & Suethanapornkul, S. (2013). The cognition hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning*, 63(2), 330–367. <https://doi.org/10.1111/lang.12008>
- Major, C. H., & Savin-Baden, M. (2010). *An introduction to qualitative research synthesis: Managing the information explosion in social science research*. London: Routledge.
- Marcaro, E. (2020). Systematic reviews in applied linguistics. In J. McKinley & H. Rose (Eds.), *The Routledge Handbook of Research Methods in Applied Linguistics* (pp. 230–239). London: Routledge.
- Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, 39, 861–904. <https://doi.org/10.1017/s0142716418000036>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.7717/peerj.4598/supp-5>

- Müller, L.-M., Howard, K., Wilson, E., Gibson, J., & Katsos, N. (in press). Bilingualism in the family and child well-being: A scoping review. *International Journal of Bilingualism*. <https://doi.org/10.1177%2F1367006920920939>
- Norris, J. M., & Ortega, L. (2007). The future of research synthesis in applied linguistics: Beyond art or science. *TESOL Quarterly*, *41*(4), 805–815. <https://doi.org/10.1002/j.1545-7249.2007.tb00105.x>
- Oswald, F. L., & Plonsky, L. (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics*, *30*, 85–110. <https://doi.org/10.1017/s0267190510000115>
- Plonsky, L. (2012). Replication, meta-analysis, and generalizability. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 116–132). New York: Cambridge University Press.
- Plonsky, L. (2017). Quantitative research methods. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 505–521). New York, NY: Routledge.
- Plonsky, L., & Brown, D. (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*(2), 267–278. <https://doi.org/10.1177/0267658314536436>
- Plonsky, L., & Ziegler, N. (2016). The CALL-SLA interface: Insights from a second-order synthesis. *Language Learning & Technology*, *20*, 17–37. <https://doi.org/10125/44459>
- Swinkels, A., Briddon, J., & Hall, J. (2006). Two physiotherapists, one librarian and a systematic literature review: Collaboration in action. *Health Information and Libraries Journal*, *23*(4), 248–256. <https://doi.org/10.1111/j.1471-1842.2006.00689.x>
- Tulloch, B., & Ortega, L. (2017). Fluency and multilingualism in study abroad: Lessons from a scoping review. *System*, *71*, 7–21. <https://doi.org/10.1016/j.system.2017.09.019>

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Supplementary Material