A methodological review of qualitative research syntheses in CALL: The state-of-the-art[1]

Sin Wang Chong, Queen's University Belfast, UK
Hayo Reinders, KMUTT, Thailand; Anaheim University, USA

Abstract

Qualitative research in computer-assisted language learning (CALL) has received more attention recently, most evidently from the publications of two special issues on qualitative research in CALL by *CALICO Journal* (2015) and *Language Learning & Technology* (2018). The inherent strength of qualitative studies in CALL is that they celebrate depth of interpretation by focusing on specific contexts and a small number of participants. Yet this strength can sometimes be perceived as a limitation when transferability and applicability of findings are considered. However, this potential shortcoming can be overcome when findings of multiple studies are meaningfully synthesised, for example, in a qualitative research synthesis (QRS). QRS is a relatively new form of research synthesis in CALL which systematically summarises qualitative findings guided by research questions and a predetermined research protocol. To report current methodological practices and identify areas of strengths and weaknesses, this exploratory review analyses the methodological sections of 16 QRS in CALL using a recent, CALL-based QRS methodological framework comprising seven stages (Author 1 et al., 2021). Findings reported include implementation of the seven methodological stages, and convergent and divergent methodological practices. Suggestions, especially about qualitative data extraction and synthesis, are made in light of the findings.

Keywords: *CALL; qualitative research; research synthesis; research*

---

1. Qualitative research and research synthesis in computer-assisted language learning (CALL)

Selecting an appropriate research methodology is regarded as one of the three main challenges in CALL research (Levy et al., 2015). Methodological challenges in CALL research stem in part from the fact that CALL is transdisciplinary and draws on a plethora of methodological traditions and practices. Among the various research methodologies, qualitative research and research synthesis are relatively recent yet important additions. Qualitative research has been increasingly valued in CALL. For example, *Language Learning & Technology* published a special issue on qualitative studies in CALL in 2018 (edited by Mike Levy and Paul J. Moore). Earlier, a special issue on the same topic was also published in *CALICO Journal* in 2015 (edited by Ursula Stickler and Regine Hampel). In their editorial, Levy and Moore (2018) identified two overarching aims of qualitative research in CALL: "to help clarify and detail the contextual factors—from macro to micro—that impact upon the success (or otherwise) of CALL implementations" and "to inform the design, development, and evaluation of new CALL artefacts (language learning apps, software, learning systems, CALL tasks) through a deeper understanding of the user's perspective and the dynamic, moment-by-moment processes of CALL task completion" (2018, p. 1). Qualitative research contributes to CALL research by providing in-depth accounts of the implementation process of educational technologies both inside and outside the language classroom. It also sheds light on individual and contextual factors which shape technology-mediated interactions (Levy & Moore, 2018). Equally important, qualitative studies in CALL place language learners in the limelight, enabling their voices to be heard regarding their experiences in learning a language with new technologies (Levy, 2015). Stickler and Hampel (2015) point to the need for CALL research to embrace "cultural

relativism" in order for studies to uphold ecological validity (p. 381), which can be achieved through qualitative studies.

Research synthesis, which refers to a stand-alone systematic literature review, is a type of secondary research which is increasingly prominent in applied linguistics and language education research, especially in terms of the synthesis of quantitative data in meta-analysis (Plonsky & Ziegler, 2016). Other types of research synthesis noted include bibliometric review (e.g., Gong et al., 2018), methodological review (e.g., Plonsky & Kim, 2016), scoping review (e.g., Jiang et al., 2020), systematic literature review (e.g., Ma et al., 2017), and mixed review (e.g., Tullock & Ortega, 2017)[2]. The distinction between research synthesis and literature review is that the former employs a set of systematic methodological protocols which guide literature search and analysis (Ellis, 2015). In this sense, research synthesis can be compared with primary studies because of its emphasis on systematicity, transparency, reliability, and replicability.

In CALL, research synthesis is a new form of research. Following a research tradition more inclined towards positivist, quantitative research, the most common type of research synthesis is meta-analysis (Lin, 2015). Another kind of research synthesis, qualitative research synthesis (QRS), referring to a repertoire of techniques used to systematically synthesise qualitative data, is likely to be the least known to the CALL community. To date, there are but a handful of QRS published in the past decades (a comprehensive search for the present methodological review resulted in only 16 QRS in CALL (See Section 2). QRS has its origin in medical research where it is used to synthesise clinical experiences of patients.

---

[2] See a presentation by Author 1 in [month, year] at [name of university removed] (link removed for peer review), which drew on a paper under review by Author 1 et al., for the 13 types of secondary research in applied linguistics and TESOL (see preprint here [link removed for peer review]).

QRS contributes to the research base by "bringing together the rich and detailed findings of qualitative research studies and thus to offer a new interpretation of a research question" (Cooke et al., 2012, p. 1435).

Applied to CALL, QRS is useful "to aggregate qualitative findings of naturalistic, classroom-based studies, which are often criticised because of their lack of generalisability" (Author 1 et al., 2021, p. 3). It is especially suitable for summarising practitioner research, including action research, appreciative inquiry, and exploratory research, which investigate small numbers of participants, innovative practices, and unique educational contexts (Author 1, 2020). Additionally, QRS can be combined with other forms of research synthesis. For example, qualitative evidence can be synthesised alongside a meta-analysis which aims to examine the effectiveness of a particular pedagogical intervention both sequentially and convergently, resulting in a mixed review (Noyes et al., 2019). While QRS is a useful form of research synthesis and one that can inform practice and policy, it involves an extremely complex methodological process and there is a lack of explicit and consensual view towards the acceptable standard of QRS. For instance, published QRS in other disciplines are found to be weak in reporting processes and decision-making related to literature search, study appraisal, and data synthesis (Hannes & Macaitis, 2012). The lack of QRS appraisal tools can be attributed to the fact that QRS does not refer to a single methodology. In fact, QRS methodologies are highly diversified; for example, Flemming and Noyes (2021) noted that there are over 30 QRS methodologies, making the assessment of QRS quality more difficult. One of the most widely used QRS appraisal tools is the GRADE-CERQual ('Confidence in the Evidence from Reviews of Qualitative research') approach (Lewin et al., 2018) which assesses quality of QRS from five perspectives: (1) methodological limitations, (2) coherence, (3) adequacy of data, (4) relevance, and (5) dissemination/publication bias (p. 1).

In the field of CALL, QRS is a budding research genre, and more knowledge needs to be garnered to develop a better understanding of its methodological practices and standards. With this in mind, the current methodological review analyses 16 QRS in CALL to review the state-of-the-art of QRS methodologies and identify areas of strengths and weaknesses. It is important to note that this review employs an exploratory and inductive approach to unravelling standards and expectations for conducting QRS in CALL. In other words, we do not begin with a top-down appraisal tool to assess the quality of QRS in CALL because, as noted above and later in the analysis, QRS encompasses a variety of methodological approaches across and within disciplines. Instead, what we attempt to do is to report existing methodological practices systematically using a recent, CALL-based generic QRS methodological framework (Author 1 et al., 2021). From there, we identify convergent and divergent practices. Based on these observations, we propose some baseline and preferred methodological practices for conducting QRS in CALL.

2. Methodology

A methodological review is a type of systematic secondary research (i.e., research synthesis) which focuses on summarising the state-of-the-art methodological practices of research in a substantive field or topic (e.g., Plonsky & Kim, 2016). In this methodological review, the focus is on the methodological practices of QRS in CALL. In other words, this methodological review is a "review of reviews" or a "meta-review". Focusing our analysis on the methodology section of QRS in CALL, this methodological review aims to answer the following research questions:

1. What are the methodological stages reported in QRS in CALL?
2. What methodological literature do QRS in CALL draw upon?

2.1 Analytical framework

The analysis of methodological stages of QRS in CALL is guided by the QRS framework proposed by Author 1 et al. (2021) published in *TESOL Quarterly*. The decision to employ this methodological framework is twofold. First, to the best of our knowledge, this is the first framework on QRS in TESOL and language education. Second, this methodological framework is developed based on recent QRS practices in CALL, exemplified in a QRS example on technology-mediated task-based language teaching (Author 1 & Author 2, 2020). The methodological framework is introduced in Figure 1.

[**Figure 1.** A methodological framework for conducting qualitative research synthesis in CALL (Author 1 et al., 2021)]

Design research questions: A QRS, like primary studies, is guided by a number of research questions. Adhering to the nature of qualitative research, these research questions tend to focus on experiences, perceptions, and beliefs.

Identify keywords for conducting literature search: A search protocol needs to be developed which comprises search strategy (i.e., which databases to search) and search string(s) (i.e., keywords used to search for articles).

Conduct literature search: At this stage, researchers implement the search protocol. This is often done by more than one researcher to enable comparison of search results.

Evaluate literature using inclusion criteria: This stage concerns the evaluation of searched literature, a quality assurance mechanism to ensure the included literature is relevant to the scope and focus of the review. Some QRS also include exclusion criteria. Inclusion and/or exclusion criteria are used to screen the searched literature. Usually, two levels of

screening are employed: first-level screening which focuses on titles and abstracts, and second-level screening which includes full texts. Second-level screening is usually performed on articles when their eligibility remains unclear after title and abstract screening.

Extract qualitative data: Researchers develop a data extraction form based on the research questions and extract relevant information from the included articles. It is important that researchers do not alter wordings of the information, to enable comparison between extraction results between researchers. Moreover, it is advised that sources of the extracted information (e.g., paragraph number, page number) be retained so that the information can be reviewed quickly.

Synthesise qualitative data: Resembling the later stage of qualitative coding, the extracted qualitative data are collated into themes and sub-themes. Qualitative data synthesis, like qualitative data analysis, is guided by specific methodological approaches including thematic analysis, grounded theory, meta-ethnography.

Report synthesised qualitative data: This concerns the writing up of the findings and discussion section of the QRS. This is usually done thematically, with researchers presenting each theme substantiated by synthesised evidence. Synthesised evidence includes at least two types: *macro* and *micro*. Macro-evidence presents the "big picture" of the synthesised results, for example, by reporting the percentages or numbers of included studies endorsing a particular theme. Micro-evidence describes representative and/or interesting examples of a theme. Unlike primary qualitative studies, it is important for QRS, which aims to synthesise state-of-the-art, to focus on macro-evidence.

2.2 Search protocol of the present review

The search for QRS in CALL was conducted on 16 September 2020 on five major international refereed journals in CALL: *CALICO*, *Computer Assisted Language Learning*,

Language Learning & Technology, *System*, and *ReCALL*. The keyword used was

("qualitative research synthesis"). Although we decided to focus on the five specialist

journals dedicated to CALL in our search, we were aware that a significant number of CALL

studies are published in journals in the fields of language education, second language

acquisition, and applied linguistics. Therefore, when second-tier search was conducted, we

employed a snowballing technique, that is, we referred to reference lists of the QRS identified

to locate additional QRS published in non-CALL journals. This resulted in two non-journal

publications (a book chapter and a report) and two QRS published in non-CALL journals.

A search result comprising 496 articles was reviewed and articles were included for

further analysis if they reported research syntheses of qualitative data (syntheses reporting

quantitative data were excluded). Moreover, it is important to bear in mind that QRS is a type

of systematic secondary research, so traditional secondary research types (e.g., narrative

reviews) were also excluded. Since this meta-review focuses on methodologies of QRS, there

is no limitation on their substantive topics as long as they are CALL-related. After screening

the 496 entries, nine entries were found to match the focus of this meta-review (see Table 1

for the breakdown of search results). From the nine entries, a second-tier search was

conducted using a snowballing technique, resulting in a total of 16 QRS. The rationale for

including a second-tier search was that many QRS did not include "qualitative research

synthesis" in their titles and texts. The snowballing process enabled us to identify these

additional QRS. For example, Debski (2003) is a QRS with the title "Analysis of research in

CALL (1980-2000) with a reflection on CALL as an academic discipline". The search

process is summarised in Figure 2 in a Preferred Reporting Items for Systematic Reviews and

Meta-Analyses (PRISMA) 2020 diagram (Page et al., 2020); bibliographical information of

the included QRS is presented in Appendix 1.

[**Table 1.** Breakdown of the initial search results (n=496)]

[**Figure 2.** PRISMA diagram showing the flow of literature search]

2.3 Length of methodology section in QRS

Table 2 collates information about the lengths of the methodology sections of the 16 QRS in CALL. In these 16 reviews, the number of words authors dedicated to the methodology section ranges from 293 to 2312. In terms of percentages, the number of words on methodology accounts for 3% to 17% of the total article (SD = 3.9; Mean = 8.9%). Categorising the 16 QRS suggests a number of factors which potentially affect the length of the methodology section. First, ordering the QRS according to years of publication, Figure 3 reveals an emergent pattern that QRS published after 2017 tend to include a lengthier methodology section. When the 16 QRS are categorised by publication types, by and large, the methodological sections of QRS published in academic journals (Mean = 8.8%) seem to be longer than those published as book chapters/reports (Mean = 6%), even though book chapters/reports are usually more lenient with word limits. One possible explanation is that articles published in international refereed journals are usually subject to more rigorous peer review and researchers are expected to provide a more thorough description of the methodological steps taken in the QRS process.

Table 3 shows the percentage of the words of the methodology section in relation to the total number of words of the QRS by journals. The means of the journals are: *CALICO* (11%), *Computer Assisted Language Learning* (8%), *Educational Technology & Society* (4%*), *Language Teaching* (3%*), *Language Learning & Technology* (17%*), *ReCALL* (9.25%), *System* (11%*[3]). The methodology sections of QRS published in *Language Learning & Technology*, *CALICO*, and *System* are the longest in relation to the total length of

---

[3] The asterisk indicates that there is only one QRS from the journal included in this review.

the articles. The two QRS published in *Language Teaching* and *Education Technology & Society* have the lowest percentages. Macaro et al. (2012) included a relatively short methodology section in comparison with the total length of the review for good reasons. First, this QRS was submitted to the "state-of-the-art review" section of *Language Teaching*, which is one of the lengthiest sections of the journal (the total length of this QRS is 18866 words). Second, the scope of the submissions to this "state-of-the-art" review section intends to be broad. Macaro et al. (2012) provided "an in-depth review of 47 post-2000 studies investigating the efficacy of technology in the teaching of L2 English" (p. 1). The number of studies included in this QRS is large when compared with other QRS in CALL (e.g., Author 1 & Author 2, 2020, reviewed a total of 16 studies; Chen, 2016, reviewed a total of 20 studies). It is understandable that the researchers devoted a large portion of the manuscript to reporting extensive findings. On the other hand, Ciftci (2016) was published in *Educational Technology & Society*, which is a high-impact, open-access journal in educational technology (Impact Factor 2019: 2.086). These two exceptional cases of QRS published in journals which have unusually short methodology sections are both published in journals outside of CALL (although the topics are CALL-related). Furthermore, QRS published in CALL journals consistently maintain a higher methodology/full-text word ratio than those published as book chapters. Information about lengths of methodology sections of the 16 QRS is placed here instead of as part of the findings because its intention is to provide an overview about the sections. It must be stressed that length of the section is not a reliable way to infer the quality of QRS methodology. For example, Shadiev et al. (2017) and Chen (2016) were both published in *Computer Assisted Language Learning*, a highly ranked CALL journal; while the percentage of the methodology section is 3 in Shadiev et al. (2017), it is 12 in Chen (2016). Moreover, some QRS extended their methodological discussions to other sections or

maintained a methodological reflexivity throughout their papers which cannot be accounted for by simply analysing the methodology section.

[**Table 2.** Length of methodology section]

[**Figure 3.** Number of words in the methodology section (%)]

[**Table 3.** Methodology/full-text word percentage by CALL journals]

3. Findings and discussion

3.1 Similarities in methodological stages in QRS in CALL

Based on the QRS methodological framework by Author 1 et al. (2021) and in response to research question 1, Table 4 identifies seven methodological stages for conducting QRS in CALL and the reported practices noted in the 16 QRS. Among the seven stages, four are consistently applied in all 16 QRS: *design research questions*, *identify scope of search*, *decide on inclusion and/or exclusion criteria*, and *evaluate literature using inclusion and/or exclusion criteria*.

[**Table 4.** Methodological stages reported in 16 QRS in CALL]

3.1.1 Designing research questions

All 16 QRS included research questions that guided the research synthesis, although all the research questions were mentioned at the end of the introduction or literature review, not in the methodology section. The number of research questions in each QRS ranged from one to six. These QRS most often included three research questions (n=7), followed by four (n=3), two (n=3), one (n=2), and six (n=1). These research questions can be categorised into four types: (1) research questions on technologies (n=16); (2) research questions on impact (n=10); (3) research questions on research (n=6); (4) open-ended research questions (n=5).

Research questions focusing on technologies cover three areas: (1a) types of technology used (e.g., Duman et al., 2014); (1b) characteristics of the technology used (e.g., Zhao, 2003); and (1c) justifications for using the technology (Stockwell, 2007). As for research questions on impact, three sub-themes were identified: (2a) usefulness and limitations of technological interventions (e.g., Author 1 & Author 2, 2020); (2b) impacts on language learning (Elgort, 2018; Sauro, 2011); (2c) type of evaluation tool (Elgort, 2018). The third type of research question deals with CALL research; it focuses on (3a) research methodologies (e.g., Chwo et al., 2018); (3b) research topics and trends (e.g., Debski, 2003); (3c) participants and contexts (e.g., Ciftci & Savas, 2018); (3d) theoretical frameworks (Debski, 2003; Duman et al., 2014); (3e) bibliographical information (Shadiev et al., 2017); (3f) duration of research (Ciftci, 2016); and (3g) future research directions (Ciftci, 2016). Lastly, some QRS included an open-ended research question which focuses on emerging themes during the synthesis process (e.g., "What are the main themes emerging from the GT analysis and what is the metatheory for the synthesis?" (Chen, 2016, p. 367)).

3.1.2 Identifying scope of search

Moreover, scope of the search was reported in all 16 QRS. This pertains to the identification of databases and/or journal websites on which the researchers conducted their search. Although there are other parameters (e.g., language of publication, year of publication, type of publication), this goes beyond the scope of the present review, which provides a bird's-eye view of QRS methodology. The number of databases used in a single QRS ranges from one to 11. The number of journals searched in a single QRS ranged from four to 33. One QRS also included conference proceedings in their search (Chen, 2016). Appendices 2 and 3 list the journals and databases searched in the 16 QRS respectively, ranked from the least-frequently searched to the most-frequently searched. In total, the 16

QRS cover 27 databases and 69 journals. When selecting which databases to conduct their search on, QRS researchers considered (1) disciplinary relevance (e.g., ERIC), (2) quality of publications (e.g., SSCI), (3) publication bias (e.g., System for Information on Grey Literature in Europe), (4) accessibility of publications (e.g., DOAJ), and (5) coverage of the databases (e.g., EBSCO, JSTOR). Regarding journal websites, QRS researchers consulted four types of journals: (1) flagship CALL journals (e.g., *ReCALL*), (2) language education journals (e.g., *TESOL Quarterly*), (3) applied linguistics journals (e.g., *Applied Linguistics*), (4) language learning and acquisition journals (e.g., *Studies in Second Language Acquisition*). In addition, there were attempts in the 16 QRS to mitigate publication bias by including not only international refereed journals, which mostly publish articles focusing on the Anglophone contexts, but also regional and local journals (e.g., *RELC Journal*, *Taiwan Journal of TESOL*).

3.1.3 Inclusion and/or exclusion criteria

The 16 QRS in CALL employed inclusion and/or exclusion criteria to shortlist studies relevant to their research questions. The number of inclusion criteria used ranged from three to nine. Usually, these QRS employed three inclusion criteria (n=7) (e.g., Zhao, 2003) followed by four (n=4) (e.g., Duman et al., 2014). One to nine exclusion criteria were used. When developing inclusion and exclusion criteria, the following areas were considered: (1) types of publication (n=14) (e.g., include only journal articles), (2) years of publication (n=13) (e.g., 2000-2012), (3) relevance to research questions (n=12) (e.g., use of technology in task-based language teaching), (4) rigour of studies (n=6) (e.g., peer-reviewed articles), (5) participants (n=5) (e.g., ESL/EFL learners), (6) contexts (n=4) (e.g., higher education), (7) research design (n=4) (e.g., qualitative studies), and (8) languages of publication (n=4) (e.g., English).

3.2 Divergences in methodological stages in QRS in CALL

3.2.1 Identifying keywords for conducting literature search

Among the 16 QRS in CALL, two did not report the keywords or search strings used to conduct the literature search (Debski, 2003; Stockwell, 2007) because their objective was to review broad research trends. Debski (2003) focused on all publications between 1980 and 2000 in seven CALL journals and Stockwell (2007) reviewed all publications in the four flagship CALL journals (*CALICO Journal*, *Computer Assisted Language Learning*, *Language Learning & Technology*, *ReCALL*) published from 2001 to 2005. Interestingly, Zhao (2003), who also analysed the general trends in CALL research, conducted the search based on one key term, "computer-assisted language learning". The remaining QRS (n=14) usually included more than one search term or string, most of which were keywords related to the substantive focus of the reviews (e.g., task-based language teaching in Author 1 and Author 2, 2020; mobile-assisted language learning in Chwo et al., 2018). Some of the included QRS made use of more fine-grained search strategies including Boolean operators (Elgort, 2018), truncation (Avgousti, 2018), or synonyms (Macaro e tal., 2012). The above seems to suggest that the inclusion of keywords or search strings is contingent on the scope and purpose of the QRS. For QRS which have more extensive focuses, keyword search is usually not employed. On the other hand, QRS aiming to review the state-of-the-art of a sub-field in CALL are more likely to include more sophisticated search strings.

3.2.2 Extracting qualitative data

Data extraction in QRS can be understood using qualitative data analysis terminology - it is similar to initial coding of qualitative data. What is different is that usually a more

deductive approach is employed in QRS data extraction through the development of an extraction form detailing the types of information to be coded. On the contrary, initial coding of qualitative data can take on a more inductive approach in which new codes are identified through line-by-line and repeated reading of data. Extraction of qualitative data needs to be reported in QRS for two reasons. First, it facilitates the comparison of data to be included in the analysis through, for example, calculating inter-rater reliability or discussing data extracted by different reviewers. Second, reporting data extraction clarifies what counts as qualitative data. For example, Author 1 & Author 2 (2020) mentioned that both transcribed verbatim and interpretations of qualitative findings were treated as "data" in their QRS.

13 of the included QRS reported how data extraction was performed, while three did not (Chwo et al., 2018; Sauro, 2011; Zhao, 2003). For the sake of illustration, Zhao (2003) reported extensively on the literature selection process but did not describe how relevant data were extracted from the shortlisted articles. For those QRS that reported qualitative data extraction, two approaches are noted: *tabular* and *textual*. Lewis and O'Dowd (2016) presented information extracted from the included articles in Table 2.1 including author, year, title, bibliographical details, and computer-mediated communication tools. Ciftci (2016) and Ciftci and Savas (2018) tabulated similar information in Appendix A. The most comprehensive data extraction table is found in Table 2 of Avgousti (2018), which covered 34 categories of data (e.g., participants' target language, country, research context), focusing mostly on research design. Appendix A of Debski (2003), which comprised 11 categories of data (e.g., theoretical foundation, research design, duration of data collection, research methods), is also an extensive list. While an example of the data extraction form was not included, Duman et al. (2014) included a dedicated paragraph under the heading (the data collection instrument) describing the processing of developing and validating the data

extraction form, amongst others, through consulting three experts. In a similar vein, Shadiev et al. (2017) did not attach a data extraction form but briefly described the major categories of data, including number of articles, research topic, technology, and research methodology (p. 288). Following the same descriptive approach, Stockwell (2007) included three sections (3.1-3.3) to discuss the types of data for inclusion, focusing on technologies and language skills, technologies and learning goals, and technological features.

Data extraction in the 13 QRS in CALL was mostly performed by one reviewer, while others were done by two (n=3) or four reviewers (n=1). Three of the QRS did not specify the number of reviewers involved in extracting qualitative data. The most common approach to resolving disagreement between coders is through discussions. Chen (2016), for example, described in detail how the author and the invited coder resolved disagreements. Information reported include duration and frequency of meetings, synthesis and research experience of coders, and activities conducted during the meetings. Calculation of inter-coder reliability, a method recommended by some research synthesis methodology texts (e.g., Author 1 et al., 2021; O'Connor & Joffe, 2020), was not reported in these 16 QRS. It appears that QRS researchers in CALL are inclined towards a more qualitative approach to establishing trustworthiness and credibility than a statistical one.

3.2.3 Synthesising qualitative data

Qualitative data synthesis, using the language of qualitative data analysis, resembles higher-order coding which combines more descriptive codes into more conceptual ones. This is the methodological stage where discrepancies are the most apparent in the 16 QRS in CALL - only nine of the included QRS reported synthesis of qualitative data. Synthesis of qualitative data was informed by various qualitative data analytical approaches, namely grounded theory (n=3) (e.g., Author 1 & Author 2, 2020), thematic analysis (n=2) (e.g.,

Chwo et al., 2018), open coding (n=1) (Ciftci, 2016), content analysis (n=1) (Dunman et al., 2014), narrative synthesis (n=1) (Lock & Andrews, 2004), and topic-specific framework (n=1) (Elgort, 2018 used the technology-mediated vocabulary development framework to synthesise qualitative data). Authors of these nine QRS were aware of the "messiness" in the definitions of various approaches to qualitative data analysis and that these approaches sometimes overlap with one another (e.g., open coding and grounded theory, content analysis and thematic analysis). To clarify the methodological traditions which informed their data synthesis process, QRS researchers provided operational definitions of the specific approach to data synthesis. Definitions were given not only in relation to the analytical approach but also the employment of relevant methodological jargon. For instance, data synthesis of Chwo et al. (2018) was informed by thematic analysis; terminologies associated with thematic analysis such as descriptive analysis and reflective analysis were mentioned (p. 64). In a similar fashion, Author 1 and Author 2 (2020) clearly defined grounded theory by referring to works by Charmaz (2006), and by defining related terms (e.g., axial coding, constant comparative method).

Moreover, we note that a few of the included QRS focus not only on the conceptual understanding of their data synthesis approaches but also on how the approaches are operationalised. This is achieved through illustrations using coding schemes (in the form of tables) and conceptual diagrams. For example, a coding scheme was included in Author 1 and Author 2 (2020) (Table 2) to illustrate the developments of initial codes, sub-categories, descriptive categories, and analytical categories. The table included examples of initial codes (extracted from included studies) and number of studies endorsing each category. Ciftci (2016), on the other hand, created a conceptual diagram near the end of the review (Figure 4) to illustrate how open coding was operationalised and how new themes were identified.

Compared to other methodological stages, data synthesis in QRS in CALL is the least-frequently enacted stage. However, it is the most distinctive feature of QRS, as opposed to the more quantitative type of research synthesis (meta-analysis) and the more descriptive types of research synthesis (e.g., scoping review, systematic literature review).The inherent difference between QRS and other types of research synthesis lies in the fact that QRS can "reveal deep insights into disparate literature for future research" (Chen, 2016, p. 387) through synthesising qualitative data into "a more abstract level in which multidimensions, varieties, and complexities are disclosed (Cifci & Savs, 2018, p. 281). It is only through higher-order qualitative data analysis, informed by carefully selected qualitative methodological traditions and approaches, that multidimensionality and complexity in QRS are upheld. In other words, we feel that there is a need for not only *objectifying* coding (itemising qualitative data) but also *abstracting* the coding (combining items into ideas or concepts) in QRS.

4. Methodological literature informing QRS in CALL

Addressing research question 2, this section focuses on the citation patterns of the literature that inform the methodologies of the 16 QRS in CALL. The analysis of citation patterns has become an increasingly important area in synthetic studies in applied linguistics and language education, especially in bibliometric reviews (e.g., Hyland & Jiang, 2021). Among the many focuses of tracking citation patterns, one is on highly-cited publications (e.g., Lei & Liu, 2019). Despite not being a bibliometric review, the present review encompasses the analysis of frequently cited publications in the methodology section of the 16 QRS. It is believed that the analysis of citation patterns can unravel traditions and sources

influencing the methodological decisions of QRS in CALL. It also provides an overview to CALL researchers who attempt to conduct QRS to consider available methodological texts.

Figure 4 shows the number of references cited in the methodology section of the 16 QRS. In general, it is observed that QRS published more recently cited more methodological references than those published earlier. For instance, Avgousti (2018) and Author 1 and Author 2 (2020), which were published in the past two to three years, are the two QRS which included the highest number of methodological references. On the other hand, Zhao (2003), Debski (2003), and Lock and Andrews (2004), which were published almost a decade ago, included the least number of citations. This demonstrates the heightened awareness of CALL researchers that QRS is unlike traditional secondary research (e.g., narrative review); rather it celebrates systematicity, replicability, and transparency (Ellis, 2015). The increasing number of methodological citations also indicates an attempt of CALL QRS to mirror acceptable standards in evidence synthesis, both within and beyond the discipline.

[**Figure 4.** Number of references related to QRS methodology]

A closer look at the citations in the methodology section of the 16 QRS reveals seven areas where CALL researchers feel the need to support their methodological decisions using literature: (1) introduction to QRS; (2) justification of number of studies; (3) scope/search strategy; (4) inclusion/exclusion criteria; (5) data extraction; (6) data synthesis; (7) methodological framework. Tables 5 to 7 show the seven purposes of citation of different types of methodological texts, including journal publications, books, book chapters, and others (e.g., reports), as well as the disciplines of the cited works. "N/A" in discipline indicates that the text is a purely methodological one or is published in a journal or book

which focuses on methodologies without specifying the substantive discipline. A quick glance at the tables reveals that methodological influences of QRS in CALL are eclectic, drawing on methodological traditions from a wide range of academic disciplines, namely applied linguistics, language education, medicine, psychology, education, sociology, technology, and social sciences. The following paragraphs focus on the top three areas where most methodological citations are found. The remainder can be found in Tables 5 to 7.

Among the seven, the most-cited area is data synthesis, which is the crucial methodological step in QRS and also the most divergent one (see Section 3.2.3). In total, 18 references were included in the QRS. Interestingly, none of these are CALL publications, but are instead publications in healthcare, education, technology, and sociology. This shows that this methodological step in QRS is still in its infancy in CALL and CALL researchers rely on methodological texts in other (sometimes distant) disciplines to guide its operationalisation. This echoes our earlier analysis that this is the least-standardised methodological step, practised only by nine out of the 16 QRS included (see Section 3.2.3).

Next, methodological references were frequently cited to introduce QRS as a distinctive type of research synthesis (10 references). Because QRS is relatively new, nearly half of the citations are drawn from publications in the healthcare, medicine, and nursing disciplines. There were only two references to QRS published in CALL. The third-most-commonly cited area is the description of scope and search strategy (eight references), with half from applied linguistics/language education (3 references) and CALL (1 reference). The purpose of citation here is to support the researchers' decision to focus on specific research databases and/or journals by referring to published QRS as benchmarks. Therefore, it makes

sense to refer to published QRS in the home field (CALL) or cognate fields (applied linguistics, language education).

[**Table 5.** Journal articles cited in the methodology sections of 16 QRS]

[**Table 6.** Books or book chapters cited in the methodology sections of 16 QRS]

[**Table 7.** Methodology texts cited in the methodology sections of 16 QRS]

5. Implications and conclusion

QRS is an emergent type of research synthesis in CALL, especially when compared to its quantitative, more established counterpart, meta-analysis. This methodological review surveyed 16 QRS in CALL published between 2003 and 2020, aiming to present the state-of-the-art of QRS methodology. The analysis follows the most recent and only known QRS methodological framework published in *TESOL Quarterly* (Author 1 et al., 2021) comprising seven stages.

Design research questions, identify scope of search, decide on inclusion and/or exclusion criteria, and evaluate literature using inclusion and/or exclusion criteria are stages reported in all QRS. On the other hand, methodological stages which show considerable differences in the included QRS are: identify keywords for conducting literature search, extract qualitative data, and synthesise qualitative data. Some QRS did not report keywords or search strings used to conduct the literature search because their focus was on the general research trends of the field. As argued in Sections 3.2.2 and 3.2.3, qualitative data extraction and synthesis are distinctive stages and both are indispensable in QRS. QRS reported data extraction using descriptive texts or tables but the number of categories of data extracted varied greatly from a

few to as many as 34. It is also found that data extraction of QRS in CALL was usually performed by one reviewer without addressing inter-coder reliability. As for qualitative data synthesis, where most divergence is shown, several qualitative research methodologies were referred to (e.g., grounded theory, content analysis, thematic analysis). Most of the QRS which reported data synthesis included definitions of the methodological approaches underlying data synthesis; very few (e.g., Author 1 & Author 2, 2020; Ciftci, 2016) not only described conceptual understanding of these approaches but also illustrated how qualitative data synthesis was operationalised. The diversity of methodologies employed to synthesise qualitative findings may be attributed to the range of topics focused on in the 16 QRS, which is an acute representation of the fact that CALL itself is a relatively young and disjointed area of research.

QRS is an emerging yet important research genre in CALL. In this review, 12 out of the 16 QRS were published after 2010; At the same time, the increasing number of QRS is likely due to the gradual recognition of the values of primary qualitative research in the field which not only captures the efficacy of technological interventions but also the sociocultural influences and the experiences of stakeholders (Levy & Moore, 2018). As Levy (2015) rightly pointed out, there is untapped potential in qualitative research (including QRS) in CALL to reveal how technologies aid language learners in naturalistic classroom settings, rather than in contexts which are restricted to "strictly controlled lab-style conditions and a PC monitor" (p. 566). Additionally, QRS, like other forms of qualitative research, can delve into the "complexities involved in closing in on the learners' experience" (p. 566). QRS is particularly ideal for unpacking the notion of "complexities" in CALL research by aggregating in-depth narratives and experiences of individual learners from a plethora of educational milieus, possessing diverse language proficiency levels, technological savviness,

and in various age groups, using a wide range of technological innovations. QRS allows for the presentation of not only the depth and richness of qualitative data but also its breadth, making qualitative findings more transferrable and generalisable.

The methodological contributions of this review are twofold: to *define* and *refine* QRS methodological practices. Through an in-depth analysis of the methodological sections of 16 QRS, this review, using a realist and inductive approach, *defines* quality of QRS in CALL. Specifically, it throws light on baseline and preferred methodological practices of QRS in CALL in six methodological stages (Table 8).

[**Table 8.** Baseline and preferred methodological practices of QRS in CALL]

Finally, and most importantly, this review aims to *refine* methodological practices of QRS in CALL through suggesting areas which can be improved. Most notably, CALL researchers who attempt to conduct QRS need to pay attention to qualitative data extraction and qualitative data synthesis, which are two distinctive methodological stages. Data extraction precedes data synthesis and needs to be done through identifying the types or categories of data to be extracted. These data types are determined by the research questions, although QRS researchers usually extract bibliographical information of the included studies (e.g., year of publication, learner characteristics, contextual characteristics). The extraction process needs to be documented in the methodology section or attached as an appendix. It is crucial to consider how reliability is maintained in data extraction, for instance, by increasing the number of coders, the provision of coder training, and carrying out coding comparisons. When all necessary data are extracted, they can be synthesised, that is, combined into more conceptual and higher-order themes that respond to the research questions. When performing

qualitative data synthesis, it is important to consider and clearly define the qualitative research methodology which underpins the synthesis (e.g., grounded theory), and to illustrate how the methodological approach is operationalised (e.g., by including a coding scheme).

An academic discipline is defined by the methodological traditions it draws upon (Tight, 2013). It is obvious from this review that QRS in CALL draws on an eclectic body of methodological traditions, very much aligned with Chapelle's (1997) observation that CALL researchers utilise "cross-disciplinary sources for perspectives and research methods" (p. 19). While it is still early to conclude whether QRS in CALL require some distinctive methodological practices, it is the aspiration of this review to enhance methodological awareness of CALL researchers conducting QRS and initiate methodological dialogues. Additionally, as one of the goals of QRS is to make qualitative research findings more accessible to not only researchers but also other stakeholders (e.g., teachers, policymakers), we hope that this methodological review can shed light on a "methodological threshold", or a set of baseline methodological practices, for conducting QRS in CALL, which can ultimately lead to the development of QRS appraisal tools for researchers and consumers of research (e.g., a tool reminiscent of the GRADE-CERQual).

Like other types of review, this methodological review is limited because its analysis is based on what is reported in the QRS. It is possible that some important methodological information or decisions were made without being documented in the article (e.g., because of word limits). It is therefore important to conduct a Delphi study to interview CALL researchers who have conducted QRS regarding the rationale of their methodological decisions. (6,669 words)

References

Author 1, A. (2020).

Author, 1, A., & Author 2, A. (2020).

Author 1, A. et al. (2021).

Avgousti, M. I. (2018). Intercultural communicative competence and online exchanges: a systematic review. *Computer Assisted Language Learning, 31*(8), 819–853. doi:10.1080/09588221.2018.1455713

Chapelle, C. A. (1997). CALL in the year 2000: Still in search of research paradigms? *Language Learning & Technology,1*(1), 19-43. http://llt.msu.edu/vol1num1/chapelle/default.html

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. doi: 10.1080/09588221.2014.960942

Chwo, G. S. M., Marek, M. W., & Wu, W.-C. V. (2018). Meta-analysis of MALL research and design. *System, 74*, 62–72. doi:10.1016/j.system.2018.02.009

Çiftçi, E. Y. (2016). A review of research on intercultural learning through computer-based digital technologies. *Journal of Educational Technology & Society, 19*(2), 313-327.

Çiftçi, E. Y., & Savas, 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. doi: 10.1017/s0958344017000313

Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO. *Qualitative Health Research, 22*(10), 1435–1443. doi:10.1177/1049732312452938

Debski, R. (2003). Analysis of research in CALL (1980–2000) with a reflection on CALL as an academic discipline. *ReCALL, 15*(2), 177–188. doi:10.1017/s0958344003000429

Duman, G., Orhon, G., & Gedik, N. (2014). Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL, 27*(2), 197–216. doi:10.1017/s0958344014000287

Ellis, R. (2015). Introduction: Complementarity in research syntheses. *Applied Linguistics, 36*(3), 285-289. doi: 10.1093/applin/amv015

Elgort, I. (2017). Technology-mediated second language vocabulary development: A review of trends in research methodology. *CALICO Journal, 0*(0). doi:10.1558/cj.34554

Flemming, K., & Noyes, J. (2021). Qualitative Evidence Synthesis: Where Are We at? *International Journal of Qualitative Methods*. doi: 10.1177/1609406921993276

Gong, Y., Lyu, B., & Gao, X. (2018). Research on teaching Chinese as a second or foreign Language in and outside Mainland China: A bibliometric analysis. *The Asia-Pacific Education Researcher, 27*(4), 277–289. doi:10.1007/s40299-018-0385-2

Hannes, K., & Macaitis, K. (2012). A move to more systematic and transparent approaches in qualitative evidence synthesis: update on a review of published papers. *Qualitative Research, 12*(4), 402–442. doi:10.1177/1468794111432992

Jiang, M. Y., Jong, M. S., Lau, W. W., Chai, C., Liu, K. S., & Park, M. (2020). A scoping review on flipped classroom approach in language education: Challenges, implications and an interaction model. *Computer Assisted Language Learning*, 1–32. doi:10.1080/09588221.2020.1789171

Levy, M. (2015). The role of qualitative approaches to research in CALL contexts: Closing in on the learner's experience. *CALICO Journal, 32*(3), 554-568. doi: 10.1558/cj.v32i3.26620

Levy, M., Hubbard, P., Stockwell, G., & Colpaert, J. (2015). Research challenges in CALL. *Computer Assisted Language Learning, 28*(1), 1-6. doi: 10.1080/09588221.2014.987035

Levy, M., & Moore, P. J. (2018). Qualitative research in CALL. *Language Learning & Technology, 22*(2), 1-7. doi: 10125/44638

Lewin, S., Booth, A., Glenton, C., Munthe-Kaas, H., Rashidian, A., Wainwright, M., … Noyes, J. (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. *Implementation Science, 13*(S1). doi:10.1186/s13012-017-0688-3

Lewis, T., & O'Dowd, R. (2016). Online intercultural exchange and foreign language learning: A systematic review. In: R. O'Dowd & T. Lewis (eds.), *Online intercultural exchange: Policy, pedagogy, practice* (pp. 21-66). London: Routledge.

Lin, H. (2015). A meta-synthesis of empirical research on the effectiveness of computer-mediated communication (CMC) in SLA. *Language Learning & Technology, 19*(2), 85–117. doi:10125/44419

Locke T., & Andrews R. (2004). *A systematic review of the impact of ICT on literature-related literacies in English 5-16*.Retrieved from http://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/eng_rv3.pdf?ver=2006-03-02-124833-123.

Ma, X., Gong, Y., Gao, X., & Xiang, Y. (2017). The teaching of Chinese as a second or foreign language: a systematic review of the literature 2005–2015. *Journal of Multilingual and Multicultural Development, 38*(9), 815–830. doi:10.1080/01434632.2016.1268146

Macaro, E., Handley, Z., & Walter, C. (2011). A systematic review of CALL in English as a second language: Focus on primary and secondary education. *Language Teaching, 45*(1), 1–43. doi:10.1017/s0261444811000395

Noyes, J., Booth, A., Cargo, M., Flemming, K., Harden, A., Harris, J., Garside, R., Hannes, K., Pantoja, T., & Thomas, J. (2019). Chapter 21: Qualitative evidence. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, &, V. A. Welch (eds.), *Cochrane Handbook for Systematic Reviews of Interventions version 6.0* (updated July 2019). Cochrane. Retrieved from www.training.cochrane.org/handbook.

Page, M.J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C.D. et al. (2021) *The PRISMA 2020 statement: an updated guideline for reporting systematic reviews*. BMJ 2021(372), n71. doi: 10.1136/bmj.n71

Plonsky, L., & Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics, 36*, 73–97. doi:10.1017/s0267190516000015

Plonsky, L., & Ziegler, N. (2016). The CALL-SLA interface: Insights from a

second-order synthesis. *Language Learning & Technology, 20*, 17–37.

doi.org/10125/44459

Sauro, S. (2011). SCMC for SLA: A research synthesis. *CALICO Journal, 28*(2), 369–391.

doi:10.11139/cj.28.2.369-391

Shadiev, R., Hwang, W.-Y., & Huang, Y.-M. (2017). Review of research on mobile language

learning in authentic environments. *Computer Assisted Language Learning, 30*(3-4),

284–303. doi:10.1080/09588221.2017.1308383

Stickler, U., & Hampel, R. (2015). Qualitative research in CALL. *CALICO Journal, 32*(3),

380-39. doi: 10.1558/cj.v32i3.27737

Stockwell, G. (2007). A review of technology choice for teaching language skills and areas in

the CALL literature. *ReCALL, 19*(2), 105–120. doi:10.1017/s0958344007000225

Tight, M. (2013). Discipline and methodology in higher education research. *Higher*

Education Research & Development, 32(1), 136–151.

doi:10.1080/07294360.2012.750275

Tullock, B., & Ortega, L. (2017). Fluency and multilingualism in study abroad:

Lessons from a scoping review. *System, 71*, 7–21. doi: 10.1016/j.syste

M.2017.09.019

Zhao, Y. (2013). Recent developments in technology and language learning: A literature

review and meta-analysis. *CALICO Journal, 21*(1), 7–27. doi:10.1558/cj.v21i1.7-2.
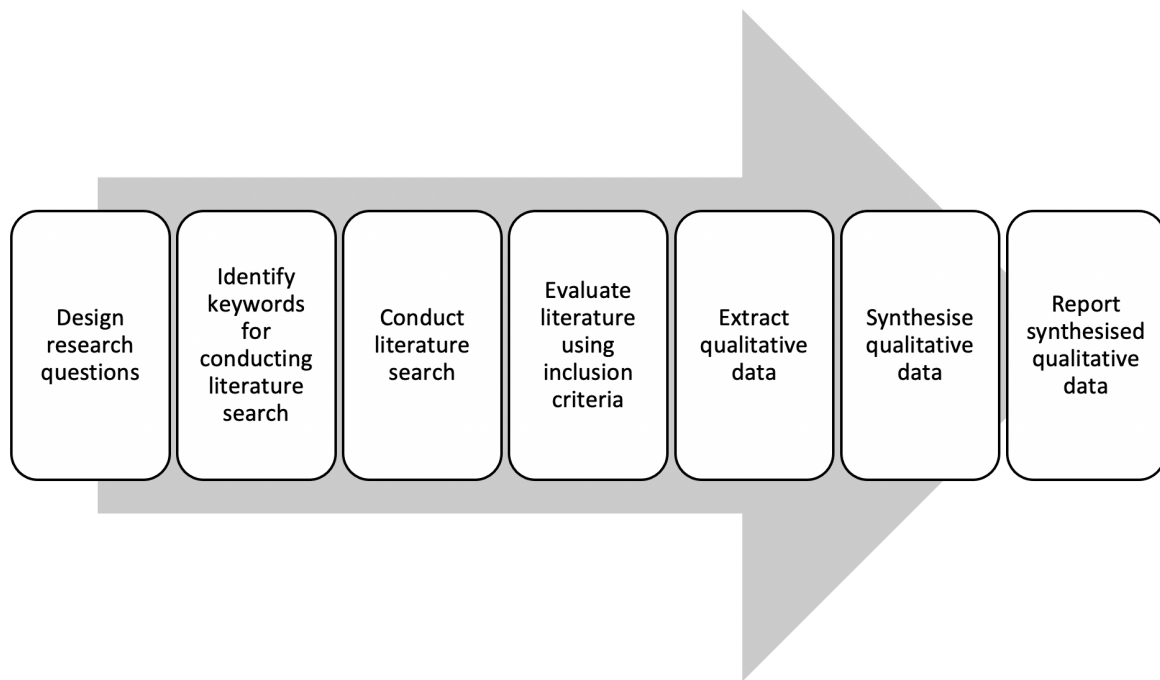
Figure 1. A methodological framework

Table 1. Breakdown of the initial search results (n=496)

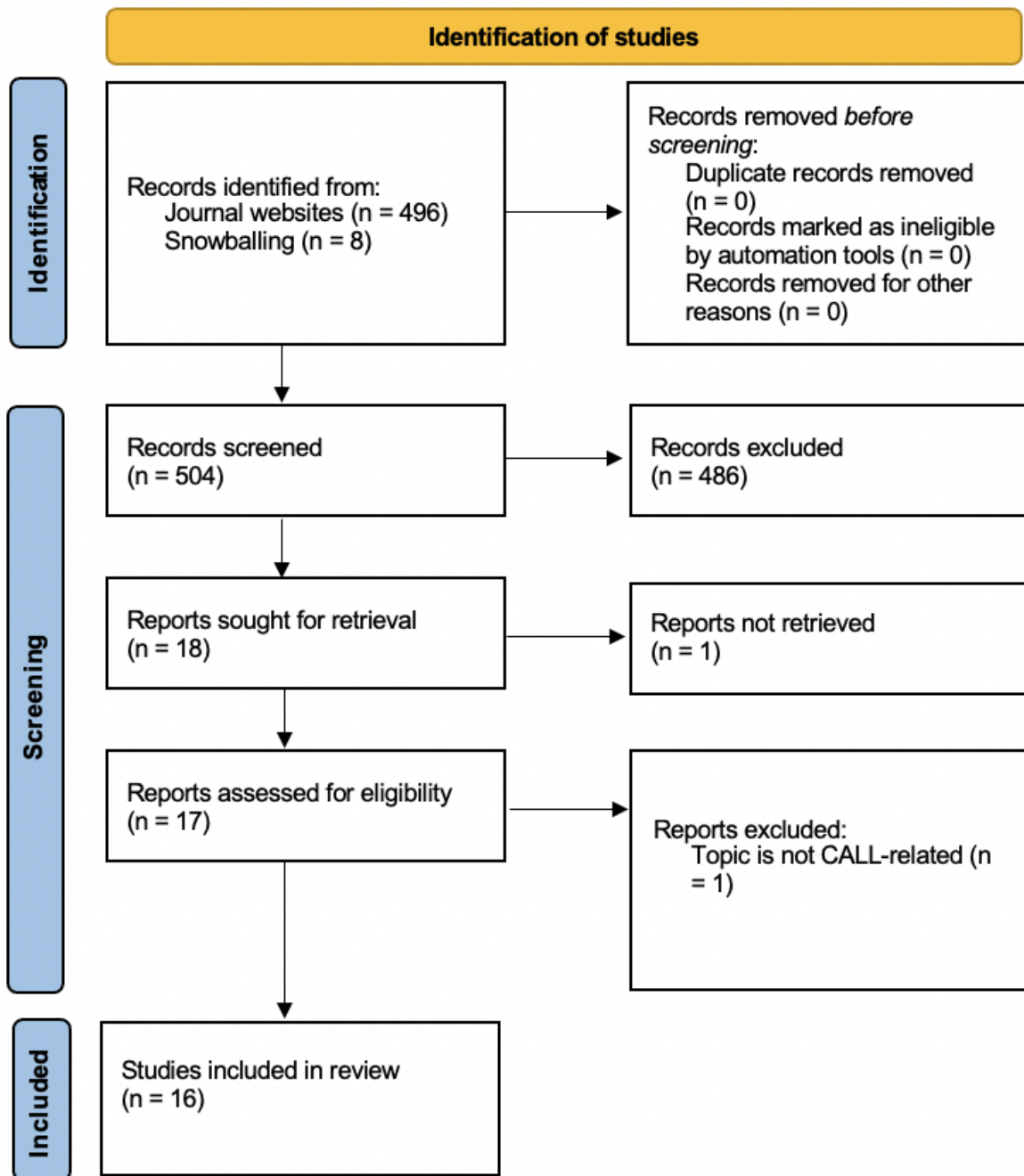| CALL Journal	Search results (n =)	QRS included (n =)
CALICO	43	3
Computer Assisted Language Learning	67	2
Language Learning & Technology	42	1
ReCALL	275	2
System	69	1

Figure 2. PRISMA diagram showing the flow of literature search

Table 2. Length of methodology section

	Qualitative research synthesis	No. of words of methodology section	No. of words of whole article	%[4]
1	Shadiev et al. (2017)	293	9142	3
2	Macaro et al. (2012)	494	18866	3
3	Lewis & O'Dowd (2016)*	791	18005	4
4	Ciftci (2016)	309	7176	4
5	Stockwell (2007)	674	7980	8
6	Ciftci & Savas (2018)	797	10188	8
7	Lock & Andrews (2004)^	2312	27830	8
8	Avgousti (2018)	1385	14924	9
9	Elgort (2018)	909	10411	9
10	Duman et al. (2014)	927	9482	10
11	Chwo et al. (2018)	908	8323	11
12	Sauro (2011)	1192	11102	11
13	Debski (2003)	616	5584	11
14	Chen (2016)	2088	17379	12
15	Zhao (2003)	1278	9100	14
16	Author 1 & Author 2 (2020)	1450	8652	17

Note: * refers to QRS entry published as a book chapter while ^ refers to QRS entry published as an online report.

---

[4] This is calculated through dividing the number of words of the methodology section by the number of words of the whole article, then converting the result into percentage.
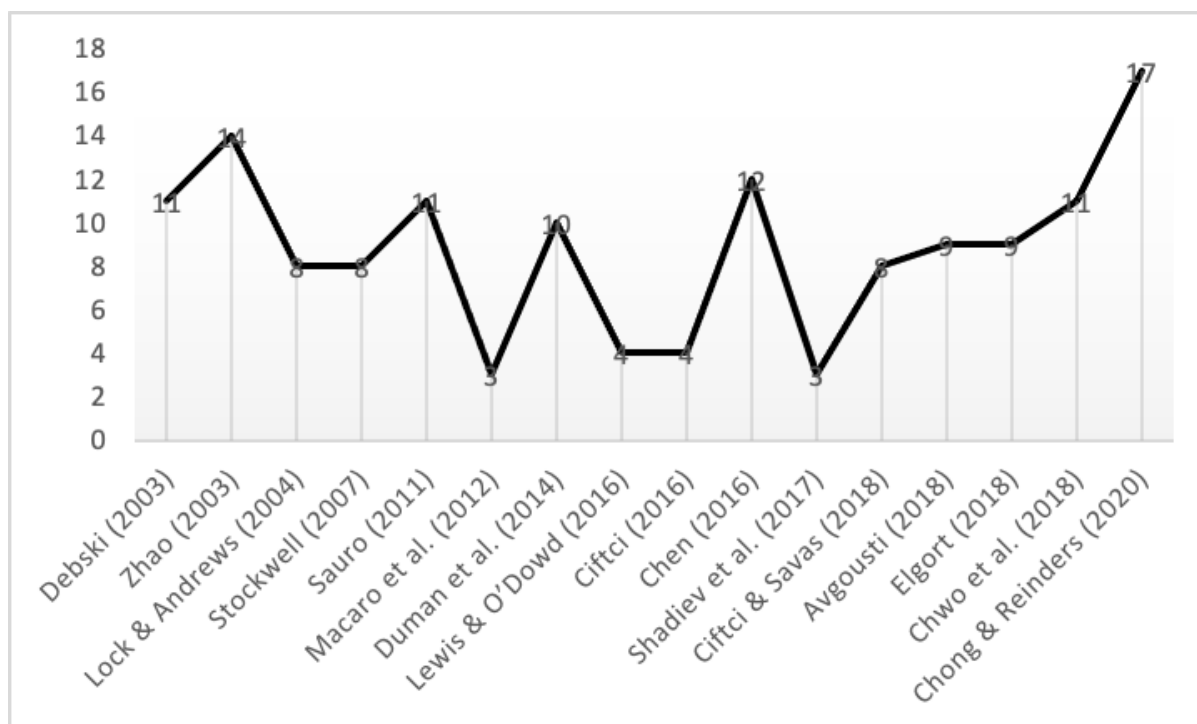
Figure 3. Number of words in the methodology section (%)

Table 3. Methodology/full-text word percentage by CALL journals

QRS	Methodology/full-text word percentage	Journal
Zhao (2003)	14%	*CALICO*
Sauro (2011)	11%	*CALICO*
Elgort (2018)	9%	*CALICO*
Chen (2016)	12%	*Computer Assisted Language Learning*
Shadiev et al. (2017)	3%	*Computer Assisted Language Learning*
Avgousti (2018)	9%	*Computer Assisted Language Learning*
Ciftci (2016)	4%	*Educational Technology & Society*

Macaro et al. (2012)	3%	*Language Teaching*
Author 1 & Author 2 (2020)	17%	*LLT*
Debski (2003)	11%	*ReCALL*
Stockwell (2007)	8%	*ReCALL*
Duman et al. (2014)	10%	*ReCALL*
Ciftci & Savas (2018)	8%	*ReCALL*
Chwo et al. (2018)	11%	*System*

Table 4. Methodological stages reported in 16 QRS in CALL

	Design research questions	Identify keywords for conducting literature search	Identify scope of search	Decide on inclusion and/or exclusion criteria	Evaluate literature using inclusion and/or exclusion criteria	Extract qualitative data	Synthesise qualitative data
Avgousti (2018)	✓*	✓	✓	✓	✓	✓	✓ (thematic synthesis)
Chen (2016)	✓*	✓	✓	✓	✓	✓	✓ (grounded theory)
Author 1 & Author 2 (2020)	✓*	✓	✓	✓	✓	✓*	✓ (grounded theory)
Chwo et al. (2018)	✓*	✓	✓	✓	✓		✓ (thematic analysis)
Ciftci (2016)	✓*	✓	✓	✓	✓	✓	✓ (open coding)
Ciftci & Savas (2018)	✓*	✓	✓	✓	✓	✓	✓ (grounded theory)

Debski (2003)	✓*		✓	✓	✓	✓	
Duman et al. (2014)	✓*	✓	✓	✓	✓	✓	✓ (content analysis)
Elgort (2018)	✓*	✓	✓	✓	✓	✓	✓ (TMVD)
Lewis & O'Dowd (2016)	✓*	✓	✓	✓	✓	✓*	
Lock & Andrews (2004)	✓*	✓	✓	✓	✓	✓	✓ (narrative synthesis)
Macaro et al. (2012)	✓*	✓ (not specified)	✓	✓	✓	✓*	
Sauro (2011)	✓*	✓	✓	✓	✓		
Shadiev et al. (2017)	✓*	✓	✓	✓	✓	✓	
Stockwell (2007)	✓*		✓	✓	✓	✓	
Zhao (2003)	✓*	✓	✓	✓	✓		

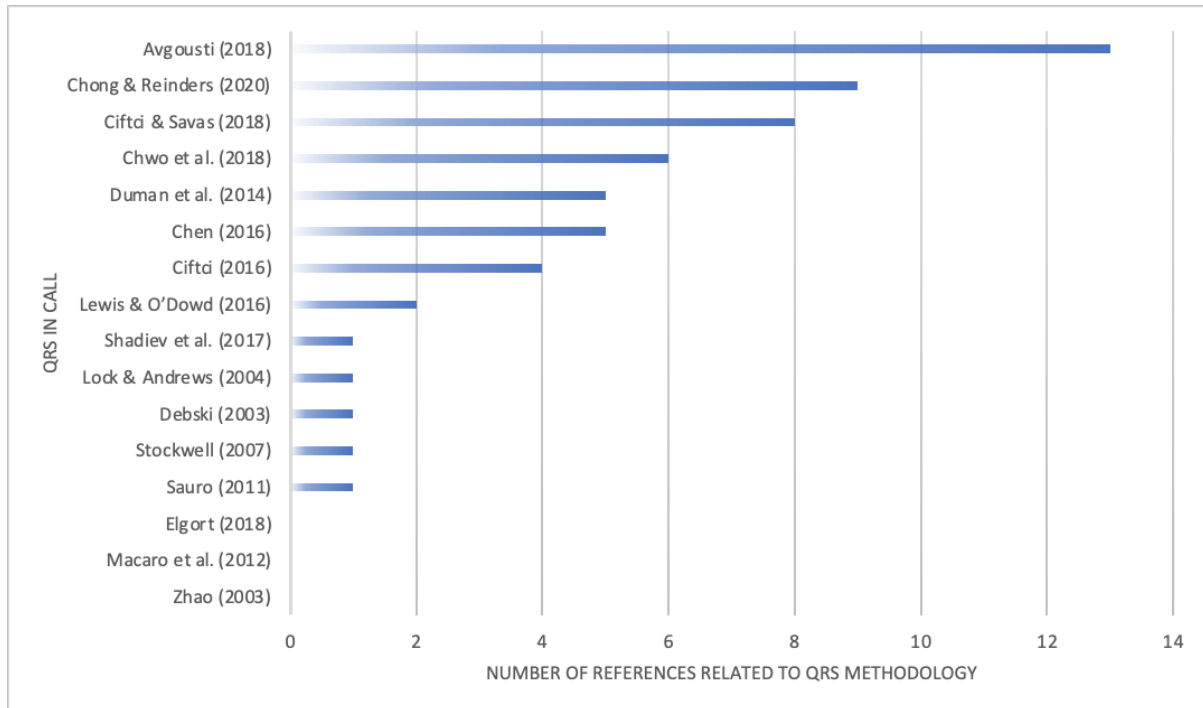* Information not mentioned in the methodology section

Figure 4. Number of references related to QRS methodology

Table 5. Journal articles cited in the methodology sections of 16 QRS

Purpose of citation	N	Discipline of works cited
Introduction to QRS	9	CALL (n=2) healthcare/Medicine/Nursing (n=4) Technology (n=2) Education (n=1)
Justification of number of studies included	2	CALL (n=2)
Scope/search strategy	5	CALL (n=1) Applied Linguistics/Language Education (n=3) Psychology (n=1)
Inclusion/exclusion criteria	4	CALL (n=3) Applied Linguistics/Language Education (n=1)
Data extraction	2	Education (n=2)
Data synthesis	7	Healthcare/Medicine/Nursing (n=3) Education (n=2) (Educational) Technology (n=2)

Table 6. Books or book chapters cited in the methodology sections of 16 QRS

Purpose of citation	N	Discipline of works cited
Inclusion and exclusion criteria	1	Applied Linguistics (n=1)
Data synthesis	4	Psychology (n=2) Applied Linguistics/Language Education (n=2)

Table 7. Methodology texts cited in the methodology sections of 16 QRS

Purpose of citation	N	Discipline of works cited
Introduction to QRS	1	N/A (n=1)
Methodological framework	2	N/A (n=2)
Scope/search strategy	3	N/A (n=2) Social Sciences (n=1)
Inclusion and exclusion criteria	2	Social Sciences (n=1) Language Education (n=1)
Data extraction	3	Social Sciences (n=2) N/A (n=1)
Data synthesis	11	Sociology (n=4) N/A (n=7)

Table 8. Baseline and preferred methodological practices of QRS in CALL

Methodological stage of QRS	Baseline methodological practice	Preferred methodological practice
Design research questions	• QRS in CALL usually include more than one research question. • These research questions can be about: technologies (e.g., use of technologies), impacts (e.g., learners' experiences), and research (e.g., methodologies).	• Reviewers can consider including an open-ended research question to allow for unexpected but relevant ideas to be included in QRS.
Identify keywords	• Reviewers include keywords and/or search strings relevant to the topic of the QRS and its research questions.	• Reviewers justify the use of more broad-brush versus more fine-grained search strategies (e.g., in relation to the scope and focus of the QRS).
Conduct literature search	• Reviewers search from multiple sources, most commonly from online databases focusing on journal articles. • When deciding on the search strategy, reviewers consider disciplinary relevance, quality of publications, accessibility of	• Reviewers consider publication bias, that is, the publications included in the QRS are a representative sample. For instance, to ensure representativeness of research contexts, reviewers conduct search on not only international

	publications, and coverage of the databases.	refereed journals but regional journals.
Evaluate literature using inclusion/exclusion criteria	• Reviewers usually use multiple inclusion criteria.	• Reviewers can develop inclusion criteria which cover important areas of concerns: types of publication, years of publication, relevance to research questions, rigour of studies, participants, contexts, research design, and languages of publication.
Extract qualitative data	• Reviewers prepare a data extraction form with pre-determined focuses relevant to the research questions. • Data extraction is performed by one reviewer.	• More than one reviewer can extract relevant data independently. Differences in understanding can be resolved through discussions.
Synthesise qualitative data	• Reviewers identify and define the methodological approach used to synthesise qualitative findings (e.g., grounded theory).	• Reviewers not only demonstrate a conceptual understanding of the methodological approach but provide examples of how the approach is operationalised (e.g., by including coding schemes or describing how themes are identified).