

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

A Comprehensive Guide to Conducting Systematic Reviews for Research and Evidence Synthesis in the Field of Applied Linguistics

Ngo Cong-Lem, PhD

Faculty of Foreign Languages, Dalat University

Table of Contents

Introduction.....	3
Terminologies and Categorization of Literature Reviews	5
<i>Relevant Terminologies.....</i>	<i>5</i>
<i>Categorisation of Literature Reviews.....</i>	<i>5</i>
Systematic Review as a Comprehensive Evidence Synthesis Method.....	8
Stages of Conducting a Systematic Review	10
<i>Formulating Research Questions and Objectives</i>	<i>10</i>
<i>Literature Search Strategy.....</i>	<i>12</i>
<i>Screening and Selection of Studies.....</i>	<i>13</i>
<i>Data Extraction and Data Synthesis.....</i>	<i>13</i>
<i>Quality Appraisal for Included Studies and Risk of Bias.....</i>	<i>14</i>

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

<i>Reporting and Discussing Review Findings</i>	15
Addressing Bias in Systematic Review and Whether SR is Appropriate	16
Developing a Checklist for Appraising Studies in AppLing Systematic Review	17
Systematic Reviews in the Age of Generative Artificial Intelligence	20
Enhancing Applied Linguistics through Systematic Reviews: Policy and Practice	22
Conclusion	23
References	23

Abstract

This proposed book chapter serves as a comprehensive guide for anyone interested in conducting systematic reviews (SRs) within the field of applied linguistics. Systematic reviews play a crucial role in the field, allowing researchers to analyze and synthesize existing research, ultimately leading to evidence-based decisions and advancements in applied linguistics knowledge. The chapter offers detailed guidance on the various stages of conducting an SR in applied linguistics. First, it delves into a systematic methodological approach for categorizing different types of literature reviews. This lays the foundation for understanding the specific type of review you intend to conduct. Next, it comprehensively explores the various stages involved in conducting an SR. This includes formulating clear and focused research questions,

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

developing effective search strategies to identify relevant literature, meticulously screening and selecting studies that meet the predefined criteria, extracting and synthesizing data from the selected studies, critically appraising the quality of the included studies, and finally, reporting the findings in a clear and concise manner. Furthermore, the chapter proposes a unique checklist specifically tailored for appraising studies within the applied linguistics field. This checklist serves as a valuable tool for researchers to ensure they are comprehensively evaluating the quality and relevance of the studies they include in their SR. Finally, the chapter explores the potential of using generative artificial intelligence (AI) tools to streamline and enhance the process of conducting systematic reviews. It discusses how such AI tools could potentially be employed in various stages of the SR process, highlighting the opportunities and challenges associated with their integration. By encompassing these key aspects, this book chapter aims to equip researchers with the knowledge and tools necessary to conduct rigorous and insightful systematic reviews in the field of applied linguistics.

Keywords: *Systematic Review, Applied Linguistics, Evidence-Based, Appraisal Tool, Generative Artificial Intelligence*

Introduction

As the field of applied linguistics (AppLing) continues to evolve, the volume of topics and publications has grown significantly. This expansion underscores an increasing need for literature reviews, which serve to distill complex bodies of research into coherent, evidence-based narratives. Such narratives are invaluable in informing both current and future research endeavors.

Literature review (LR), also known as secondary research, research synthesis, or evidence synthesis, serves as comprehensive examinations of the existing literature on a particular topic (Briner & Denyer, 2012; Chong & Plonsky, 2023; Ellis, 2015). Ellis (2015) offers a broad definition of research synthesis as "any attempt to review the literature related to a specified topic, either narrowly or broadly defined" (p. 285). This definition underscores the versatility of literature reviews in providing a comprehensive understanding of a topic.

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

LR plays a critical role in academic research for several reasons. Firstly, it provides scholars with a comprehensive understanding of existing research, including emerging themes, gaps in the literature, and future research directions. By doing so, it helps to minimize the occurrence of ineffective or redundant research, commonly referred to as "research waste" (Isaacs & Chalmers, 2023, p. 1). This contributes to the more efficient use of resources and efforts, ultimately advancing the field more effectively. Moreover, LR not only informs individual primary studies but also shapes the broader research landscape. Secondly, SR is increasingly essential as a means of synthesizing evidence to inform policy and practice (Beerkens, 2018; Siddaway et al., 2019). By evaluating and synthesizing existing research, LR provides a robust foundation for policy-making and practice implementation.

Systematic reviews (SRs), a specific type of literature review (LR) that adopts a transparent, systematic, and replicable approach, are becoming increasingly prevalent in the AppLing field. Compared to traditional narrative reviews, SRs offer advantages due to their transparent and systematic methodology, including clearly defined literature search procedures and explicit inclusion/exclusion criteria. A scientometric review of 42 AppLing journals by Liu and Hu (2024) revealed a growing trend toward literature reviews as a preferred method of research synthesis in the field. However, the rigorous execution of SRs, a specific type of literature review, can be a daunting and stressful task for researchers, as noted by Siddaway et al. (2019). This book chapter aims to address this challenge by providing a comprehensive methodological guide specifically for conducting SRs in AppLing research. This guide will equip researchers with the knowledge and tools to produce trustworthy and credible evidence on AppLing topics.

This chapter's structure commences with a discussion of LR terminologies, including the categorization of different LR types. It then delves into the stages of conducting a SR, followed by recommendations for addressing bias in SRs and the potential use of generative AI. Additionally, the chapter introduces a checklist for appraising the quality of studies included in an SR, drawing upon established tools.

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

Terminologies and Categorization of Literature Reviews

Before delving into SRs and their applications, it is essential to establish a solid understanding of the related terminologies and categorizations, which serve to position SRs within the broader spectrum of literature reviews (LRs).

Relevant Terminologies

LR can generally be categorized into two main types: traditional non-systematic reviews and systematic reviews. Traditional non-systematic reviews are typically conducted by established scholars in a specific research area. Their aim is often to provide a narrative or "tell a story" about the empirical landscape, while also identifying research gaps and offering recommendations (Norris & Ortega, 2006, p. 5). However, these reviews often lack a clearly defined or replicable methodology, leading to criticism for their potential bias and lack of rigor and transparency (Macaro, 2019; Norris & Ortega, 2007).

LR is a comprehensive term that encompasses several subcategories, including research synthesis and evidence synthesis. While often used interchangeably, these terms emphasize different aspects of the review process. On the one hand, research synthesis typically focuses on the systematic analysis and integration of prior research to identify trends, patterns, or gaps within the existing literature (Thomas & Harden, 2008). It involves combining diverse findings from multiple studies into a coherent narrative, ultimately contributing to a deeper understanding of a particular topic or field. On the other hand, evidence synthesis primarily emphasizes the utilization and evaluation of research evidence to assess the effectiveness of a specific tool, intervention, or approach (e.g., Glasziou et al., 2014). This involves critically appraising the quality and relevance of research evidence to help guide decision-making or practice.

Categorisation of Literature Reviews

LR comes in various forms, each serving different purposes and employing different methodological approaches. Norris and Ortega (2006) identified several types, such as the authoritative tour, the comprehensive bibliographical review, the vote-counting review, the historiographical review, the integrative review, and the critical review (cited in Ellis, 2015, p.

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

285). Similarly, Chong and Plonsky (2023) offered a typology of 13 types of secondary research in AppLing:

1. Critical review
2. Meta-analysis
3. Methodological synthesis
4. Mixed review
5. Narrative review
6. Qualitative research synthesis
7. Research agenda
8. Research into practice
9. Scoping review
10. State-of-the-art review
11. Systematic literature review
12. Historical review
13. Bibliometric review

Chong and Plonsky (2023) also introduced four criteria to further classify these review categories based on their purposes (research or practice-oriented), review process (less or more systematic), text (mono- or multimodal), and structure (more or less standardized). Each type of LR serves different goals. For instance, meta-analysis aims to quantitatively analyze and synthesize data across multiple studies, while a scoping review aims to provide a broad overview of a research area. Similarly, a narrative review tells a story about the research landscape, and a systematic literature review employs a rigorous methodology to synthesize research findings. The choice of LR type depends on the research question, the nature of the available literature, and the desired outcomes. Researchers should carefully select the most appropriate type of LR to achieve their research goals effectively.

While Chong and Plonsky's (2023) categorisation of reviews is comprehensive, it tends to be more purpose-oriented, focusing on aspects like research agenda or methodological synthesis. However, a key element that's lacking is a categorisation of reviews based on their approach to analysing and synthesising research literature. This is because a literature review (LR) can be

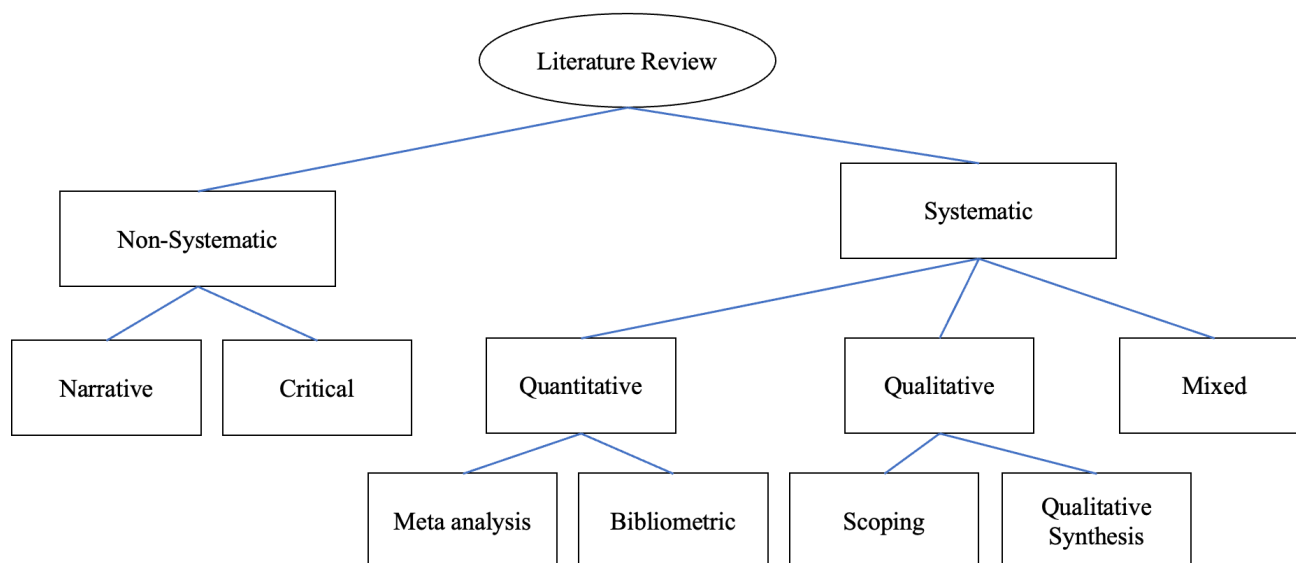
See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

conducted using various synthesis methods, each with its own implications. For example, when reviewing the effectiveness of extensive reading in improving English language outcomes for English as a foreign language learners, scholars may choose between conducting a meta-analysis or a qualitative systematic review. In such cases, the primary purpose of the review remains largely the same, but the methodological approach to synthesising evidence differs significantly. As a result, there is a pressing need to propose a new scheme for classifying reviews at a more overarching level, based on their methodological approaches to literature search and data synthesis. This can aid scholars in making informed decisions about which types of reviews are most relevant to their research questions and objectives. Figure 1 illustrates the categorisation of LRs based on their analytical approach to synthesising research literature.

Figure 1

A Classification of Literature Reviews Based on Their Methodological Approaches



As delineated in Figure 1, LRs can be fundamentally divided into systematic and non-systematic reviews. Also, within non-systematic reviews, which are commonly characterized with traditional

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

methods, narrative and critical reviews are two representative subcategories. As for the SR division, it can be further sub-divided into those utilizing a quantitative approach, such as meta-analyses and bibliometrics, those using qualitative content analysis, and those who adopt mixed methods LR, which is "a synthesis in which researchers combine qualitative, quantitative, and mixed methods studies, and apply a mixed methods approach in order to integrate those studies, for the broad purpose of breadth and depth of understanding and corroboration" (Heyvaert et al., 2011, p. 662). While quantitative and qualitative approaches to data analysis in LR are common, mixed-methods LR represents a more recently emerging approach in LR that aims to provide a more balanced view of evidence from both quantitative and qualitative perspectives (e.g., Bedle, 2017; Mansfield et al., 2024). This is indicative of evolution in LR methodologies. The list of categories of reviews presented in Figure 1 is by no means an exhaustive list of possible types/purposes of reviews (e.g., Chong & Plonsky 2023). This categorization of reviews is not new and has been discussed in previous literature in other fields (e.g., Siddaway et al., 2019).

Systematic Review as a Comprehensive Evidence Synthesis Method

In contrast to traditional reviews, SR has evolved into a methodology that enables a more rigorous and credible form of evidence synthesis. Siddaway et al. (2019) contends that SR is "characterized by being methodical, comprehensive, transparent, and replicable" (p. 751). Building on previous literature (e.g., Siddaway et al., 2019), four key components/dimensions of systematic reviews involve the comprehensiveness of the literature search, reliability, replicability, and quality of evidence.

First, a SR should be comprehensive in its search for literature. Siddaway et al. (2019) states that SR involves "a comprehensive search to locate all relevant published and unpublished work on a subject" (p. 747). Second, considered a more stringent scientific approach, the reliability of an SR is of critical importance. In this regard, SR affords such an approach "with an emphasis on practices that would be systematic, transparent, and replicable, just as other scientific methodologies strive to be" (Norris & Ortega, 2007, p. 807).

Third, replicability is another crucial characteristic of an SR. Siddaway et al. (2019) contend that systematic reviews "are characterized by a methodical and replicable methodology and

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

presentation" (p. 747). SR is increasingly employed as a more favorable approach for its transparency, stringency, and replicability. As such, the evidence or findings are more trustworthy and credible. However, it is important to acknowledge that while SR reduces bias in comparison with traditional reviews, it cannot eliminate bias (Macaro, 2019).

Fourth, the quality of included studies should be taken into account when interpreting the findings, discussion, and recommendations in the SR. Siddaway et al. (2019) assert that there is a necessity for "a critique of the extent, nature, and quality of evidence in relation to a particular research question" (p. 747).

Figure 2

A Framework for Evaluating Systematic Reviews



These attributes make SR a more rigorous and trustworthy approach to literature review. However, SR is not a panacea; it serves solely as a means of synthesizing the evidence and research that is available (Siddiway, 2019).

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

Stages of Conducting a Systematic Review

Based on our categorization of LRs, the term "systematic review" can encompass both meta-analysis and qualitative systematic review. However, in previous literature, "meta-analysis" is often implicitly understood as quantitative systematic reviews, whereas the term "systematic review" is commonly used to refer to qualitative systematic review. This usage of the terms is common and is also reflected in the chapter titles of this book. This section provides detailed guidance on how to plan and conduct a systematic review in the sense that it is a qualitative systematic review.

Formulating Research Questions and Objectives

The first pivotal step in conducting a SR is to formulate clear and precise research questions and objectives. This step not only sets the trajectory for the entire systematic review process but also ensures that the investigation remains focused and purposeful. According to Siddaway et al. (2019), the two main purposes of a systematic review are "to establish to what extent existing research has progressed toward explaining a problem and to clarify the extent to which a new or existing theory explains the existing evidence" (p. 768). More specifically, the research questions in a SR can target these four areas/intentions: general effectiveness, comparative analysis, specific population, and impact/influence.

Guidance and examples for formulating SR's RQs are presented in Table 1.

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

Table 1

Types, Formulas, and Examples of Research Questions for Systematic Reviews

Type of SR Research Questions	Formula	Examples
General effectiveness	<ul style="list-style-type: none"> • What is effectiveness of X (technology/pedagogy) in improving Y (learning outcome) for Z population? • To what extent do X (technology/pedagogy) contribute to Y (learning outcome) in Z context? 	<ul style="list-style-type: none"> • What is the effectiveness of extensive reading in improving vocabulary for English-as-a-Foreign-Language (EFL) learners?) • To what extent to web-based portfolio contribute to self-regulated learning in online language courses?
Comparative effectiveness	<ul style="list-style-type: none"> • Compared to A method, how effective is B in fostering C outcome among D population? • What are the relative advantages and disadvantages of A versus B for measuring C (learning outcome) in D learner group? 	<ul style="list-style-type: none"> • Compared to traditional reading teaching approach, how effective is extensive reading in fostering reading comprehension achievement among EFL learners? • What are the relative advantages and disadvantages of standardised assessment and portfolio as a form of alternative assessment for measuring L2 writing achievement among EFL learners?
Population	<ul style="list-style-type: none"> • What are the most effective X (technology/pedagogy) for Y population? • How can X be adapted to meet the specific needs and cultural backgrounds of Y population? 	<ul style="list-style-type: none"> • What are the most effective language learning strategies for Vietnamese EFL learners? • How can language learning materials be adapted to meet the specific needs and cultural backgrounds of adult immigrant learners?

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

Impact	<ul style="list-style-type: none"> • What impact does X have on Y in Z setting? • How does the use of X (technology/pedagogy) influence Y (learning outcomes) in/outside Z setting? 	<ul style="list-style-type: none"> • What impact does flipped classroom have on students' L2 proficiency in Asian educational contexts? • How does the use of portfolio-based learning approach influence willingness to communicate among EFL learners?
--------	---	--

Literature Search Strategy

Literature search is crucial for SRs as it determines the scope, quality, and relevance of the literature, which in turn determines the data for the review. Searching for literature is often conducted on major scientific databases such as Scopus and Web of Science using relevant keywords. These keywords are carefully selected and combined into search strings to extract the most relevant documents. This process may involve testing different keywords and search strings to find the most effective ones.

Three other strategies for gathering more relevant literature include: manual search directly from a journal, input from experts, and snowballing from the references of the included studies. The use of different search strategies and a careful design of inclusion and exclusion criteria demonstrate the comprehensiveness of the SR and its trustworthiness in terms of the evidence presented in the review. This is because publication bias is a significant issue that SR aims to address. The comprehensive search, including unpublished literature, aims to address this issue.

Next, a key aspect of SR is the establishment of inclusion and exclusion criteria, which is largely based on the purpose of the review and the research questions. The goal is to have a set of criteria that dictate what studies are relevant and what are not. Examples of inclusion criteria can target the type of research design (e.g., experiment, qualitative, mixed-methods studies), type of publication (e.g., empirical article, reviews, book chapter), study focus, population, and setting. The exclusion criteria are mostly the opposite of those in the inclusion criteria.

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

SR also involves searching for grey literature, which comprises non-research types of publications such as reports from organizations.

Screening and Selection of Studies

The screening and selection of studies is a pivotal stage in the systematic review process, requiring careful attention to predefined inclusion and exclusion criteria. This stage ensures that the selected body of evidence is not only relevant but also closely aligned with the research objectives, thereby safeguarding the integrity and validity of the review. Before screening begins, it is common practice to remove duplicate records retrieved from different search databases using software such as EndNote or Covidence.

In SRs, screening commonly encompasses two primary stages: title and abstract screening, followed by full-text screening. The first stage, title and abstract screening, involves a thorough assessment of the titles and abstracts of the identified records to determine their relevance to the review's overarching purpose and scope. Records that meet the relevance criteria proceed to the next phase. The second stage, full-text screening, entails retrieving the full texts of the selected articles and then meticulously analyzing them to ascertain their alignment with the review's scope. This stage marks the final screening step before proceeding with data extraction and analysis.

It is recommended that screening be conducted in duplicate by two reviewers to minimize bias and enhance accuracy. When conflicts arise between reviewers' decisions, they often collaborate to resolve them, or a third reviewer is consulted to make a final decision.

Data Extraction and Data Synthesis

Following the screening stage, the reviewers should have a definitive number of included studies – these will be used for the next step, data extraction. Data extraction refers to the process in which the reviewers extract relevant segments of information from each included article that can be used to answer the research question or objective of the review.

Data extraction can be facilitated by creating a table with columns for the target information the reviewers want to capture. Common fields include authors, year of publication, study aim, methods, findings, conclusions, and limitations. The reviewers can use readily available

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

software like Microsoft Word, Excel, Google Docs, or Google Sheets. Alternatively, dedicated systematic review software like Covidence offers both default and customizable data extraction templates, but requires a subscription fee.

Once the reviewers complete their data extraction into a table or template, data synthesis can begin. This process shares similarities with qualitative research methods, focusing on identifying emerging themes from the extracted data. Depending on the purpose of the systematic review, different qualitative analysis methods can be chosen, such as thematic analysis and grounded theory. For example, if a SR aims to build a model or theory, grounded theory would be appropriate. Conversely, if the purpose is to identify major themes within the literature surrounding a specific phenomenon, thematic analysis would be a suitable choice. According to Thomas and Harden (2008), thematic analysis can be used to “bring together and integrate findings of multiple qualitative findings” (p. 1). They described three stages of thematic analysis: initial 'line-by-line' coding of text, development of 'descriptive themes', and generation of 'analytical themes' (Thomas & Harden, 2008, p. 1). While descriptive themes stay close to primary studies, analytical themes extrapolate beyond, creating novel interpretive constructs, explanations, or hypotheses.

Quality Appraisal for Included Studies and Risk of Bias

The number of studies included in a SR shouldn't be confused with their quality. This critical distinction necessitates the evaluation of both the individual quality of the included studies and the overall quality of evidence within the review. Critical appraisal tools (CATs) come in as checklists with questions tailored to specific study designs, aiming to delve into their methodological aspects and assess their quality.

A two-pronged approach is crucial for understanding and analyzing the quality of evidence. Firstly, at the primary level, it is important to consider the study design, which determines the degree of confidence we can place in claims of causal relationships and the generalizability of the study's findings. Traditionally, a hierarchy of decreasing confidence exists, with experimental studies at the top, followed by quasi-experimental, pre-post, survey-based, and lastly, qualitative research studies (as outlined by the National Institute of Clinical Studies, 2009).

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

Secondly, it is essential to assess the study's performance on relevant CATs. While CATs are widely used in healthcare SRs, their application is less common in AppLing research. However, efforts to develop such tools exist, as exemplified by Ganji and Derakhshan (2020). While their study offers a valuable effort in developing a tool for appraising studies in AppLing, the tool requires further refinement to address two key limitations. Firstly, the tool's development process primarily relies on analyzing existing journal guidelines to formulate its questions. This method, while offering a starting point, requires further empirical validation. Secondly, the extensive list of questions in the tool could be cumbersome and time-consuming to work through, especially for systematic reviews that include a large number of included studies.

AppLing reviewers should also consider employing CATs that have been validated and used in previous reviews. These include the Mixed Methods Appraisal Tool (MMAT) (Hong et al., 2019) and JBL critical appraisal tools (Barker et al., 2023) for evaluating primary studies and the AMSTAR II tool by Shea et al. (2017) for appraising systematic reviews.

In summary, this section has highlighted the importance of considering the quality of evidence collected and reported in the SR, including both the level of confidence and performance on CATs. This is crucial because the conclusions and recommendations in SRs should be based on the quality of the evidence they present.

Reporting and Discussing Review Findings

Presenting the findings of a SR can be done in various ways, with common approaches including grouping findings by themes or by the original research questions. However, established research acknowledges that "there is no one right way" to present an SR (Siddaway et al., 2019, p. 765). Ultimately, the reviewer's choice of presentation method depends on two key factors: the goals of the review itself, and what the reviewer believes will be the most readable and navigable format for the intended audience.

When showcasing findings in a SR, it's essential to adhere to several general principles. Firstly, aim to go beyond simply summarizing existing findings. Instead, strive for a critical synthesis that builds upon the existing knowledge and provides new or enhanced understandings of the phenomenon under investigation. This means not only presenting what has been found in

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

previous research but also critically analyzing and interpreting these findings in the context of your study's objectives.

Secondly, prioritize a clear and logical structure for your presentation. A well-organized structure not only makes it easier for readers to follow your argument but also enhances the credibility of your findings. Consider using visuals like tables, figures, or graphs to present complex data in a more accessible format. These visuals can help to illustrate key points and trends, making your findings more engaging and understandable for your audience.

When reaching the Discussion section, reviewers have a vital responsibility to explicitly draw upon the evidence they have accumulated to support their conclusions and recommendations. As Siddaway et al. (2019) emphasizes, "[a]ny conclusions and recommendations for practice or policy should be based on the evidence and tempered by the flaws and weaknesses in the evidence" (p. 768). This highlights the importance of acknowledging the limitations of the evidence base while still drawing insightful conclusions for practical applications.

Addressing Bias in Systematic Review and Whether SR is Appropriate

It is first and foremost to reiterate the point that has been mentioned at the start of this chapter that literature review is not a neutral act, even for systematic review. There are steps that reviewers can take to further enhance the rigor and robustness of the SR. In addition to taking certain steps to ensure the four components of SR framework (i.e., comprehensiveness, reliability, replicability and quality assessment of included studies), scholars should also attempt to address potential bias in their review by considering these actions: (1) involving more than one member, (2) providing rich information into the steps, choices, and justifications, and (3) clarifying potential bias and/or background of the authors.

First, having more than one author in the review team allows for reflexive analysis and allow members to cross check one another interpretations to enhance objectivity in decision and interpreting the findings. Second, providing transparent procedure and justifications of their review choices can help enhance the transparency and trustworthiness of the review. Some steps that researchers can employ may involve providing their coding scheme, report coding

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

inter-rater reliability, and describing strategies for resolving disagreements in analysing and synthesizing studies (Chong & Plonsky, 2023).

Lastly, qualitative SRs involves qualitative content analysis where the theoretical views and background of the reviewers should be clarified. The reviewers should clarify potential bias due to their background, prior experiences and theoretical perspectives in conducting qualitative systematic reviews. For instance, Chong and Plonsky (2023) contend that “it is important to minimize bias by providing a brief statement on the reflective process of the reviewers”, citing Hanks (2019) as an example where the latter acknowledge his personal stance and discuss in the review how the reviewers’ perspective evolves as they interact with other researchers, professionals and doctoral students. The discussion of the reviewers’ background also enhance the transparency and trustworthiness of the findings and conclusions in the reviews.

Developing a Checklist for Appraising Studies in AppLing Systematic Review

There are several reasons why the use of multiple critical appraisal tools (CATs) can potentially introduce challenges in interpreting the quality of the evidence base in SRs. One obstacle is the inconvenience associated with using a diverse array of tools for different study designs within a single review. This creates a cumbersome and inefficient process for researchers conducting the review. Another challenge is the difficulty in comparing the quality of studies across designs. Since various CATs employ different criteria, it becomes challenging to draw accurate comparisons and effectively gauge their overall quality. The section proposes a study appraisal checklist that potentially helps address these limitations, offering a more streamlined and efficient approach for evaluating study quality in systematic reviews.

Table 2.

A Checklist for Evaluating Studies in Applied Linguistics

Checklist Questions	Yes	Partial Yes	No	Unclear	Not applicable
----------------------------	------------	------------------------	-----------	----------------	---------------------------

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

Q1. Are the study purposes and/or research questions clear, appropriate, and effective?

Q2. Is the literature review section effective in justifying the gaps and theoretical framework (if applicable) used in the study?

Q3. Is the research design appropriate and effective in addressing the study purpose(s) or research question(s)?

Q4. Is the selection and randomization of the participants if the study adopts experimental design appropriate and reliable?

Q5. Are the data collection methods appropriate, validated and effective for addressing the research questions?

Q6. Is the data analysis appropriate and follows established practices in the previous literature and takes into account confounding factors if applicable?

Q7. How well are ethical considerations addressed and documented in the study?

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

Q8. Is the result presented appropriately in a way that effectively addresses the research questions?

Q9. Does the discussion effectively link the existing study and its findings with previous literature and highlight its contributions?

Q10. Are the study limitations, confounders, and bias acknowledged and discussed in terms of their impact on the study findings or conclusions?

Q11. Was a systematic, transparent, and comprehensive literature search conducted?*

Q12. Were the screening and data extraction of studies performed in duplicate?*

Q13. Was quality appraisal performed for the included studies and potential bias of the individual studies taken into consideration when interpreting the findings of the review?*

*Note. questions with * are those only applicable for evaluating review studies.*

When appraising studies using the checklist, a 'Yes' answer signifies that the study demonstrates appropriate actions or justifications, a 'Partial Yes' indicates the reviewer's partial satisfaction, 'Unclear' signifies unidentifiable information, and 'Not Applicable' indicates the

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

question's irrelevance to the study. The three asterisked questions are reserved for reviews, not primary studies.

Calculating the score assigns 1 point for 'Yes', 0.5 points for 'Partial Yes', 0 points for 'Unclear', and excludes 'Not Applicable' from the total. This results in a maximum quality appraisal score (QA) of 10 for primary studies and 13 for reviews.

It is crucial to remember that the quality assessment scores are a proxy indicator, not an absolute measure, of a study's quality. They should be interpreted cautiously, considering how the study addresses the individual questions of the appraisal tool and its applicability to the specific research question. Similar to screening and data extraction, conducting quality assessments in duplicate is recommended to minimize bias.

Acknowledging that synthesizing a list of questions to evaluate studies may not be a neutral, objective process, the author wishes to clarify their background to provide context for this checklist. This aligns with the principles of addressing bias in systematic reviews, as discussed in the next section.

The author of this book chapter has prior education in AppLing and currently works as an evidence review specialist for a large Australian university. Their extensive experience includes conducting systematic reviews and utilizing diverse qualitative appraisal tools in both health science and social science fields. However, it is important to acknowledge that the proposed checklist, while synthesized from validated tools, has not itself undergone formal validation. This represents a potential limitation that should be addressed in future research.

Systematic Reviews in the Age of Generative Artificial Intelligence

The rise of generative artificial intelligence (GenAI) tools, such as OpenAI's ChatGPT, Microsoft's Bing Chat, and Google's Bard, is poised to revolutionize the systematic review process, offering potential improvements in efficiency and accuracy across various stages.

While current literature acknowledges limitations and emphasizes the need for human oversight, GenAI tools hold great promise in facilitating more efficient and resource-saving systematic reviews. Researchers have explored the use of GenAI chatbots in assisting with various stages

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

of the process, from formulating research questions and developing search strategies to data extraction, data analysis, and assessing risk of bias (Mahuli et al., 2023; Qureshi et al., 2023).

GenAI chatbots can prove valuable in several aspects of the systematic review process. It can significantly aid in formulating clear and precise research questions by suggesting synonyms and related terms, thereby facilitating the construction of effective search strings. Additionally, GenAI is capable of extracting key details from primary studies, such as mean values, standard deviations, and sample sizes, a capability demonstrated by Mahuli et al. (2023). Furthermore, GenAI can summarize and synthesize studies into coherent themes and patterns. For instance, a study by Nguyen-Trung (2024) tested ChatGPT-4's potential use for thematic synthesis, where the tool demonstrated speed and efficiency in generating codes, clustering codes, and developing key themes, while demonstrating the ability to continuously improve its thematic analysis skills. Moreover, GenAI can even be prompted to utilize specific tools for assessing the risk of bias, as shown in Mahuli et al.'s (2023) study. In terms of reporting the findings of the review, the chatbots can be employed to draft the structure for the review manuscript, offering researchers a valuable starting point.

However, it is important to recognize that GenAI does have its limitations. However, further research is needed to comprehensively evaluate its effectiveness and potential biases compared to traditional methods. While the generated summaries offer researchers a valuable starting point, they require human review and refinement for accuracy and context, as emphasized by Qureshi et al. (2023). Therefore, while GenAI can significantly enhance certain aspects of the systematic review process, it remains crucial for researchers to exercise caution and human judgment in its application, ensuring that the final outputs are accurate, relevant, and contextually appropriate.

Overall, GenAI serves as a valuable tool to enhance researchers' capabilities and streamline the systematic review process. By leveraging its strengths while acknowledging its limitations, researchers can conduct more thorough and efficient reviews, ultimately expediting scientific advancement.

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

Enhancing Applied Linguistics through Systematic Reviews: Policy and Practice

SRs hold immense potential to bridge the gap between research and practice in AppLing. By synthesizing and evaluating existing research evidence, SRs can inform both evidence-based policymaking and effective teaching practices in the AppLing field (Alsowat, 2020; Davies, 2000; Mitchell, 2000; Pachler, 2003).

On the one hand, SRs offer a more objective and robust foundation for policy development compared to solely relying on expert opinions or anecdotal evidence. Basing policies on the accumulated research evidence summarized in SRs fosters transparency and accountability in the policymaking process, leading to more informed decisions that ultimately benefit language learning experiences (Davies, 2000; Gillam Sandra & Gillam Ronald, 2006; Mitchell, 2000).

On the other hand, SRs empower educators and researchers in AppLing by providing them with readily accessible resources and insights. These resources include syntheses of effective teaching methods condensed and clarified from various pedagogical approaches, allowing teachers to readily apply evidence-based strategies in their classrooms (Davies, 2000). For instance, SRs on learner needs and diverse learning styles can inform teachers of the necessity to tailor their approaches to cater to the specific needs of different student groups, fostering personalized and effective learning environments. Additionally, SRs can serve as a method for teachers and researchers to stay informed about the latest advancements and emerging trends in the field, encouraging ongoing reflection and enhancement of their teaching practices (e.g., Mitchell, 2000). This ensures that educators remain updated on the latest knowledge and can adapt their teaching methods to best serve the evolving needs of their students.

It is crucial to acknowledge that teachers should also have the ability to independently pose and conduct systematic inquiries. Additionally, fostering collaboration between researchers and practitioners is essential. Foreign language educators should be equipped to appraise the quality of research evidence and make informed judgments about its application within their specific contexts (Mitchell, 2000; Pachler, 2003). Such capacity can be referred to as educational and research literacies (Pachler, 2003). Pachler (2003) also notes that “it is only by way of breaking up the false dichotomy of researcher versus practitioner that evidence-informed practice can become a sustained reality” (p. 13).

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

In conclusion, SRs serve as powerful tools for bridging the gap between research and practice in AppLing. As an effective and reliable evidence synthesis approach, they have the potential to contribute to improving language learning experiences and outcomes for learners of all backgrounds by informing evidence-based policymaking and empowering effective teaching practices, which are still in their infancy in AppLing.

Conclusion

In conclusion, this chapter provides a comprehensive roadmap for researchers embarking on systematic reviews in applied linguistics. By addressing key components such as research question formulation, search strategy design, study selection, data extraction, quality assessment, and reporting, this chapter equips scholars with the tools to conduct rigorous and insightful reviews. This chapter also introduces a checklist, developed based on the author's knowledge and experience of using various validated tools in health science. While tailored for the AppLing field, further validation is necessary. Additionally, methods for using GenAI for enhancing and streamlining the systematic review process have also been discussed. As researchers continue to embrace this systemic methodology, it is anticipated that systematic reviews will continue to shape and elevate the landscape of evidence synthesis in applied linguistics and beyond.

References

- Alsowat, H. H. (2020). Evidence-Based Practices of English Language Teaching: A Meta-Analysis of Meta-Analyses. *English Language Teaching*, 13(11), 75-93.
- Barker, T. H., Stone, J. C., Sears, K., Klugar, M., Leonardi-Bee, J., Tufanaru, C., Aromataris, E., & Munn, Z. (2023). Revising the JBI quantitative critical appraisal tools to improve their applicability: an overview of methods and the development process. *JBI Evid Synth*, 21(3), 478-493. <https://doi.org/10.11124/JBIES-22-00125>
- Bedle, S. (2017). *Reading strategies in an efl context: a mixed methods research synthesis* [Master's thesis, University of Glasgow]. https://www.teachingenglish.org.uk/sites/teacheng/files/steven_bedle_university_of_glasgow_dissertation.pdf

See the published version at:

Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>

- Beerens, M. (2018). Evidence-based policy and higher education quality assurance: progress, pitfalls and promise. *European Journal of Higher Education*, 8(3), 272-287. <https://doi.org/10.1080/21568235.2018.1475248>
- Briner, R. B., & Denyer, D. (2012). 112 Systematic Review and Evidence Synthesis as a Practice and Scholarship Tool. In *The Oxford Handbook of Evidence-Based Management* (pp. 0). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199763986.013.0007>
- Chong, S. W., & Plonsky, L. (2023). A typology of secondary research in Applied Linguistics. *Applied Linguistics Review*, 0(0). <https://doi.org/10.1515/applirev-2022-0189>
- Davies, P. (2000). Approaches to evidence-based teaching. *Medical Teacher*, 22(1), 14-21. <https://doi.org/10.1080/01421590078751>
- Ellis, R. (2015). Introduction: Complementarity in Research Syntheses. *Applied Linguistics*, 36(3), 285-289. <https://doi.org/10.1093/applin/amv015>
- Ganji, M., & Derakhshan, A. (2020). Developing a Checklist for Evaluating Research Articles in Applied Linguistics. *Teaching English Language*, 14(2), 239-268. <https://doi.org/10.22132/tel.2020.121858>
- Gillam Sandra, L., & Gillam Ronald, B. (2006). Making Evidence-Based Decisions About Child Language Intervention in Schools. *Language, Speech, and Hearing Services in Schools*, 37(4), 304-315. [https://doi.org/10.1044/0161-1461\(2006/035\)](https://doi.org/10.1044/0161-1461(2006/035))
- Glasziou, P. P., Chalmers, I., Green, S., & Michie, S. (2014). Intervention synthesis: a missing link between a systematic review and practical treatment(s). *PLoS Med*, 11(8), e1001690. <https://doi.org/10.1371/journal.pmed.1001690>
- Hanks, J. (2019). From research-as-practice to exploratory practice-as-research in language teaching and beyond. *Language Teaching*, 52(02), 143-187. <https://doi.org/10.1017/s0261444819000016>
- Heyvaert, M., Maes, B., & Onghena, P. (2011). Mixed methods research synthesis: definition, framework, and potential. *Quality & Quantity*, 47(2), 659-676. <https://doi.org/10.1007/s11135-011-9538-6>

See the published version at:

- Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>
- Hong, Q. N., Pluye, P., Fabregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M. P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M. C., & Vedel, I. (2019). Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. *J Clin Epidemiol*, *111*, 49-59 e41. <https://doi.org/10.1016/j.jclinepi.2019.03.008>
- Liu, Y., & Hu, G. (2024). Research Trends in Applied Linguistics (2017–2021): A Scientometric Review of 42 Journals. In H. Meihami & R. Esfandiari (Eds.), *A Scientometrics Research Perspective in Applied Linguistics* (pp. 45-81). Springer Nature Switzerland. https://link.springer.com/10.1007/978-3-031-51726-6_3
- Macaro, E. (2019). Systematic reviews in applied linguistics. In J. McKinley & H. Rose (Eds.), *The Routledge Handbook of Research Methods in Applied Linguistics* (1 ed., pp. 230-239). Routledge. <https://www.taylorfrancis.com/books/9781000734034/chapters/10.4324/9780367824471-20>
- Mahuli, S. A., Rai, A., Mahuli, A. V., & Kumar, A. (2023). Application ChatGPT in conducting systematic reviews and meta-analyses. *Br Dent J*, *235*(2), 90-92. <https://doi.org/10.1038/s41415-023-6132-y>
- Mansfield, L., Daykin, N., O'Connell, N. E., Bailey, D., Forde, L., Smith, R., Gifford, J., & Ashdown-Franks, G. (2024). A mixed methods systematic review on the effects of arts interventions for children and young people at-risk of offending, or who have offended on behavioural, psychosocial, cognitive and offending outcomes: A systematic review. *Campbell Systematic Reviews*, *20*(1), e1377. <https://doi.org/https://doi.org/10.1002/cl2.1377>
- Mitchell, R. (2000). Anniversary article. Applied linguistics and evidence-based classroom practice: the case of foreign language grammar pedagogy. *Applied Linguistics*, *21*(3), 281-303. <https://doi.org/10.1093/applin/21.3.281>
- Nguyen-Trung, K. (2024). ChatGPT in Thematic Analysis: Can AI become a research assistant in qualitative research? <https://doi.org/10.31219/osf.io/vefwc>

See the published version at:

- Cong-Lem, N. (2024). A comprehensive guide to conducting systematic reviews. In H. Bui (Ed.), *Considerations and techniques for applied linguistics and language education research* (pp. 115-135). IGI Global. <https://doi.org/10.4018/979-8-3693-6482-6.ch006>
- Norris, J. M., & Ortega, L. (2006). The value and practice of research synthesis for language learning and teaching. In *Synthesizing research on language learning and teaching* (Vol. 350).
- 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>
- Pachler, N. (2003). Foreign language teaching as an evidence-based profession? *The Language Learning Journal*, 27(1), 4-14. <https://doi.org/10.1080/09571730385200031>
- Qureshi, R., Shaughnessy, D., Gill, K. A. R., Robinson, K. A., Li, T., & Agai, E. (2023). Are ChatGPT and large language models "the answer" to bringing us closer to systematic review automation? *Syst Rev*, 12(1), 72. <https://doi.org/10.1186/s13643-023-02243-z>
- Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. *Annu Rev Psychol*, 70, 747-770. <https://doi.org/https://doi.org/10.1146/annurev-psych-010418-102803>
- Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol*, 8, 45. <https://doi.org/10.1186/1471-2288-8-45>