

UNIVERSITA' DEGLI STUDI DI VERONA

DEPARTMENT OF

Human sciences

DOCTORAL PROGRAM IN

Education

Cycle 38°

TITLE OF THE DOCTORAL THESIS

The evidence claims in the Systematic Reviews of qualitative studies in
education: A Framework for Reading and Understanding

S.S.D. PAED-02/B

Coordinator: Prof. Lorenzo Bernini

Signature _____



Tutor: Prof.ssa Marcella Milana

Signature _____



Doctoral Student: Dott./ssa Rita Marzoli

Signature _____



TABLE OF CONTENTS

ABSTRACT	4
INTRODUCTION	7
1 LITERATURE REVIEW	12
1.1 Educational research between evidence and interpretation: contexts and paradigms.....	12
1.2 Synthesis in qualitative educational research: purposes, possibilities, and challenges.....	16
1.3 Evidence claims and argumentation in reviews	23
1.4 Conceptual frameworks in qualitative systematic reviews	27
1.5 Gaps in the literature	30
2 THEORETICAL FRAMEWORK	33
2.1 An integrated conceptual framework	34
2.2 Toulmin’s <i>Model of argumentation</i>	36
2.3 Gough’s <i>Fitness for purpose of an evidence claim framework</i>	38
2.4 Justification for integrating Toulmin and Gough.....	42
3 METHODOLOGY	48
3.1 Operationalising the integrated conceptual framework	48
3.2 Categories derived from Gough’s framework.....	53
3.3 Categories derived from Toulmin’s model	55
3.4 Protocol development.....	59
3.5 Codebook development.....	61
4 CONSTRUCTING CLAIMS: A COMPARATIVE ANALYSIS OF SIX SYSTEMATIC REVIEWS	64
4.1 Search strategy and study selection.....	64
4.2 SR#1 – Facilitators and barriers to movement integration in elementary classrooms: a systematic review (Michael et al., 2019).....	70
4.3 SR#2 – Learning spaces, agency and notions of improvement: an interpretive meta- ethnography (Savin-Baden et al., 2008).....	71
4.4 SR#3 – Curriculum and learning in Aboriginal and Torres Strait islander education: a systematic review (Harrison et al., 2019).....	72
4.5 SR#4 CLIL teachers’ Needs and Professional Development: A Systematic Review (Kim & Graham, 2022)	74
4.6 SR#5 – Education policy and practice on intimate partner violence among young people in the UK: a systematic review (Sheng, 2020)	76

4.7	SR#6 – Non-native-English-speaking teacher identity in TESOL programs: a systematic review (Swearingen, 2019).....	78
4.8	Comparative synthesis across systematic reviews (SR#1–SR#6).....	80
5	DISCUSSION	87
5.1	The Nature of Evidence Claims in Qualitative Synthesis.....	87
5.2	Implicit Argumentation: Patterns and Consequences	89
5.3	Reflections on the Adapted Framework and Codebook.....	91
5.4	Synthesis Insights Across Reviews	93
	<i>Concluding remarks</i>	95
6	CONCLUSION	96
6.1	Summary of the Research	96
6.2	Key Findings	96
6.3	Contribution to Knowledge.....	97
6.4	Reflections on the Framework	98
6.5	Limitations	99
6.6	Directions for Future Research	100
6.7	Final Thoughts.....	102
	APPENDIX A - THE PROTOCOL	104
	APPENDIX B INTEGRATED CONCEPTUAL FRAMEWORK - CODEBOOK	113
	APPENDIX C CODEBOOK ANALYSIS	115
	REFERENCES	134
	<i>DISCLAIMER ON THE USE OF AI-ASSISTED TOOLS</i>	144

Abstract

La presente tesi nasce dal riconoscimento di una sfida persistente e di un ambito che richiede una comprensione più approfondita. Sebbene le revisioni sistematiche qualitative in ambito educativo producano affermazioni importanti e spesso influenti, esse raramente esplicitano i processi di ragionamento che le sostengono. I passaggi dai risultati dei singoli studi ai temi di sintesi, dalla sintesi alle affermazioni, e da queste ultime alle raccomandazioni di policy rimangono spesso impliciti o poco definiti. Questa opacità indebolisce sia la credibilità delle revisioni sia la loro utilità nei processi decisionali. Per colmare tale lacuna, la tesi sviluppa e applica un quadro concettuale integrato per l'analisi delle affermazioni di evidenza nelle revisioni sistematiche qualitative. Questo quadro combina due approcci complementari: il Modello di argomentazione di Toulmin (1958) e il Framework per la "fitness for purpose of an evidence claim" di Gough (2021). Il modello di Toulmin illumina la struttura logica degli argomenti, scomponendoli nelle diverse componenti del ragionamento: claim, dati, warrant, qualificatori e controargomentazioni. Il framework di Gough contestualizza tali elementi in rapporto allo scopo, al pubblico di riferimento e al contesto d'uso. Insieme, i due modelli permettono di mantenere una duplice attenzione sia alla coerenza interna dell'argomentazione, sia alla rilevanza esterna delle affermazioni di evidenza. Il disegno di ricerca prevede la selezione di un campione intenzionale di revisioni sistematiche qualitative in educazione, la loro analisi attraverso il framework integrato e una riflessione su ciò che questa analisi mette in luce rispetto a come l'evidenza viene costruita e comunicata. In questo modo, la tesi offre tre principali contributi:

1. Innovazione metodologica: propone uno strumento di valutazione critica delle revisioni sistematiche qualitative che va oltre le checklist di qualità esistenti.
2. Approfondimento teorico: avanza la comprensione delle dimensioni epistemologiche e argomentative della sintesi in educazione.
3. Rilevanza pratica: mette a disposizione di ricercatori, professionisti e decisori politici uno strumento per una valutazione mirata della leggibilità e dell'applicabilità delle affermazioni di evidenza che sorreggono i *claim*.

In ultima analisi, questo lavoro è guidato dalla convinzione che le revisioni sistematiche non siano contenitori neutri di conoscenza, ma pratiche argomentative che contribuiscono a definire ciò che viene riconosciuto come evidenza. Rendendo tali pratiche più visibili e

riflessive, la sintesi qualitativa in educazione può servire meglio sia la ricerca scientifica sia i processi decisionali democratici.

The present thesis arises from recognising a persistent challenge and an area requiring deeper understanding. While qualitative systematic reviews in education produce important and often influential claims, they rarely make explicit the reasoning processes that underpin them. The transitions from study findings to synthesis themes, from synthesis to claims, and from claims to policy recommendations are often left implicit or underspecified. This opacity undermines both the credibility of reviews and their usefulness for decision-making.

The thesis develops and applies an integrated conceptual framework for analysing evidence claims in qualitative SRs to address this gap. This framework combines two complementary traditions: Toulmin, (1958) *Model of argumentation* and Gough (2021) *Fitness for purpose of an evidence claim framework*. Toulmin's model sheds light on the logical structure of arguments, breaking them down into the various components that constitute reasoning: claims, data, warrants, qualifiers, and rebuttals. Gough's framework contextualises these elements in terms of their intended purpose, audience and context of use. Together, these two models enable a dual focus on both the internal coherence of argumentation and the external relevance of evidence claims.

The research design involves selecting a purposive sample of qualitative systematic reviews in education, analysing them through the integrated framework, and reflecting on what this reveals about how evidence is constructed and communicated. By doing so, the thesis contributes in three main ways:

1. Methodological innovation: offering a tool for the critical appraisal of qualitative systematic reviews that goes beyond existing quality checklists.
2. Theoretical insight: advancing understanding of the epistemological and argumentative dimensions of synthesis in education.
3. Practical relevance: it provides researchers, practitioners, and policymakers with a tool for targeted evaluation of the readability and applicability of the evidence claims that underpin their claims.

Ultimately, this project is motivated by the conviction that systematic reviews are not neutral containers of knowledge but argumentative practices that shape what is recognised as

evidence. By making these practices more visible and reflexive, qualitative synthesis in education can better serve both scholarly inquiry and democratic decision-making.

Introduction

Over the past two decades, there have been significant changes in the objectives and approaches of educational research. Researchers are increasingly expected to interpret the complexity of educational phenomena and contribute to decision-making processes by providing practitioners and policymakers with robust, evidence-based insights. This dual role reflects the broader 'evidence movement' trend, whereby research is expected to inform practice in a timely, reliable and relevant manner (Oakley, 2002). Systematic reviews in particular have gained prominence as tools for organising, evaluating and communicating educational knowledge (Gough et al., 2017) and are increasingly seen as a means of meeting the demands of evidence-based policy and practice.

A systematic review is defined as 'a review of existing research using explicit, accountable and rigorous research methods' and involves a sequence of structured activities: clarifying the research question, identifying, describing and appraising relevant studies, synthesising the findings into a coherent account and establishing what evidence claims can reasonably be made (Gough, Oliver & Thomas, 2017, pp. 5–6). These reviews are designed to reduce bias, promote transparency and enhance the reliability of conclusions drawn from multiple sources. They are grounded in the belief that examining what is already known is a critical step in reducing the risk of ineffective or even harmful interventions — what Chalmers (2003) described as the danger of 'doing more harm than good' when research is not properly reviewed.

However, the expansion of systematic review methodology into education has sparked critical debate. As Hammersley (2015) argues, the assumption that policy can or should be directly based on research is unrealistic and may distort the nature and purpose of educational enquiry. He points out that researchers, policymakers, and practitioners operate according to different rationales and constraints, meaning their relationship is inherently mediated and often conflictual. From this perspective, educational knowledge cannot simply be 'applied' in a straightforward, instrumental fashion — especially given the normative, contextual, and political dimensions of education that resist reduction to generalisable claims.

Furthermore, Biesta (2020) reminds us that the value of educational research cannot be judged solely by its instrumental utility. Instead, it must also be evaluated in terms of its contribution to democratic debate about the purposes and meanings of education. In this sense, the function of research is not merely to determine 'what works', but also to critically interrogate 'what is educationally desirable'. Accordingly, research synthesis becomes an epistemological and

ethical endeavour involving the construction of meaningful, context-sensitive knowledge claims rather than merely a technical exercise in aggregating findings.

Within this evolving landscape, this thesis is guided by the following research question:

What are the characteristics of knowledge claims made in qualitative systematic reviews in education, and how are these claims supported by the review authors?

To delve deeper into this *how*, the study examines the extent to which the reasoning that underpins these evidence claims is articulated explicitly or remains implicit within the reviews.

This research question lies at the heart of this thesis's purpose and intellectual commitment, as it addresses the fundamental issue of how knowledge claims are constructed, justified and communicated in qualitative systematic reviews in education. This concern underpins the entire analytical framework and motivates the study's methodological approach. As systematic reviews are increasingly used to inform policy and practice, there is a growing expectation that they should be both methodologically rigorous and epistemologically transparent. However, Gough, Oliver, and Thomas (2017) argue that this dual expectation often clashes with the epistemic diversity of the social sciences, where different traditions have different assumptions about what constitutes credible evidence, how it should be interpreted, and how it should be synthesised. In educational research, these tensions are exacerbated by the field's entanglement with normative values, political contexts, and lived experiences (Biesta, 2007; Hammersley, 2001).

Scholars such as Dixon-Woods, Agarwal, Jones, Young and Sutton (2005), and Thomas and Harden (2008), have emphasised the importance of making reasoning in qualitative syntheses more visible. They argue that the credibility of synthesis findings depends not only on procedural transparency, but also on the interpretive logic connecting data, concepts and conclusions. However, the interpretive nature of qualitative evidence means that this reasoning is often tacit and embedded in thematic choices or conceptual translations, rather than being formalised through clear argumentative structures. This makes it difficult for readers - whether researchers, policymakers, or practitioners - to assess how claims were constructed and what warrants their authority.

Therefore, investigating the explicit or implicit reasoning behind evidence claims becomes crucial for enhancing the trustworthiness, usability, and conceptual integrity of qualitative reviews. This issue also speaks to a broader concern about how the field of education positions

itself in relation to scientific authority and public accountability (Gough, 2021; Sandelowski & Barroso, 2007). In this context, transparency should not be reduced to checklist compliance, but rather understood as a reflexive process of making visible the assumptions, values and inferences that underpin knowledge claims.

As MacLure (2005) provocatively suggests, the pursuit of clarity in research communication, often celebrated as a marker of rigour, can also suppress the complexity, ambiguity and multiplicity intrinsic to educational enquiry. When reviews earnestly strive for simplified narratives or generalisable messages, they may mask the uncertain, entangled and politically charged dimensions of educational research. This thesis takes MacLure's call to resist overly coherent representations seriously, instead interrogating how claims to knowledge are made persuasive, whether by being explicit, strategically implicit, or somewhere in between.

This thesis proposes that a deeper understanding of the argumentative and justificatory structures in qualitative systematic reviews, in the educational field, can enrich their methodological quality and epistemological robustness.

To address these challenges, the aim of this thesis is twofold. First, it investigates how knowledge claims are constructed, justified, and communicated in systematic reviews that synthesise qualitative educational research. Second, it develops and applies a framework specifically designed to analyse these claims. The framework was developed through an iterative and reflexive process in which conceptual framing was not imposed from the outset but emerged through sustained engagement with theoretical literature, empirical data, and the researcher's evolving interpretive lens. This approach aligns with the reflexive logic of framework development described by Ravitch and Riggan (2021), who emphasise the value of constructing conceptual frameworks through a dynamic and situated inquiry process. In this study, the close reading and analysis of a purposive sample of systematic reviews informed and was simultaneously shaped by the gradual construction of the framework.

This framework results from the combination and adaptation of two established approaches: Toulmin's *Model of argumentation* (1958), which provides a structure for identifying the components and logical architecture of arguments, and Gough's *Fitness for purpose of an evidence claim framework* (2021), which focuses on assessing the appropriateness and contextual adequacy of evidence claims in relation to the stated goals of a review. Throughout this thesis, this combined approach will be referred to as the *integrated framework*.

By integrating Toulmin's analytical model with Gough's evaluative framework, the resulting tool enables a close reading of how claims are formulated and supported within systematic

reviews. It allows for the identification of key argumentative components - such as data, warrants, and qualifiers - and links them to the broader aims and context of the review. This makes it possible to trace how review authors justify their conclusions and to what extent these justifications are coherent with the epistemological and practical purposes of the review.

In this way, the thesis contributes both methodologically and conceptually. Methodologically, it offers a structured yet flexible tool for analysing argumentation and justification in qualitative syntheses. Conceptually, it enhances our understanding of how educational knowledge is constructed and legitimated through review processes. The framework developed here is applied to a purposive sample of systematic reviews, and the results are used to reflect on the transparency, reasoning, and contextual grounding of the evidence claims they produce.

Structure of the thesis

The thesis is organised into six chapters.

Chapter 1: Literature Review

This chapter examines the emergence and evolution of systematic reviews in the social sciences and education, with particular focus on the integration of qualitative research. It traces debates over methodological standards, the rise of evidence-based agendas, and the challenges of evaluating qualitative evidence within review frameworks initially designed for quantitative studies. The chapter also reviews existing frameworks and tools used to appraise systematic reviews and identifies a gap in approaches that address the reasoning behind evidence claims, particularly in the context of qualitative syntheses.

Chapter 2: Theoretical Framework

This chapter presents the conceptual foundations of the study by outlining Toulmin's *Model of argumentation* and Gough's *Fitness for purpose of an evidence claim framework*. Toulmin's model is introduced as a tool for analysing the internal structure of reasoning, while Gough's framework addresses the broader contextual and purposive dimensions of evidence use. The chapter explains the rationale for integrating these two approaches and describes how their combination serves as a lens for examining how systematic reviews of qualitative studies construct, justify, and communicate their knowledge claims.

Chapter 3: Methodology

This chapter details the research design, data sources, and analytical procedures adopted in the study. It describes the criteria for selecting a purposive sample of systematic reviews, the development and structure of the codebook derived from the integrated framework, and the

process of applying this tool through close reading and qualitative comparative analysis. Attention is given to issues of reflexivity, reliability, and transparency in the coding and interpretation of argumentative elements within the reviews.

Chapter 4: Results

This chapter presents the findings of the analysis, offering both individual case summaries and a comparative overview across the selected systematic reviews. It identifies patterns in how review authors formulate claims, construct warrants, use data, and qualify or limit their conclusions. The analysis reveals varying degrees of explicitness and coherence in argumentative structures, as well as differing uses of evidence depending on the review's stated purposes. Key thematic trends and divergences are highlighted to show how reasoning is shaped by methodological and contextual factors.

Chapter 5: Discussion

In this chapter, the results are interpreted in light of the conceptual framework and wider debates about qualitative synthesis. It reflects on the value of applying Toulmin and Gough's approaches to the study of systematic reviews, assessing their contribution to understanding the reasoning practices behind evidence claims. The discussion also considers the implications for review methodology, educational research quality, and the role of argumentation in evidence-informed decision-making. Limitations of the study are acknowledged, and the framework's potential for adaptation in other review contexts is explored.

Chapter 6: Conclusions

The final chapter summarises the main contributions of the thesis, highlighting how the integrated framework offers a new perspective for interrogating the reasoning underpinning evidence claims in qualitative systematic reviews. It discusses the relevance of this approach for both researchers and users of evidence, particularly in contexts where the purposes of reviews are varied and the logic of synthesis is interpretive. The chapter concludes by suggesting directions for future research, including the potential for further conceptual development and application of the framework in other domains of educational inquiry.

1 Literature review

Systematic reviews have become an increasingly prominent method for synthesising existing research in education and the broader social sciences. Originating within the health and medical sciences field, where evidence-based decision-making requires the aggregation of results from randomised controlled trials, systematic reviews have gradually evolved to encompass a wider variety of research traditions, including qualitative inquiry. As they are adapted across disciplines, the methodological assumptions, evaluative standards, and epistemological underpinnings of systematic reviews are also reshaped. This process has introduced both opportunities and tensions, particularly when synthesising interpretive, context-dependent, and theory-laden studies, as is often the case in educational research.

This chapter aims to situate the present study within the scholarly landscape of systematic review methodology, focusing on qualitative evidence syntheses in education. Section 1.1 addresses the relationship between evidence and interpretation in educational research; Section 1.2 examines how systematic review approaches have been adapted in the field of education. It traces the development of qualitative synthesis methods and considers the implications of these methods. Section 1.3 investigates how evidence claims in systematic reviews of qualitative studies are constructed, analysing the logics of synthesis and the role of interpretation in shaping review findings. Section 1.4 turns to the evaluation of these claims, drawing on methodological frameworks and argumentation theory to examine how claims are justified and warranted. Finally, Section 1.5 identifies key gaps in the literature that justify the development of a new integrative framework for reading and understanding evidence claims in systematic reviews of qualitative educational research.

1.1 Educational research between evidence and interpretation: contexts and paradigms

The way in which different research designs are valued and ranked reflects underlying assumptions about what constitutes legitimate knowledge in education. As Guba and Lincoln (1994) argue, every research tradition is based on specific ontological and epistemological principles that determine what constitutes credible evidence and how it should be produced. Within evidence-informed policy frameworks, particularly those influenced by post-positivist and rationalist assumptions, research tends to be evaluated according to its perceived methodological rigour. This has led to the creation of methodological hierarchies that rank research designs based on their evidentiary strength. Methods associated with causal inference

and control are at the top of these hierarchies, while interpretive or context-sensitive designs are positioned lower and are often regarded as less reliable or generalisable.

In the context of evidence-based paradigms, randomised controlled trials (RCTs) are often placed at the top of these hierarchies, seen as the most reliable means of establishing causal claims. Qualitative, interpretive, or case-based studies, by contrast, are typically positioned lower and viewed as less generalisable or objective (Slavin, 2002). These hierarchies reflect a post-positivist epistemology, which values control, prediction, and replication.

However, this ranking system is not without criticism. Scholars such as Biesta (2007) and Hammersley (2001) argue that applying these hierarchies indiscriminately in education, where values, context, and meaning are central, is both philosophically reductive and practically limiting. They call for recognition of epistemological pluralism, where the credibility of evidence depends not solely on method but on its appropriateness to the research purpose and context.

The rise of evidence-based education (EBE) has been a major force in shaping this methodological landscape. Drawing inspiration from evidence-based medicine, EBE promotes decision-making informed by research deemed methodologically rigorous, especially RCTs (Slavin, 2002). While this has encouraged greater attention to evidence quality and knowledge mobilisation, it has also reinforced narrow definitions of rigour and contributed to the marginalisation of research questions and approaches that do not align with experimental or aggregative logic.

Critics of this EBE movement have highlighted its conceptual limitations. For example, Biesta (2007) warns that focusing exclusively on 'what works' risks overlooking the more fundamental question of 'what is it supposed to work for?' (p. 6). From a pragmatist perspective, educational research must consider not only effectiveness, but also purpose, values and public interest. Similarly, Hammersley (2001) argues that even the most procedural forms of review cannot eliminate the need for judgement and interpretation. Decisions about which studies to include, how to define relevance and how to present findings are all inherently evaluative.

Adding further nuance to this debate, Taylor et al. (2021) compare several evidence frameworks employed in educational research, including the What Works Clearinghouse (WWC), Blueprints for Healthy Youth Development and Social Programs that Work (SPW). Their analysis reveals significant differences in how these frameworks define and

operationalise key elements such as intervention fidelity, study design and effectiveness. For example, the WWC provides outcome-specific effectiveness ratings based on fixed-effects meta-analyses of randomised controlled trials (RCTs) and quasi-experimental studies, whereas Blueprints includes dissemination readiness and researcher independence as formal evaluation criteria. SPW, on the other hand, emphasises replication and long-term outcomes, often excluding interventions that do not use RCTs or closely matched designs.

Importantly, Taylor et al. (2021) emphasise that these discrepancies can result in inconsistent, or even contradictory, ratings of the same intervention, depending on the framework employed. Such fragmentation can undermine trust in research syntheses and limit their usefulness to practitioners and policymakers. The authors advocate greater harmonisation and transparency across frameworks, proposing that evidence-based education standards should include clearer guidance on replication, contextual sensitivity and theoretical coherence, elements that are often underrepresented in procedural or checklist-driven approaches.

Pellegrini & Vivaret (2020) offer a complementary perspective, critically examining the implementation of evidence-informed policy across European contexts. While supporting the broader aim of basing educational policy on research, the authors caution that efforts remain fragmented and inconsistent. They prefer the term 'evidence-informed' to 'evidence-based' to signal a more flexible, pluralistic approach that accommodates methodological diversity and context-aware reasoning. Drawing in particular on the Italian context, they emphasise how systemic challenges, such as limited national investment in educational research, weak infrastructures for knowledge mobilisation and insufficient alignment between research and policy, have hindered the development of coherent, research-informed educational strategies. More broadly, they highlight other barriers, including a lack of coordination among initiatives and difficulties in measuring complex competencies (e.g. creativity, citizenship and critical thinking). These observations resonate with broader critiques of Italy's educational research system, which has historically suffered from fragmented funding, a lack of continuity in reform efforts and limited institutional support for bridging the gap between research and practice (Benadusi & Giancola, 2016). Pellegrini and Vivaret (2020) argue that evidence cannot be treated as a universally applicable 'toolkit', but must be interpreted, adapted and mediated for different cultural, institutional and pedagogical contexts. This further challenges the neutrality and standardisation assumptions in the design and use of systematic reviews.

Significant contributions to this field also come from the Italian context, notably through the work of (Ghirotto, 2020). Ghirotto calls for a critical distinction between reviews and

syntheses, emphasising that while reviews involve the selection and aggregation of data, syntheses entail a genuine interpretive process aimed at producing new theoretical understandings. Ghirotto notes that, although reviews are typically viewed as aggregative activities, synthesis can also be interpretative, especially when guided by a research question that requires the construction of meaning. This distinction is particularly pertinent in education, where systematic reviews should not merely be procedural classifications of evidence; rather, they must explicitly demonstrate the reasoning processes through which findings are selected, interpreted and related to one another. In this sense, Ghirotto proposes that qualitative reviews should be accompanied by both methodological transparency and conceptual coherence in order to meaningfully contribute to educational knowledge.

To see systematic reviews as neutral or apolitical is to ignore the layered judgments, selections and interpretations involved. Although they are often presented as objective procedures governed by transparent criteria and replicable methods, this can mask the paradigm-dependent reasoning underlying every stage of the review process. As Hammersley (2001) argues, even the most procedural and technical decisions, such as setting inclusion criteria, defining relevance or coding findings, are infused with theoretical and epistemological assumptions. These assumptions reflect broader commitments regarding what constitutes legitimate knowledge, what constitutes evidence and which methods are considered more credible or authoritative within a given field of study.

Furthermore, Dixon-Woods et al. (2006) and MacLure (2005) highlight that systematic reviews are never conducted in isolation; they are embedded in institutional, political and policy contexts that influence the research questions and how the findings will be utilised. In education, for instance, the demand for reviews to inform policy decisions or demonstrate 'what works' often prioritises quantifiable outcomes and 'rigorous' methodological studies, often at the expense of more contextualised, qualitative or critical perspectives. This can lead to the marginalisation of research grounded in interpretivist paradigms, feminist theory, or postcolonial critique. This is not because such work lacks value, but because it does not conform to prevailing standards of reviewability or evidence hierarchies.

MacLure (2005) reminds us in particular that the rhetoric of neutrality - the idea that systematic reviews can serve as impartial aggregators of existing knowledge - can function as a discursive mechanism that suppresses the tensions, ambiguities and interpretive work inherent in synthesising complex and diverse studies. If treated as purely technical exercises,

systematic reviews may contribute to epistemic closure, where only certain types of research and reasoning are rendered visible and legitimate.

Recognising that systematic reviews are constructed, situated and value-laden does not undermine their utility; rather, it calls for greater reflexivity and transparency in their design, conduct and interpretation. This also creates space for alternative frameworks, such as those explored in this thesis, that consider the argumentative, theoretical and contextual aspects of knowledge claims in qualitative research synthesis. Only by acknowledging these dimensions can systematic reviews in education become procedurally, epistemologically, and ethically rigorous.

1.2 Synthesis in qualitative educational research: purposes, possibilities, and challenges

Educational research has long been intertwined with qualitative traditions that prioritise meaning-making, context and the perspectives of those involved. Qualitative inquiry in education is rooted in interpretivist and constructivist paradigms. It studies learning and teaching as situated social practices, paying attention to the lived experiences of students and teachers, the mediating role of culture and institutions, and the value-laden nature of schooling. Typical research designs, such as ethnography, case studies, narrative inquiry, grounded theory and phenomenology, favour naturalistic settings and the iterative and abductive logic of analysis. They also emphasise thick descriptions that connect local accounts to conceptual insights (Denzin & Lincoln, 2018; Merriam & Tisdell, 2017; Stake, 1995). These characteristics have direct implications for synthesis: rather than merely aggregating results, qualitative syntheses frequently aim to configure and reinterpret findings to explain how and why educational phenomena take shape under particular conditions, and with what meanings for participants. Concerns about the usability of educational research have also been addressed through reporting guidance, such as the REPOSE Guidelines, which emphasise that readers need clear information about aims, context, methods, and conclusions in order to judge what findings are “valid for” and usable in practice (Newman & Elbourne, 2004).

The need to synthesise research findings in education has become central to the call for more transparent, accessible, and policy-relevant evidence (D. Gough et al., 2019) Gough et al., 2017; Petticrew & Roberts, 2008). The early development of systematic reviews was dominated by models from evidence-based medicine, particularly through the work of the

Cochrane Collaboration¹, which promoted procedures designed to ensure replicability, control bias, and support statistical aggregation. These models informed parallel initiatives in education, such as the Campbell Collaboration², which adapted Cochrane’s standards for the social sciences. Although these frameworks contributed significantly to improving methodological rigour, they were designed primarily for synthesising quantitative data and evaluating the effectiveness of interventions through standardised outcomes. However, as many researchers have noted, the transfer of these models to education has revealed important limitations. Educational research often addresses open-ended, value-laden, and context-dependent questions that cannot be reduced to “what works” formulations. In response, a growing movement has emerged to include qualitative research in systematic reviews, not merely as contextual or illustrative evidence, but as a source of conceptual insight in its own right (Dixon-Woods et al., 2006; Sandelowski & Barroso, 2007). This shift entails significant methodological and epistemological implications.

In these settings, systematic reviews aim to minimise bias by applying structured search strategies, transparent inclusion criteria and formalised appraisal techniques (Petticrew & Roberts, 2008). However, as these methods migrated into the social sciences - and in the educational research area addressed in this study - they encountered a complex epistemological terrain. Unlike the standardised experimental designs typical of medical research, educational research is often interpretive, context-sensitive and theoretically diverse. This diversity has necessitated the substantial adaptation of systematic review methodology (Gough et al., 2017).

In educational research, systematic reviews are increasingly recognised as processes of knowledge production deeply embedded in values, perspectives and socio-political contexts. As Guba and Lincoln (1994) argue, what counts as credible evidence is shaped by ontological and epistemological assumptions. Efforts to impose hierarchies of evidence risk favouring certain paradigms - typically post-positivist - at the expense of others. For instance, randomised controlled trials (RCTs), which are often considered the gold standard in evidence-based education (EBE), may not be suitable for addressing the interpretive and exploratory objectives of qualitative research (Slavin, 2002). Critics such as Biesta (2007)

¹ Cochrane (founded in 1993) is a global, independent network that produces and maintains systematic reviews—primarily in health and medicine—to inform evidence-based decision making. Reviews follow pre-registered protocols, comprehensive search and appraisal procedures, and standardized reporting; they are published in the Cochrane Library (<https://www.cochrane.org/about-us>).

² Campbell Collaboration (founded in 2000) is an international network that produces and disseminates systematic reviews of social and educational interventions to inform policy and practice (<https://www.campbellcollaboration.org/>).

and Hammersley (2001) argue that such hierarchies can be epistemologically reductive and practically limiting, particularly when applied uncritically to educational issues requiring context- and purpose-sensitive reasoning.

This concern is particularly pertinent in qualitative synthesis, a key strategy in educational research amid growing calls for evidence-based policy and practice. Unlike quantitative syntheses, which seek statistical aggregation, qualitative syntheses aim to interpret, re-contextualise and conceptualise findings across studies. Gough et al. (2017) distinguish between two major synthesis traditions: aggregative reviews, which use pre-specified frameworks to summarise findings; and configurative reviews, which seek to generate new insights through interpretation and reconfiguration. While not mutually exclusive, these traditions reflect different assumptions about what systematic reviews are meant to achieve. In educational research, configurative approaches—more aligned with qualitative and critical traditions—are often better suited to understanding how, why, and under what conditions phenomena occur (Zawacki-Richter et al., 2020).

A variety of methods have emerged to support the interpretive approach to synthesising qualitative research, each offering different strategies for dealing with complexity, meaning and conceptual development. One of the most significant contributions to this field is meta-ethnography, which was developed by Noblit and Hare (1988). Rather than simply summarising the findings of individual studies, meta-ethnography involves a process of 'translating' concepts and metaphors to generate a new, higher-order interpretation. This translation is conceptual rather than literal and assumes that meaning is co-constructed by researchers and participants, and is always shaped by the contexts in which the original studies were conducted. The outcome is a synthesis that is greater than the sum of its parts, providing new theoretical insights that are grounded in empirical data.

Building on this interpretive foundation, (Sandelowski & Barroso, 2007) proposed a typology of qualitative synthesis methods, distinguishing between metasummaries and metasyntheses. Metasummaries focus on aggregating and quantifying the frequency of findings across studies. They are more descriptive in nature and aim to capture the prevalence of certain themes or categories. Metasyntheses, by contrast, involve a deeper level of interpretation and conceptual integration. They aim to produce novel theoretical constructs or understandings that emerge through the comparative and reflexive reading of multiple qualitative texts. This typology has proven influential in educational research,

where studies often vary in their methodological approaches and conceptual depths, making it essential to distinguish between different types of summary and synthesis.

Another widely used method is thematic synthesis, which was developed by (Thomas & Harden, 2008). This approach combines techniques from thematic analysis, which is commonly used in primary qualitative research, with principles from grounded theory, such as constant comparison and inductive coding. Thematic synthesis typically proceeds in three stages: (1) coding text line by line; (2) developing descriptive themes; and (3) generating analytical themes that move beyond the content of the original studies. This layered approach combines systematic procedural transparency with interpretive richness, making it particularly suited to reviews intended to inform educational policy or practice. Its structured yet flexible process has been successfully applied in areas such as student motivation, curriculum reform and inclusive education.

These methods all share a commitment to preserving the interpretive nature of qualitative research while equipping researchers with the tools to synthesise diverse findings transparently and in a conceptually rigorous manner. They all represent a response to the challenge of producing credible, transferable knowledge from studies that are often situated, heterogeneous and grounded in different theoretical traditions.

However, the increasing number of synthesis methods also presents challenges. As Zawacki-Richter et al. (2020) emphasise, while methodological pluralism is important, it can introduce ambiguity regarding quality appraisal and the criteria for evaluating evidence claims. Configurative syntheses often resist standardisation, and conventional appraisal tools may not adequately capture the interpretive reasoning involved in constructing findings. Reviewers must therefore consider not only rigour and coherence, but also make explicit the theoretical, epistemological and ethical choices that influence their work (Dixon-Woods et al., 2006; MacLure, 2005).

This complexity highlights the need for frameworks that support procedural transparency and conceptual integrity. Several such frameworks have been proposed. Gough's (2007) Weight of Evidence (WoE) model, for example, offers a multidimensional appraisal strategy that incorporates methodological quality, appropriateness to the review question and relevance to the topic. This flexible model enables more nuanced judgements to be made across heterogeneous studies—an especially important feature in education, where reviews often bridge multiple paradigms. More recently, standards such as GRADE-CERQual (Lewin et al., 2018) and ENTREQ (Tong et al., 2012) have introduced criteria for assessing

confidence in qualitative review findings, while the RETREAT (Booth et al., 2016) framework provides guidance on selecting synthesis methods that align with the purpose of the review.

Taylor et al. (2021) shed further light on the subject by comparing how various organisations, including the What Works Clearinghouse (WWC), Blueprints for Healthy Youth Development and Social Programs that Work, evaluate evidence in education. Their analysis highlights methodological differences, such as criteria for replication, independence, and long-term effects, and the consequences of inconsistent standards for end users. Interventions may receive different ratings depending on the framework applied, which could confuse policymakers or practitioners. The authors advocate greater harmonisation across frameworks, calling for standards that distinguish between study rigour and intervention effectiveness, emphasise implementation fidelity, and permit domain-specific ratings. These recommendations are highly relevant to the use of qualitative evidence in education.

Despite these advances, however, several persistent tensions remain. Firstly, qualitative synthesis must address how to evaluate and combine studies with different theoretical perspectives. Secondly, many reviews still fail to clearly articulate how interpretive claims are generated and justified, which can undermine their credibility and usefulness (Dixon-Woods et al., 2006). Thirdly, as Booth et al., Sutton and Papaioannou (2016) argue, systematic reviews must strike a balance between breadth and depth, transparency and reflexivity, and aggregation and interpretation.

As Hannes & Macaitis (2012) observes, integrating qualitative evidence into systematic reviews requires the development of new tools and logics. Unlike quantitative synthesis, which often seeks convergence, qualitative synthesis values variation, interpretation, and conceptual abstraction. This difference is not merely a matter of technique: analyses of systematic reviews suggest that narrative syntheses and meta-analyses operate through incommensurate forms of argumentation, with different standards for what counts as persuasive support for a conclusion (Melendez-Torres et al., 2017).

Early methodological discussions already highlighted both the promise and the practical difficulties of incorporating qualitative studies into systematic reviews, particularly regarding appraisal, relevance judgements, and synthesis across heterogeneous designs (Dixon-Woods et al., 2001). The work of Dixon-Woods et al. (2006), drawing on an ESRC-funded project, highlights that incorporating qualitative findings challenges every stage of

the review process—from formulating review questions to assessing relevance, appraising quality, and synthesising results. They argue that this is not a matter of technical adjustment but of fundamentally rethinking what it means to review and integrate evidence. In contrast to the aggregative logic of quantitative meta-analysis, qualitative synthesis typically follows a configurative approach. Methods such as meta-ethnography, critical interpretive synthesis, and thematic synthesis aim to build conceptual frameworks, develop theories, or challenge prevailing assumptions (Thomas & Harden, 2008; Dixon-Woods et al., 2006). This diversity is reflected in published review reports, which vary substantially in aims, methods, and levels of analytic transparency, reinforcing the need for clearer conceptual and methodological tools for synthesis and appraisal (Dixon-Woods et al., 2007). Recent methodological work in education confirms both the rapid growth and the heterogeneity of qualitative synthesis approaches, highlighting ongoing challenges in transparency, reporting, and methodological consistency that directly affect how synthesis claims can be interpreted and appraised (Maeda et al., 2022). Computational mapping of education-related research territories also shows that thematic emphases and key terms change over time, with consequential shifts in how problems are framed and how knowledge claims circulate, reinforcing the need for synthesis approaches that are sensitive to disciplinary language and conceptual drift (Nylander et al., 2024).

As Sandelowski and Barroso (2007) note, qualitative metasynthesis is a “construction of constructions”—a layered interpretive process in which reviewers engage critically with already-interpretive findings. Rather than generalisability, the emphasis is on trustworthiness, resonance, and conceptual transferability. This shift also implies a different understanding of validity: from asking whether a synthesis is valid in an absolute sense to asking what it is valid for, and for whom—a purposive, context-sensitive framing consistent with social constructionist accounts of qualitative validity (Aguinaldo, 2004). Moreover, these interpretive approaches resist the idea that synthesis should necessarily standardise or simplify findings. Instead, they embrace complexity and positionality. For example, reviews rooted in critical paradigms may seek to expose silences in the literature, amplify marginalised perspectives, or interrogate ideological assumptions embedded in primary studies. The aims of synthesis, therefore, extend beyond summary—they include critique, re-framing, and theorisation. The expanding body of synthesis work has also led to the emergence of overviews of reviews or “second-order” syntheses, which aim to integrate findings across multiple systematic reviews (Gessler & Siemer, 2020). As Gough and Thomas (2016) point out, this layering brings new challenges for methodological clarity and

epistemological coherence. Integrating reviews that rest on different paradigms or conceptual frameworks raises fundamental questions about what counts as valid synthesis and how different forms of evidence can, or should, be combined. This thesis responds to these developments by offering a framework for critically appraising the evidence claims made in systematic reviews of qualitative studies in education. It is grounded in the recognition that synthesis is not merely a technical process, but a deeply interpretive one, shaped by the reviewer's assumptions, the paradigms underpinning the included studies, and the communicative strategies used to justify claims. Because qualitative synthesis is a situated interpretive practice, evaluating review claims also requires attention to how perspectives, values, and power relations shape what is made visible and what is marginalised in the construction of knowledge (Reinharz et al., 2006). Rather than attempting to define fixed hierarchies of method or certainty, this study seeks to understand how reviews construct credibility, and to provide tools for evaluating the fitness for purpose of those constructions within their intended contexts.

Qualitative systematic reviews are not neutral containers of evidence but interpretive knowledge products: they translate first-order accounts (participants' experiences as represented in primary qualitative studies) into second-order interpretations (authors' analytic constructs, themes, and explanations), and then reconfigure these into a synthesis that has its own conceptual architecture. In this sense, the systematic review can be conceptualised as a third-order interpretive construct, emerging through an iterative dialogue between primary empirical materials and the interpretive work already performed in the included studies (Toye et al., 2014). Rather than simply "adding up" findings, the review re-contextualises and re-signifies them through processes such as comparison, translation, abstraction, and theoretical framing, producing claims whose meaning and force depend on how the synthesis is organised and argued (Noblit & Hare, 1988). This interpretive layering is particularly salient in education, where review claims often aim to offer explanatory accounts, conceptual reframings, or normative critiques rather than discrete, decision-ready propositions (Sandelowski & Barroso, 2007; Thomas & Harden, 2008).

Positioning the review as a third-order construct also has methodological implications for the present study. If review outputs are interpretive arguments rather than direct representations of an underlying reality, then evaluating them requires attention to how claims are constructed, warranted, qualified, and bounded (Toulmin, 1958; Gough, 2021). The analytical framework developed here, therefore, treats the systematic review as an

authored argumentative text and examines the inferential moves through which review authors transform synthesised interpretations into evidence claims addressed to particular audiences and purposes.

This approach makes the “step in between” visible: it shows how review authors move from a set of synthesised themes to a conclusion that is presented as stronger, broader, or more usable than the underlying interpretations. Concretely, it clarifies (a) what the review treats as evidence for the conclusion (e.g., recurring themes, convergent findings, illustrative quotations from primary studies, or theoretical alignment), (b) the reasoning rule that links that evidence to the claim (the warrant, such as “recurrence implies robustness” or “cross-context convergence implies transferability”), and (c) the assumptions that allow the authors to describe the conclusion as robust, relevant, or practically useful. At the same time, it makes visible how review authors handle interpretive limits: whether they qualify their claims through hedging (“may,” “suggests,” “in some contexts”), specify conditions of applicability (for whom and under what circumstances the claim is likely to hold), and acknowledge competing interpretations or contradictory evidence. These moves are particularly important in interpretive synthesis, where conclusions are constructed through translation and judgement rather than produced by simple aggregation (Sandelowski & Barroso, 2007; Noblit & Hare, 1988).

Overall, qualitative synthesis in education is a dynamic and evolving field that reflects the epistemological and methodological diversity of the discipline itself. Moving beyond checklist-based approaches, the field requires frameworks that integrate rigour with reasoning, and transparency with conceptual richness. Whether through meta-ethnography, thematic synthesis, or framework tools for analysing, such as Gough's WoE or CERQual, a commitment to clarity, reflexivity, and a situated understanding of what constitutes evidence and why remains essential. This thesis adopts these principles in its examination of how qualitative systematic reviews construct and justify their claims, and how these claims can be better understood, evaluated, and applied in the field of education.

1.3 Evidence claims and argumentation in reviews

As reviews of qualitative research become more prevalent in education, greater attention is being given to the nature and justification of the conclusions drawn from these syntheses. According to Gough et al. (2017), an evidence claim is a statement made by a review about what the studies it has examined indicate in relation to its research question. In other words, it is a conclusion that is supported by the review's processes of searching, selecting,

appraising and synthesising the studies, and is presented in a way that is appropriate for a specific audience and purpose. Unlike quantitative reviews, where claims are often supported by statistical inference, qualitative syntheses rely on conceptual reasoning, interpretive logic and narrative coherence. This makes evaluating evidence claims particularly challenging as it requires attention to both the content of the claim and the process by which it is constructed, justified and communicated. The key question is what constitutes a trustworthy or believable claim in the context of a qualitative synthesis (D. Gough et al., 2020).

In this dissertation, “evidence claims” refer to the review authors’ synthesised assertions that extend beyond the reporting of descriptive themes to propose an interpretive conclusion, an explanatory account, a practical implication, or a normative position, presented as warranted by the body of qualitative evidence assembled in the review. An evidence claim is therefore not simply a summary of included studies; it is a communicative and inferential move through which reviewers translate synthesised material into a statement that has an intended use, audience, and scope. This concept is central to the study because the main analytical objective is to examine how qualitative systematic reviews in education construct such claims: how they frame what is being claimed, what grounds are offered, what inferential warrant connects the grounds to the conclusion, how strength is qualified, and how limitations or boundary conditions are acknowledged. Treating evidence claims as the unit of analysis allows the dissertation to move beyond procedural assessments of review “quality” and instead evaluate the reasoning and fitness-for-purpose through which qualitative syntheses become actionable, persuasive, or conceptually generative.

Having defined an evidence claim as a warranted conclusion drawn from a review of what the assembled studies indicate, the next task is to consider how such claims may be evaluated. Gough's Weight of Evidence (WoE) framework (2007) provides an initial point of reference, judging the contribution of primary studies based on three criteria: A. methodological quality; B. appropriateness to the review question; and C. relevance to the topic. These criteria are then combined into an overall judgement. In WoE, the unit of appraisal is the individual study and how well it fits the review's purposes.

Subsequent work extends this logic from the level of primary studies to the level of the review’s own conclusions. In Gough’s later formulation (2021), the focus of appraisal is on the fitness for purpose of the review’s evidence claims, considering whether the claims are adequately warranted, transparent in their reasoning and appropriate for informing decisions,

given the review's aims, audiences and intended uses. The emphasis therefore shifts from methodological checklists to the explicit structure and justification of the argument advanced by the synthesis. Gough's *Fitness for Purpose of an Evidence Claim Framework* (D. Gough, 2021) extends this logic by shifting the evaluative focus from individual studies to the review's own evidence claims. In this later formulation, the central question is a claim's fitness for purpose: whether, given the review's aims, audience, and context of use, the claim is adequately warranted, transparent in its reasoning, and appropriate to the decisions it is meant to inform. The emphasis thus moves beyond methodological checklists to the structure and justification of the argument the review advances.

Evidence in qualitative reviews is rarely presented as stand-alone findings. Instead, it is embedded in a chain of reasoning that moves from the review question, through study selection and data extraction, to thematic synthesis or conceptual interpretation. However, as several scholars have noted, the logic connecting these steps is often left implicit (Dixon-Woods et al., 2006). Reviews may report themes or conclusions without clarifying how they emerge from the data or how they are warranted. Reviews may report themes or conclusions without clarifying how they emerge from the data or how the included studies warrant them. This lack of transparency limits the reader's ability to evaluate the claims' credibility, relevance, and transferability.

To address this gap, some scholars have drawn on models from argumentation theory, particularly Stephen Toulmin's (1958) work, to examine how claims are constructed and justified in research writing. Although initially developed outside the context of systematic reviews, Toulmin's model has since been applied to the analysis of scientific reasoning, including the logic of evidence synthesis in qualitative reviews (Zawacki-Richter et al., 2020). Toulmin's (1958) model distinguishes between several components of an argument: the claim (the conclusion being argued for), the data (the evidence supporting it), the warrant (the principle connecting data to claim), the backing (support for the warrant), qualifiers (indicating the strength of the claim), and rebuttals (acknowledging counter-arguments). This model provides a valuable lens for identifying the internal logic of a review's conclusions, especially in qualitative syntheses where reasoning is more interpretive than inferential. Although Toulmin's model was not initially developed to appraise research, it has been adopted in recent years to analyse scientific texts and improve the transparency and rigour of argumentation in systematic reviews (Gough, 2021). In the context of education, Toulmin's framework can be beneficial for revealing the assumptions and reasoning structures that underpin claims about teaching practices, student experiences, or institutional

contexts. It also allows reviewers and readers alike to scrutinise what is being claimed and how and why those claims are persuasive within the text.

A further complication arises from the diversity of qualitative studies typically included in educational reviews. Syntheses may draw on research conducted in vastly different cultural, institutional, or theoretical contexts. As a result, reviewers must navigate conceptual translation, theoretical coherence, and contextual sensitivity questions. These considerations affect how claims are constructed and how much they can be generalised or transferred. Gough (2021) addresses this issue by proposing the notion of “fitness for purpose” as a guiding principle: that is, evidence claims should be judged in light of the review’s aims, users, and intended applications, rather than against a universal standard of validity.

Taken together, these perspectives suggest that synthesis in education cannot be evaluated simply based on following procedural steps. What is required are practices that are both methodologically rigorous and argumentatively transparent. Methodological rigour refers to clear and systematic processes for searching, selecting, and analysing studies, ensuring the synthesis is replicable and credible. Argumentative transparency, by contrast, refers to making visible the reasoning through which evidence is transformed into claims, how findings are linked, warranted, qualified, and presented to readers. In qualitative reviews, where statistical generalisation is not possible, the trustworthiness of conclusions depends on this dual foundation: robust methodological processes and explicit argumentative pathways that connect data, interpretation, and claims. Without these, readers are left unable to assess the credibility, relevance, or transferability of the evidence being presented.

Yet, as Booth et al. (2016) and Harden and Thomas (2005) point out, current guidance for reviewers tends to prioritise procedural checklists over reflective reasoning. This creates a methodological and epistemological gap: while procedures may ensure consistency, they do not address the interpretive work by which claims are constructed and justified.

To respond to this gap, the present study brings together Gough’s (2021) principle of evaluating evidence claims in terms of their “fitness for purpose” with Toulmin’s (1958) *Model of argumentation*. This integrated approach seeks to provide a framework through which readers can critically examine how qualitative systematic reviews in education construct, warrant, and communicate their claims, making the underlying logic of synthesis more visible and open to evaluation.

1.4 Conceptual frameworks in qualitative systematic reviews

Conducting a qualitative systematic review is not just a technical or procedural exercise; it is also a conceptual endeavour. Conceptual frameworks influence every stage of the synthesis process, from defining evidence and extracting and comparing data to constructing themes and warranting conclusions for specific audiences and applications. While quantitative reviews often rely on statistical models or effect-size estimation (Borenstein, 2009), qualitative reviews depend on frameworks attuned to interpretive reasoning, contextual complexity and theoretical development.

In this context, a conceptual framework is a structured set of assumptions and concepts that organise the logic of the review: it defines what counts as evidence, how meaning is generated across studies, and the basis on which claims are justified. It functions as both an analytic scaffold and a lens for appraisal and interpretation. However, the literature suggests that many qualitative reviews in education either omit such frameworks or apply them only loosely. This results in limited transparency regarding the epistemological and inferential basis of their conclusions (Booth et al., 2016; Dixon-Woods et al., 2006).

In response to this gap, several frameworks have been proposed to support the appraisal of qualitative research and the synthesis of its findings. One of the most widely used is the Critical Appraisal Skills Programme (CASP, 2018), which provides checklists for assessing the credibility, relevance, and rigour of individual qualitative studies. While useful as a preliminary quality filter, CASP offers limited guidance for evaluating how claims are constructed across multiple studies, and it does not directly address the problem of conceptual coherence in synthesis. Alongside tools for appraising primary qualitative studies, methods have also been proposed for judging the quality of review reports themselves, including narrative reviews, for which the International Narrative Systematic Assessment (INSA) tool offers structured criteria for transparency and reporting (La Torre et al., 2015). In addition to general appraisal tools, education subfields have also proposed domain-specific quality indicators for reviews—for example in special education—typically emphasising methodological and reporting standards. This highlights the ongoing need to complement procedural criteria with approaches that examine how review-level evidence claims are constructed and justified (Talbot et al., 2018).

Debates persist, however, about whether qualitative research should be “quality appraised” for review purposes in the same way as quantitative studies, and, if so, what criteria can be justified without reducing interpretive work to procedural checklists (Garside, 2014).

Critical appraisal is increasingly framed as a context-sensitive form of judgement rather than a purely technical checklist exercise, requiring explicit attention to what counts as credible evidence for particular purposes and audiences (Tod et al., 2022).

However, these tools largely focus on evaluating primary study reports, whereas the present thesis focuses on how review authors construct, warrant, and qualify evidence claims at the synthesis level (Munthe-Kaas et al., 2019).

A more comprehensive and increasingly influential approach is GRADE-CERQual (Confidence in the Evidence from Reviews of Qualitative Research) (Lewin et al., 2018). CERQual assesses confidence in individual review findings through judgments across four components - methodological limitations, relevance, coherence, and adequacy of data - thereby strengthening transparency in how reviewers justify the degree of confidence they assign to synthesised findings. However, CERQual has also been criticised for potentially reintroducing hierarchical notions of evidence, for limited engagement with the interpretive and theory-mediated reasoning that characterises many qualitative syntheses, and for risking an over-standardised representation of complex claims in fields such as education (Booth et al., 2016; Major & Savin-Baden, 2010).

Importantly, CERQual also reflects a particular disciplinary and applicative orientation: it was developed primarily in health and policy environments, where qualitative syntheses are often expected to inform decisions and recommendations, and where communicating “confidence” serves as decision support (Lewin et al., 2018). In educational research, the emphasis is often different: qualitative systematic reviews frequently pursue additional knowledge functions, such as interpretive explanation, conceptual reframing, or critical/normative argument, so that confidence-like language (e.g., robust themes, strong evidence, consistent findings) often signals rhetorical strength without making explicit the inferential warrants linking synthesised material to the claim. These warrants may differ across synthesis traditions: recurrence or convergence can be treated as sufficient justification; theoretical resonance can be used to infer explanatory power; breadth of contexts can be taken to imply transferability; while attention to dissonant cases can imply boundary conditions that require qualification. Clarifying these implicit inferential steps is therefore crucial for understanding how “confidence” is produced and communicated in educational qualitative synthesis. It is in this space - between the appraisal of contributing studies and the argumentative construction of review-level claims - that the framework proposed in this dissertation positions itself, by making visible both the purposive framing

of claims and the reasoning that connects evidence to conclusions. Recent education scholarship has argued that “evidence quality” should be framed in relation to usability for policymaking and practice, emphasising the fit between evidence, context, and decision needs rather than treating quality as a single universal hierarchy (Ming & Goldenberg, 2021). These challenges are particularly salient in education, where the transfer of systematic review conventions from health research has long been questioned, especially when reviews address complex, context-dependent phenomena and interpretive forms of evidence (Evans & Benefield, 2001).

Gough’s WoE framework (2007) addresses some of these limitations by offering a multidimensional appraisal system. Rather than focusing solely on internal methodological criteria, WoE includes judgments about how well a study contributes to the specific review question and how relevant the topic is to the synthesis aims. Later, Gough (2021) extended this approach by proposing the *Fitness for purpose of an evidence claim framework* (Gough, 2021), which shifts the focus from the quality of individual studies to the evaluation of the *claims made in the review itself*. This reorientation encourages reviewers to consider how well a claim aligns with the review’s intended use and target audience. It also opens space for reflexivity and judgment in evaluating evidence, a crucial consideration for qualitative educational research.

Despite the value of these frameworks, they are rarely used in an integrated or systematic manner within qualitative systematic reviews in education. Reviewers often rely on methodological tools (e.g., CASP) for initial study screening, adopt CERQual for summarising findings, and informally engage with relevance or credibility criteria without a coherent evaluative structure. As a result, the synthesis process may lack theoretical grounding, and the logic of its conclusions may remain under-specified. This is particularly problematic in educational research, where studies frequently involve diverse paradigms, value orientations, and situated interpretations of experience (Booth et al., 2016; Major & Savin-Baden, 2010). The absence of robust conceptual frameworks also limits the capacity of reviews to communicate how and why their claims should be accepted or used.

In this context, Toulmin’s (1958) *Model of argumentation* has been proposed as a promising complement to existing appraisal tools. Toulmin’s model makes the structure of reasoning explicit, identifying claims, data, warrants, and rebuttals, and thus provides a tool for unpacking the argumentative moves made within a synthesis. Toulmin’s model can help reviewers and readers assess the plausibility and internal coherence of evidence claims in

systematic reviews when used alongside frameworks like Gough's *Fitness for Purpose of an Evidence Claim*. Ultimately, conceptual frameworks in qualitative synthesis serve as evaluative tools and as instruments of clarity and accountability. They help reviewers navigate the interpretive complexities of educational research and make visible the pathways through which findings are constructed. However, current practices often rely on fragmented or underdeveloped frameworks, limiting qualitative systematic reviews' transparency, rigour, and utility. This study addresses this methodological gap by developing and applying an integrated framework, combining Gough's framework and Toulmin's model to support a more systematic and theoretically grounded approach to analysing evidence claims in qualitative educational research reviews.

1.5 Gaps in the literature

Several recurring limitations have been identified across the growing body of literature on qualitative systematic reviews. While methods for conducting and appraising such reviews have evolved considerably over the past two decades, challenges remain about review practices' theoretical integration, evaluative coherence, and epistemological transparency, particularly in education. This section outlines key gaps that justify the development of a new integrative framework for reading and understanding evidence claims in qualitative systematic reviews.

A first significant gap concerns the limited use of conceptual frameworks to guide the synthesis process. As discussed in the previous section, many reviews rely on tools such as CASP or CERQual in a fragmented and procedural manner (Booth et al., 2016; Major & Savin-Baden, 2010). While useful in isolation, these tools are often deployed without explicit integration into a broader theoretical or epistemological framework. As a result, the rationale underpinning the synthesis, how concepts are translated across studies, how themes are derived, and how claims are warranted often remains under-specified or implicit. This is especially problematic in education, where studies reflect diverse paradigms, context-specific findings, and value-laden perspectives.

A second gap lies in the treatment of reasoning and argumentation within qualitative reviews. While the inclusion criteria and synthesis procedures are typically described in detail, less attention is paid to the internal logic of how claims are built. Few reviews explicitly identify the warrant for their conclusions or explain how the included studies

support their interpretations. Dixon-Woods et al. (2006) note that many syntheses fail to articulate the analytical pathways from data to conclusion, which can obscure the basis for readers to accept the review's findings. Without a clear articulation of this reasoning, assessing the credibility or fitness-for-purpose of the claims being made becomes difficult.

Third, there is a lack of guidance on evaluating the reviews' claims, as opposed to the quality of the primary studies they include. While tools like GRADE-CERQual have begun to address this issue by introducing domains such as coherence and adequacy, they remain focused on individual review findings rather than the structure and justification of the review as an argumentative whole.

Gough's (2021) proposal to evaluate evidence claims based on their fitness for purpose represents a significant step forward, yet it has not been widely operationalised or combined with other analytical tools. Moreover, very few reviews draw on argumentation theory to clarify the structure of their conclusions. Toulmin's (1958) model offers a robust means of clarifying the relationship between claims, supporting evidence, and underlying warrants. Although scholars such as Newman and Gough (2020) have begun to explore the potential of this model for analysing evidence claims in qualitative synthesis, its use remains absent mainly from actual review practice. This leaves a methodological gap in the capacity of systematic reviews to reflect critically on how they construct and communicate knowledge.

Finally, these challenges are heightened in educational research because the field often deals with value-laden and politically sensitive topics. Studies in education frequently address complex ethical and social issues, such as student agency, equity, identity, and pedagogical reform, all of which involve competing visions of what education should achieve. Synthesising such research requires sensitivity to methodological rigour, interpretive nuance, and the situated nature of knowledge claims. However, current guidance for qualitative systematic reviews rarely addresses these dimensions, leaving reviewers without the tools to engage meaningfully with educational inquiry's conceptual and ethical demands (Major & Savin-Baden, 2010). These gaps suggest the need for a more comprehensive approach to analysing and evaluating qualitative evidence claims, integrating insights from conceptual appraisal, argumentation theory, and field-specific epistemologies. This study responds to that need by developing and applying an integrated framework based on Gough's (2021) *Fitness for Purpose of an Evidence Claim* model and Toulmin's (1958) theory of argumentation. It aims to support more transparent, reflexive, and conceptually grounded synthesis practices in qualitative educational research.

Concluding remarks

The literature reviewed in this chapter provides a foundation for the present study's central contribution: developing an integrated conceptual framework that supports the analysis of evidence claims in qualitative systematic reviews. Combining Gough's (2021) *Fitness for purpose of an evidence claim framework* with Toulmin's (1958) *Model of argumentation*, the study aims to fill an important methodological and theoretical gap. This integrated approach is designed to assist researchers and research users in making sense of the reasoning behind qualitative syntheses, evaluating the credibility and appropriateness of their claims, and ultimately enhancing the transparency and utility of systematic reviews in education.

2 Theoretical framework

Chapter 1 concluded by raising a practical concern: many qualitative reviews present their findings without clearly demonstrating how claims are constructed or whether they are appropriate for their intended audiences and purposes. Chapter 2 addresses this issue by presenting a conceptual framework for examining review claims, including what is being asserted, the grounds on which it is asserted, the warrant used, the qualifiers applied, and the rebuttals considered. The framework integrates two complementary approaches: Toulmin's model (S. 1922-2009. Toulmin, 1958), which unpacks the internal logic of a claim; and Gough's Fitness for Purpose (Section I) (D. Gough, 2021), which judges whether the type and calibration of the claim are appropriate for its stated use and users. Section 2.1 introduces this integrated framework, which was developed using a reflexive, iterative approach that treats conceptualisation as an interpretive map mediating between epistemology and methodology. The framework clarifies assumptions, supports analytical coherence and facilitates the production of credible, context-sensitive knowledge. Section 2.2 provides a reconstruction of Toulmin's model for applied educational enquiry. This model is used to analyse the internal structure of reasoning in reviews, examining how claims are supported by data (grounds), connected by reasoning (warrants), backed up by evidence (backing), qualified by modality and limited by rebuttals. This approach allows for a detailed evaluation of argumentative coherence in qualitative syntheses, where conclusions emerge through configuration rather than statistical inference.

Section 2.3 introduces Gough's framework, which shifts the focus to the external positioning of claims, including their assumptions and values, intended audiences and uses, scope, level, generalisability/transferability and engagement with alternative explanations. The central question here is not only whether an argument is well formed, but also whether its claim is fit for purpose within its context of application.

Section 2.4 justifies the integration of these models. Toulmin provides micro-level tools for tracing inferential pathways, while Gough adds macro-level criteria for evaluating relevance, appropriateness and communicative intent. Together, they provide a coherent and mutually reinforcing basis for analysing the logic and purpose of evidence claims.

The chapter concludes by laying the groundwork for Chapter 3, in which this integrated framework is operationalised into coding categories and applied to a purposive sample of qualitative systematic reviews.

2.1 An integrated conceptual framework

Conceptual frameworks play a central role in shaping the logic of inquiry in qualitative research, particularly in the synthesis of qualitative evidence. They provide the structural and interpretive scaffolding that guides data selection, interpretation, and synthesis. Rather than a mere bundle of theories, a conceptual framework weaves concepts, assumptions, expectations, beliefs, and theories into a coherent scaffold that shapes and informs the research. It functions as an interpretive map that mediates between epistemology and method, enabling the researcher to make defensible and coherent analytical choices (Ravitch & Riggan, 2017) .

This study adopts a reflexive and integrative approach to framework development, aligned with qualitative traditions emphasising transparency, coherence, and responsiveness to the nature of the research questions and the empirical materials; such reflexivity also entails attention to positionality, that is, how the researcher's standpoint and assumptions shape interpretive choices throughout the research process (Maxwell, 2013); (Anfara & Mertz, 2015); (Bourke, 2014). Rather than imposing a pre-formed model, the framework was constructed iteratively, emerging through sustained engagement with theoretical literature, review data, and the epistemological assumptions underpinning qualitative synthesis. This approach reflects what Noblit and Hare (1988) describe as the translation of meaning across conceptual systems, an essential qualitative synthesis task where evidence must be actively constructed, warranted, and positioned.

Conceptual frameworks in this tradition serve at least three interrelated purposes:

1. Clarification of Assumptions: making explicit the philosophical and theoretical commitments that shape the approach to evidence, synthesis, and interpretation (Lincoln & Guba, 1985).
2. Analytical Coherence: providing a consistent conceptual structure that enables systematic yet flexible interpretation of qualitative data (Popay et al., 2006).
3. Epistemic Mediation: bridging theory and method to support the development of situated, reflexive, and credible knowledge claims (Ravitch & Riggan, 2021).

Within this study, the conceptual framework integrates Stephen Toulmin's (1958) *Model of argumentation* and David Gough's (2021) *Fitness for purpose of an evidence claim framework*. This integration offers a dual analytical lens that is logically rigorous and

contextually sensitive. Toulmin's model focuses on arguments' internal structure and reasoning patterns, mapping how data, warrants, backing, qualifiers, and rebuttals support claims. Gough's framework, particularly Section I, addresses evidence claims' broader evaluative and communicative dimensions, including their assumptions, values, intended audiences, and purposes.

This dual-theoretical foundation is not merely additive but integrative. While Toulmin enables a micro-level analysis of how claims are logically constructed, Gough provides a macro-level framework for interpreting whether those claims are appropriate, relevant, and transparent in context. This alignment is reinforced by Gough's reference to Toulmin, describing an evidence claim as "based on a warrant linking data to the claim plus any qualifications or exceptions to the claim" (Gough, 2021, p. 7). Their conceptual compatibility allows for a rich and layered reading of how qualitative evidence is transformed into claims that carry weight within educational and policy discourse.

This integrated conceptual framework also responds to calls in the methodological literature for greater reflexivity in qualitative synthesis. Authors such as Sandelowski and Barroso (2007) and Booth, Sutton, and Papaioannou (2016) highlight the need for explicit frameworks that account not only for methodological rigour but also interpretive transparency and evaluative reasoning. In this study, the integrated conceptual framework illustrated in this chapter supports the development of an analytic strategy that interrogates what is claimed, how it is justified, and for whom it is intended.

By anchoring the study in this integrated conceptual foundation, this thesis contributes to broader methodological conversations about constructing and assessing evidence claims in qualitative reviews. It offers a transferable approach for analysing claims in terms of their content, argumentative form, epistemic basis, and rhetorical function.

The remainder of this chapter is structured as follows: Section 2.2 introduces Toulmin's *Model of argumentation* and explains why it is relevant for analysing the internal structure and reasoning patterns of evidence claims in qualitative systematic reviews. Section 2.3 then introduces Gough's *Fitness for Purpose of an Evidence Claim framework*, focusing on Section I, which addresses the evaluative and contextual dimensions of claim formulation. Section 2.4 explains the rationale for integrating these two models, emphasising their conceptual compatibility and methodological alignment with the objectives of this study. Together, these sections establish the theoretical basis for the analytical framework

presented in Chapter 4, in which the combined model is converted into coding categories for empirical analysis.

2.2 Toulmin's *Model of argumentation*

To assess the structure and coherence of evidence claims in systematic reviews of qualitative research, this study draws on Stephen Toulmin's *Model of argumentation*, first introduced in his seminal work 'The Uses of Argument' (1958). Although widely cited across disciplines, including law, philosophy, science education, and communication studies, the model is not always well known in educational research or the context of systematic review methodology. For this reason, a detailed explanation of its components and underlying assumptions is warranted before considering its specific relevance for the present study. This section begins by reconstructing Toulmin's model as originally conceived, emphasising its distinction from formal logic and its utility for analysing real-world reasoning. It then considers how the model has been adapted in applied research, particularly in education, and concludes by justifying its selection as an analytic tool for examining how evidence claims are constructed and supported in qualitative syntheses.

Toulmin contends that arguments in everyday, professional, and scientific practice seldom conform to the rigid form of syllogistic deduction. They are advanced under conditions of partial knowledge and contextual constraint, and their appraisal depends on field-specific standards - accepted methods, admissible evidence, and relevant audiences - rather than a single, universal logic (Toulmin, 1958). Accordingly, he broadens "validity" from a purely deductive criterion to a practical one: an argument is strong to the extent that (a) its grounds are acceptable within the field, (b) its warrant—the inferential rule linking grounds to claim—is recognised by that field, (c) its warrant is supported by credible backing, and (d) its qualifiers and rebuttals proportionately register uncertainty and limits. Toulmin aimed to develop a model that could account for how people argue and justify their claims in everyday and professional settings.

Toulmin's *The Uses of Argument* remains a foundational text in practical reasoning, particularly in disciplines where knowledge is context-sensitive, value-laden, and interpretive. Unlike the formal logic of syllogisms or the probabilistic logic of statistics, Toulmin's approach is grounded in the real-world construction of arguments: how people justify claims in domains such as law, ethics, education, and public policy. His model offers a pragmatic schema for analysing the structure of an argument and the relationship between its elements, enabling the evaluation of reasoning in naturalistic contexts.

At the heart of Toulmin's model is the idea that an argument comprises six key components (see Figure 2.1, adapted and redrawn from Toulmin, 1958):

1. Claim: the conclusion or assertion that the speaker/writer aims to establish.
2. Data (Grounds): the evidence or facts used to support the claim.
3. Warrant: the reasoning or principle that connects the data to the claim.
4. Backing: additional support for the warrant, especially if it is contested.
5. Qualifier: an indication of the strength or generalizability of the claim (e.g., "probably," "possibly").
6. Rebuttal: an acknowledgement of exceptions, limitations, or counterarguments.

This field-dependence is particularly relevant for qualitative systematic reviews in education, where arguments are constructed through interpretive synthesis rather than through formal deduction, by combining diverse data, theory, and context to produce meaningful and useful claims within a particular epistemological and social landscape. Toulmin's model is therefore well suited to analysing the internal structure of evidence claims in qualitative syntheses: it enables a close examination of how reviewers construct, justify, qualify, and delimit their claims. At the same time, applying Toulmin to systematic reviews requires attention to how argumentation is expressed in synthesis writing, where reasoning is often distributed across sections and key inferential links remain implicit rather than being stated as explicit premises. For this reason, this study also draws on (Khambete, 2019) adaptation of Toulmin's model, which operationalises Toulmin's elements for the analysis of evaluative and evidence-based claims in extended texts. This adaptation is particularly useful for review discourse because it treats warrants as the often tacit rules by which reviewers move from synthesised patterns to conclusions - for example, treating convergence as a marker of robustness, theoretical alignment as a basis for explanatory power, or breadth of contexts as a signal of transferability - and it supports the identification of qualifiers and rebuttals when these are embedded in hedging, stated limitations, and boundary conditions rather than presented as formal counter-arguments.

In the present study, Toulmin's model is used as the first layer of analysis to dissect the argumentation strategies embedded in systematic reviews. The model provides the terminology and conceptual tools required to identify how claims are supported by evidence, how strongly they are asserted and whether counter-evidence is considered. This is

particularly relevant in qualitative syntheses, where conclusions are formed through iterative interpretive processes, such as line-by-line coding of findings, constructing descriptive themes and generating higher-order analytical themes that transcend the primary reports, rather than through linear data aggregation (Thomas & Harden, 2008).

Furthermore, Toulmin's model has a precedent in educational research, most notably in analyses of classroom discourse and scientific reasoning that use the Toulmin Argument Pattern (TAP) (Erduran et al., 2004). demonstrating its utility where argumentation is a cognitive and communicative act shaped by disciplinary norms, audiences, and rhetorical goals. TAP operationalises Toulmin's components and shows, empirically, that while claims and data are common, warrants and backings appear less frequently and rebuttals are comparatively rare; targeted instruction increases the sophistication of these elements over time.

By attending to each of Toulmin's components, particularly the often-implicit elements such as warrant, backing, and qualifier, the analysis reveals whether a claim is made and how it is made convincing (or not). This model thus enables the unpacking of argumentative coherence, epistemic rigour, and rhetorical positioning, all of which are central to understanding the credibility and communicative function of evidence in education.

In the next section, this internal lens is complemented by Gough's framework, which allows for an external appraisal that assesses the claim's fitness for purpose in relation to values, audiences, and epistemological positioning.

2.3 Gough's *Fitness for purpose of an evidence claim framework*

To complement the internal analysis of argument structure enabled by Toulmin's model, this study also draws on David Gough's work on evidence appraisal, specifically his *Fitness for purpose of an evidence claim framework* (Gough, 2021). This framework provides a means to examine not only *how* evidence claims are structured but also *why* they are constructed in particular ways, *for whom*, and *toward what end*. Where Toulmin focuses on the anatomy of an argument, Gough brings attention to its epistemological positioning, normative orientation, and communicative intent, dimensions that are especially crucial in systematic reviews of qualitative research, where claims are inherently interpretive, audience-sensitive, and often embedded in policy or practice debates.

Gough's 2021 framework represents a conceptual evolution of his earlier influential work on the Weight of Evidence (WoE) approach (D. Gough, 2007). The WoE framework was developed to support evidence appraisal in systematic reviews, especially in the social sciences. It introduced a multi-criteria model for evaluating studies according to:

- WoE A: the methodological trustworthiness of the study,
- WoE B: the appropriateness of the study design for the review question,
- WoE C: the relevance of the study's focus to the review's purpose,
- WoE D: an overall assessment that integrates the above dimensions.

This model was a significant step in moving evidence appraisal beyond rigid hierarchies and toward a more contextual understanding of quality and relevance. It helped reviewers articulate why certain studies were more suitable for inclusion or synthesis in relation to a review's aims. However, as the field of evidence synthesis matured, particularly for qualitative and mixed-methods research, a need emerged to appraise not just primary studies, but also the evidence claims made by reviewers themselves. That is, how reviewers transform data into knowledge statements through interpretive synthesis, and how those statements are framed and communicated to audiences.

In response, Gough developed the *Fitness for purpose of an evidence claim framework*, which shifts focus from evaluating studies to evaluating the claims that reviewers make on the basis of those studies. This framework centres on a foundational principle: an evidence claim should be judged by how well it fulfils its intended purpose, within a given context, for a particular audience (Gough, 2021).

Section I of the framework (see Figure 2.2, adapted and redrawn from Gough, 2021), the focus of this study, offers tools for examining two core dimensions of a claim:

- The perspectives underlying the claim, including theoretical assumptions, normative values, and stakeholder priorities.
- The nature of the claim, including its scope, level, generalizability, certainty, and engagement with alternative explanations.

These dimensions are particularly salient in qualitative synthesis, where reviewers often must navigate multiple paradigms, ambiguous data, and user-sensitive goals (Sandelowski & Barroso, 2007). By applying Gough's lens, this study can evaluate not only what a

systematic review concludes but also how well the claim aligns with its intended audience and purpose, an essential consideration in education research, where findings are expected to inform diverse actors (teachers, policymakers, researchers) with varying informational needs.

One of the most compelling reasons for selecting Gough's framework is its theoretical continuity with Toulmin's *Model of argumentation*. Gough (2021) explicitly adopts Toulmin's terminology to conceptualise the structure of evidence claims, stating that "an evidence claim is based on a warrant linking data to the claim plus any qualifications or exceptions to the claim" (p. 7). This formulation mirrors the core components of Toulmin's model, data, warrant, qualifier, and claim, and reflects a shared commitment to analysing how claims are justified through structured reasoning. However, Gough extends Toulmin's model by situating the claim within a broader evaluative context. He emphasises not only the argument's internal logic but also the assumptions, values, intended uses, and audiences that shape its formulation. In this sense, Gough builds on Toulmin's functional argumentation analysis, adding a communicative and purposive dimension that is especially salient in applied educational and policy research.

However, while Gough builds on Toulmin's functional analysis of argument, his framework is directed at a broader evaluative task: assessing the fitness of an evidence claim in light of its intended purpose, audience, assumptions, and values. He shifts the focus from the internal mechanics of reasoning to the communicative, normative, and contextual dimensions of knowledge claims, primarily as they function in applied fields such as education, social policy, and health. In doing so, Gough offers a powerful tool for appraising whether a claim is appropriate, relevant, and meaningful within a given social or decision-making context.

Despite its strengths, Gough's framework is not sufficient for the aims of this study. Its emphasis lies on the interpretive framing and epistemic appropriateness of evidence claims, their aim, for whom, and with what underlying assumptions. However, it does not provide a systematic or granular method for examining how those claims are constructed regarding their internal argumentative structure. Gough's framework invites evaluative reflection on a claim's external coherence and contextual fitness. However, it does not offer detailed tools for tracing the logic, inferential pathways, or rhetorical architecture by which claims are justified.

This analytical gap is addressed by Toulmin's model, which excels in dissecting the micro-level structure of argumentation. Toulmin enables a close reading of how claims are built

from specific forms of data, linked through warrants, supported by backing, and qualified or challenged through rebuttals. His model reveals whether claims are logically coherent, adequately supported, and transparently qualified, crucial when analysing qualitative synthesis, where arguments are often implicit, layered, and interpretive.

This is precisely why the two models are employed in tandem. Gough's framework contributes a macro-level evaluative perspective, focused on the communicative intent and appropriateness of the claim about its context. In contrast, Toulmin contributes a micro-level analytical framework to examine the structural integrity of the reasoning process itself. Each addresses a dimension that the other does not fully encompass. Integrating both models ensures this study can engage with evidence claims as epistemic arguments and contextualised, purposive knowledge statements.

Thus, rather than choosing between the two frameworks, this study adopts an integrated conceptual approach, in which Toulmin and Gough are compatible and mutually reinforcing. The result is a robust analytical lens capable of assessing the internal logic and the external legitimacy of claims made in qualitative systematic reviews, an approach uniquely suited to the complexity and nuance of evidence construction in educational research.

In the context of this study, only Section I of the framework is used. This decision is both conceptual and pragmatic. Conceptually, Section I offers the most direct interface with Toulmin's model, focusing on the construction and justification of claims. Pragmatically, it allows for a deep, sustained analysis across multiple reviews while maintaining analytic clarity. While Gough's full framework includes additional sections on evidence type, methodological integrity, and use context, these are not within the scope of the present analysis, which concentrates on evaluating how reviewers construct claims within their studies.

Applying Gough's framework in this way offers an interpretive, values-sensitive appraisal of claims. It surfaces key questions:

- Whose priorities does the review reflect?
- How explicitly are assumptions and values articulated?
- Is the claim formulated to meet specific informational needs?
- Does it acknowledge its epistemological stance?

These questions are operationalised in Chapter 4 by developing a coding framework, including analytical categories such as *Topic*, *Importance*, *Justification of Method*, *Theoretical Framework*, and *Framework Justification*. Each of these categories maps onto the evaluative concerns raised by Gough and enables the systematic identification of how qualitative reviews justify their findings and position them within a broader educational and social discourse.

In sum, Gough's framework introduces a contextual and reflexive dimension to the analysis of evidence claims, complementing the structural lens provided by Toulmin. Where Toulmin dissects the *how* of argumentation, Gough interrogates the *why* and *for whom*. This duality, logic and purpose are essential for a robust understanding and assessment of evidence claims in qualitative systematic reviews, which must be internally reasoned and externally meaningful.

2.4 Justification for integrating Toulmin and Gough

The integration of Toulmin's *Model of argumentation* and Gough's *Fitness for purpose of an evidence claim framework* stems from both a conceptual rationale and a methodological necessity. Together, they offer a multi-dimensional lens for analysing evidence claims in systematic reviews of qualitative studies, one that addresses the logical construction of arguments and the contextual appropriateness of their formulation.

As outlined in the preceding sections, Toulmin's (1958) model provides a detailed schema for examining the internal structure of reasoning, focusing on the relationships between claims, data, warrants, backing, qualifiers, and rebuttals. It allows for mapping how reviewers justify their claims through evidence and inference. However, while it is powerful in capturing argumentative coherence, Toulmin's framework is limited in its ability to account for the broader contextual, epistemological, and purposive factors that shape the function and framing of those claims.

This is where Gough's framework becomes essential. By focusing on the fitness for purpose of evidence claims, particularly in terms of their values, audiences, theoretical perspectives, and intended use, Gough (2021) expands the evaluative focus beyond formal argumentation. He highlights how claims are embedded in situated knowledge production and communication practices, especially within socially and politically charged domains such as education. This orientation resonates with the interpretivist stance of qualitative synthesis,

where meaning and relevance are co-constructed through engagement with multiple perspectives and forms of evidence (Lincoln & Guba, 1985; Noblit & Hare, 1988).

The combination of Toulmin and Gough provides a methodological response to the challenges inherent in synthesising and evaluating claims from qualitative systematic reviews. These reviews often rely on narrative or configurative synthesis strategies, which require transparency in interpretive logic and justificatory reasoning (Popay et al., 2006; Thomas & Harden, 2008). Toulmin supports the identification of such reasoning at the micro-level of argument structure, while Gough encourages a macro-level reflection on the epistemic and practical purposes of the claim. Together, they allow for a layered analysis of both form and function.

From a research design perspective, this integration reflects what Ravitch and Riggan (2017) and Maxwell (Maxwell, 2013) describe as a reflexive and iterative construction of a conceptual framework, which evolves from the interplay between theoretical commitments, methodological concerns, and the empirical realities of the study. Rather than treating theory as static or imposed, this study approaches its conceptual framework as a dynamic interpretive tool that mediates between abstract reasoning and applied analytic practice (Anfara & Mertz, 2015).

Furthermore, this integration supports what Andrews (2005; 2010) calls a “critical-constructive approach” to argumentation in education, one that is attentive not only to how claims are logically constructed but also to how they function rhetorically, socially, and politically. In this view, the production of knowledge is never neutral; it always involves choices about what to include, what to omit, how to frame findings, and how to position the reader in relation to those findings. These choices are not merely stylistic or technical; they are political, in that they shape how particular realities are represented, who is authorised to speak, and whose experiences or perspectives are foregrounded or silenced. This orientation is consistent with traditions that frame inquiry as praxis, emphasising that research is simultaneously epistemic and political, and that methodological choices shape what kinds of realities become thinkable and actionable (Lather, 1986).

In the context of educational systematic reviews, this issue becomes particularly acute. Reviews are often positioned as authoritative syntheses of available evidence, designed to inform practice, influence policy, or guide professional judgment. Yet every review is also a product of epistemological assumptions, interpretive decisions, and representational strategies. The selection of studies, the extraction of findings, the language used to describe

phenomena, and the ways in which themes are synthesised all contribute to constructing a version of reality, one that may reflect certain priorities, values, or institutional agendas while marginalising others.

This is what is meant by the politics of representation: the recognition that research, especially synthetic and influential forms like systematic reviews, participates in shaping educational discourse by producing “truth effects” (Foucault, 1980) that can legitimate specific ways of seeing, thinking, and acting. Claims within reviews do not merely report findings; they often carry implicit normative messages about what counts as relevant knowledge, which voices matter, and what actions should follow. As such, the tools used to analyse these claims must be capable of interrogating both their structural form and their discursive positioning.

This study’s integrated framework responds directly to that challenge. Toulmin’s model enables the detailed mapping of how claims are justified, exposing the logic and structure of reasoning. Gough’s framework adds a layer of analysis that questions why those claims are framed in particular ways, for whom they are intended, and to what ends they are directed. It is in the combination of these models that the analytical tools become capable of capturing both the logic of synthesis and the politics of representation, revealing not just whether a claim is valid or plausible, but also how it functions in educational discourse as a means of persuasion, legitimation, or silencing.

This dual sensitivity is crucial in a research landscape where evidence is increasingly mobilised for policy decision-making, accountability, and reform. The framework developed in this study is therefore not only a tool for reading and evaluating methodological or argumentative rigour, but also a way of making visible the power-laden processes through which educational knowledge is constructed, validated, and circulated.

In short, in this study, the integration of Toulmin and Gough is justified by:

- Their complementary analytic focus (argumentative structure vs. evaluative purpose);
- Their conceptual resonance, with Gough building explicitly on Toulmin;
- Their epistemological fit with qualitative research, especially in interpretive syntheses;
- Their methodological utility in supporting a systematic, theory-informed, and context-aware analysis of evidence claims.

This conceptual integration laid the foundation for the coding framework described in Chapter 3, where it was operationalised into a set of analytical categories used to examine a purposive sample of systematic reviews.

Concluding remarks

This chapter sets out the theoretical foundation for analysing evidence claims in qualitative systematic reviews. It clarifies the distinct role of a conceptual framework in relation to background theory and analytical procedures, and introduces two complementary lenses: Toulmin's model, which reconstructs the internal logic of claims in terms of grounds, warrants, backing, qualifiers and rebuttals; and Gough's Fitness for Purpose, which appraises whether claims are appropriate to stated aims, audiences and values (Section I). When considered in isolation, Toulmin sheds light on the structure of arguments but not their discursive positioning, while Gough highlights purpose and audience but lacks granular tools for inference. Together, they provide a coherent lens that captures both the anatomy of reasoning and the contextual fit of claims. This dual perspective recognises that qualitative syntheses do not merely report evidence, but also construct, calibrate and position it. Chapter 4 translates this integrated framework into an operational coding scheme and applies it to a purposive sample of systematic reviews.

FIGURE 2.1. TOULMIN'S *MODEL OF ARGUMENTATION* (ADAPTED AND REDRAWN FROM ORIGINAL SOURCES).

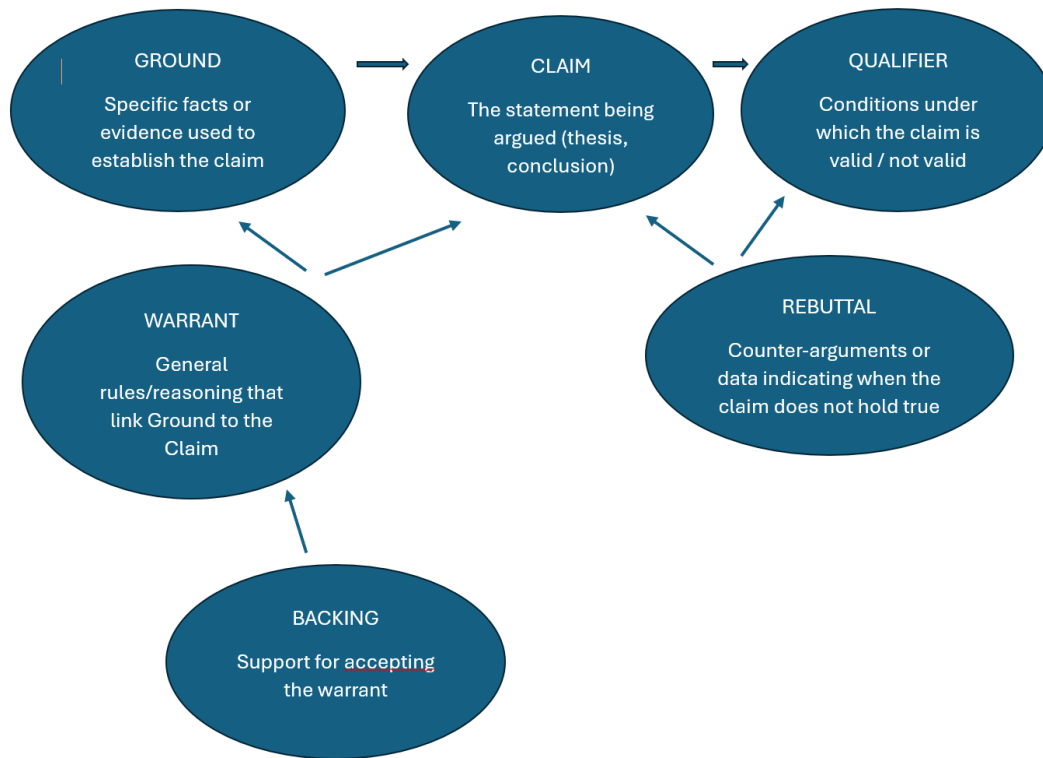
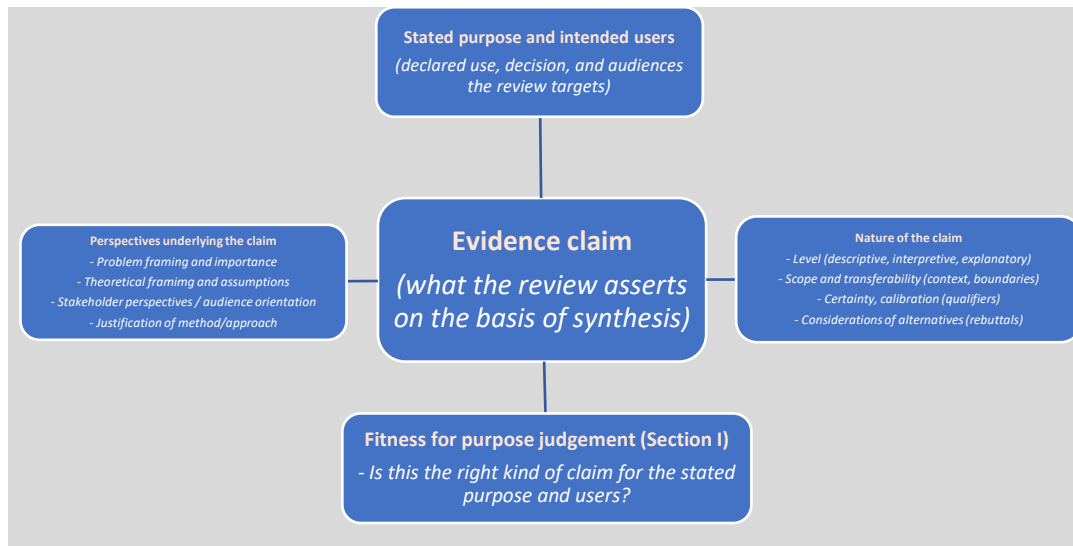


FIGURE 2.2. GOUGH'S 'FITNESS FOR PURPOSE OF AN EVIDENCE CLAIM FRAMEWORK' (ADAPTED AND REDRAWN FROM ORIGINAL SOURCES).



3 Methodology

Chapter 2 developed the integrated conceptual framework that combines Toulmin's (1958) model of argumentation with Gough's (2021) fitness-for-purpose approach. Building on this foundation, Chapter 3 sets out the study's methodology and explains how the framework was operationalised into a practical analytical strategy. Section 3.1 translates the framework into a structured set of coding categories reflecting Toulmin's argumentative elements (claim, grounds, warrant, backing, qualifier, rebuttal) and Gough's interpretive dimensions (relevance, clarity, purpose). Section 3.2 details the review protocol used to identify and select a purposive sample of six systematic reviews. Section 3.3 describes the development and iterative refinement of the codebook, informed by established literature in qualitative synthesis and content analysis, to capture both explicit and implicit dimensions of argumentation. Chapter 4 then applies this methodology to the sample, analysing how evidence claims are constructed, substantiated, and communicated in qualitative research syntheses in education.

3.1 Operationalising the integrated conceptual framework

This section describes how the combined use of Toulmin's *Model of argumentation* and Gough's *Fitness for purpose of an evidence claim framework* was translated into a practical analytical tool. This operationalisation enabled the conceptual framework to serve not only as a theoretical foundation, but also as a systematic lens through which to examine how evidence claims are constructed and communicated within systematic reviews of qualitative studies in education. The two models were used for different but complementary tasks. Toulmin's framework helped break each review's argument into its core parts and check how well those parts fit together. Gough's framework—especially Section I—kept the focus on who the claim is for, why it matters, and how its scope and limits are stated. Together, they guided the coding framework developed for this study, ensuring the analysis remained precise in its reasoning while remaining sensitive to context and purpose. From a methodological standpoint, this approach reflects what Ravitch and Riggan (Ravitch & Riggan, 2017) describe as the reflexive and iterative construction of a conceptual framework, one that is developed through ongoing engagement with theory, empirical material, and the specific demands of the research context. Rather than adopting a single pre-established model, the framework evolved through a process of adaptation, grounded in the nature of the research questions and the epistemological assumptions of qualitative synthesis. As Ravitch and Riggan emphasise, a conceptual framework must be logically aligned with the

research's aims and designed to mediate effectively between theory and method. In this study, Toulmin's model was employed to map the internal logic of how claims are constructed in systematic reviews, with a focus on the model's reasoning components, including data, warrants, backing, qualifiers, and rebuttals. This enabled a fine-grained analysis of the argumentative strategies employed by review authors. In parallel, Section I of Gough's framework was employed to assess the contextual appropriateness or "fitness for purpose" of these claims. This involved examining the values and assumptions embedded in the claims, the priorities they reflect, and the audiences they appear to address.

The decision to focus exclusively on Section I of Gough's broader framework was both conceptual and pragmatic. While Gough's full framework comprises four interrelated sections, Section I offers the most direct and coherent interface with Toulmin's model, particularly in its focus on how evidence is used to justify claims. As Gough (2021) notes, drawing explicitly on Toulmin's terminology, "an evidence claim is based on a warrant linking data to the claim plus any qualifications or exceptions to the claim" (Gough, 2021, p. 7). This overlap provided a natural point of integration between the two models. Through this combined lens, Toulmin's model contributed a detailed view of argumentative structure, while Gough's framework extended the analysis to include the intentional, audience-sensitive, and value-laden dimensions of evidence use. This allowed claims to be examined not only for how they are reasoned, but also for why they are framed in particular ways and how they are positioned within educational and policy discourses. The integration of these frameworks thus provided a comprehensive conceptual and methodological foundation for this study. The integrated framework (see Figure 3.1) lists the categories from both models—Gough's and Toulmin's—which guided the development of analytical coding questions and informed the interpretation of the findings presented in Chapter 4.

Another rationale concerns the order of application. Although Toulmin's model predates Gough's, the analysis began with Gough's categories for two reasons. Firstly, prioritising Gough's categories ensured contextual relevance and clarity of purpose: by defining the audience, purpose, values, scope and conditions of applicability of the claim at the outset, the analysis established the criteria by which argumentative adequacy would be judged later on. This approach reduced the risk of decontextualised reconstruction and aligned the evaluation with the intended use of the claim. Secondly, starting with Gough provided practical scaffolding for coding and sampling. Topic specification, importance, and method justification served as gatekeeping criteria that clarified inclusion, sharpened the unit of

analysis, and oriented subsequent, finer-grained Toulmin coding (claim, grounds, warrant, backing, qualifier, and rebuttal). In short, starting with Gough preserved contextual integrity and evaluative transparency, while proceeding with Toulmin tested how the claims were actually built and justified, given those declared purposes and audiences. In the next sections, this process is described in detail, including the development of the codebook and its application to the selected corpus of systematic reviews.

FIGURE 3.1 - THE INTEGRATED CONCEPTUAL FRAMEWORK

Category	Definition	Coding Question	Source
Topic	Focus and scope of the systematic review.	What is the topic of the SR? Is it clearly defined?	Fitness for purpose of an evidence claim framework (Gough)
Importance	Significance or relevance of the topic.	Do the authors mention the significance of the subject?	Fitness for purpose of an evidence claim framework (Gough)
Justification of Method	Rationale for using a systematic review method.	Do the authors justify the use of a systematic review?	Fitness for purpose of an evidence claim framework (Gough)
Theoretical Framework	Use and influence of a guiding framework.	Which framework do the authors say they use, and why?	Fitness for purpose of an evidence claim framework (Gough)
Framework Justification	Rationale behind the choice of theoretical framework.	What is the authors' justification for the use of the framework in question?	Fitness for purpose of an evidence claim framework (Gough)
Main Claim	Central conclusion based on synthesis.	What is the central conclusion or assertion made by the SR?	Fitness for purpose of an evidence claim framework (Gough) – Model of Argumentation (Toulmin)
Other Claims	Subordinate or related conclusions.	Are there other claims made by the authors?	Authorial integration
Claim Framing	How the claim is presented (e.g., theoretical, practical).	How do the authors frame the claim?	Fitness for purpose of an evidence claim framework (Gough) – Model of Argumentation (Toulmin)

Grounds	Data supporting the claim.	What kinds of data are used to support the claim?	Model of Argumentation (Toulmin)
Synthesis Type	Nature of synthesis (interpretative, configurative).	What kind of synthesis is used?	Authorial integration
Validity/Relevance	Evaluation of quality or appropriateness of evidence.	Do the authors address the validity and relevance of the evidence?	Authorial integration
Warrant	Logical link between data and claim.	What connects the evidence to the claim?	Model of Argumentation (Toulmin)
Backing	Theoretical/empirical support for the warrant.	What foundations justify the warrant?	Model of Argumentation (Toulmin)
Qualifier	Degree of certainty associated with the claim.	How certain or qualified is the claim?	Model of Argumentation (Toulmin)
Rebuttal	Limitations or counter-arguments acknowledged.	Do authors identify limitations or opposing interpretations?	Model of Argumentation (Toulmin)

3.2 Categories derived from Gough's framework

To operationalise Section I of Gough's *Fitness for purpose of an evidence claim framework* (2021), a set of coding categories was developed to reflect the interpretive dimensions outlined in the integrated framework (Figure 3.1). Section I addresses two interrelated aspects:

- (i) the perspectives underlying the claim, including values, priorities, and theoretical constructs; and
- (ii) the nature of the claim, which encompasses its scope, level, focus, generalisability, degree of certainty, and consideration of alternative explanations.

These dimensions were translated into five analytical categories designed to systematically examine how each review constructs its evidence claim and positions it in relation to both users' needs and its underlying epistemological orientation.

- **Topic** captures the *focus and scope* of the systematic review. This aligns directly with Gough's emphasis on the nature and focus of the claim, helping to identify what is being claimed and whether it is articulated with clarity and specificity. By coding the Topic, the analysis identifies whether the review clearly delineates the object of inquiry and its analytical boundaries, a prerequisite for assessing the clarity and precision of the claim itself. As Gough (2021) emphasises, claims must be assessed relative to their intended purpose and audience, this requires clarity about *what* the review investigates. A clearly articulated topic enables meaningful appraisal of the relevance, scope, and fit-for-purpose nature of an evidence claim. Topic specification also aligns with Toulmin's requirement that claims be assessable based on their context and field-dependent standards.
- **Importance** addresses the relevance and significance attributed to the topic by the authors. This category assesses whether the review explains *why* its topic matters and interrogates whether the review explicitly conveys the significance or urgency of its focus, thus addressing the value-laden priorities that Gough identifies as essential to understanding the relevance of a claim to its intended users. This reflects Gough's concern with fitness for purpose, particularly whether the claim responds to relevant *priorities* and stakeholder needs.
- **Justification of Method** refers to the rationale provided for adopting a systematic review as the methodological approach. This element reflects the values and assumptions underlying how evidence is assembled and justified, an issue central to the perspectives that

underpin the claim in Gough's framework. The rationale for employing a systematic synthesis is crucial to evaluating the legitimacy of the review's conclusions. Gough underscores the importance of methodological appropriateness in relation to purpose; an explicit justification reinforces coherence between the research aim and the chosen synthesis strategy, thereby enhancing the credibility of both the evidence and its associated claims. Furthermore, such justification engages with underlying assumptions about the validity of accumulation and comparison within qualitative inquiry.

- **Theoretical Framework** investigates the presence and role of conceptual or theoretical structures used to guide the review. This connects to Gough's call to interrogate the *theoretical constructs* shaping an evidence claim and their implications for the framing and interpretation of findings. Rationale: The use of an explicit theoretical or conceptual framework is a key indicator of how claims are shaped, interpreted, and delimited. Toulmin (1958) and Gough (2021) both stress that evidence claims are constructed through perspectives and assumptions, which include theoretical models and disciplinary paradigms. This category captures the presence and function of such frameworks in structuring reasoning and interpreting findings.

- **Framework Justification** refers to the rationale provided for adopting a systematic review as the methodological approach. This element reflects the values and assumptions underpinning how evidence is assembled and justified, an issue central to the *perspectives underlying the claim* in Gough's framework. The rationale for employing a systematic synthesis is crucial to evaluating the legitimacy of the review's conclusions. Gough underscores the importance of methodological appropriateness in relation to purpose; an explicit justification reinforces coherence between the research aim and the chosen synthesis strategy, thereby enhancing the credibility of both the evidence and its associated claims. Furthermore, such justification engages with underlying assumptions about the validity of accumulation and comparison within qualitative inquiry.

Taken together, these five categories enable a structured yet interpretive analytical approach that responds directly to Gough's call for appraising the "fitness for purpose" of evidence claims not only in terms of their technical adequacy but also with regard to their relevance, transparency, and coherence within a given social and epistemological context. Rather than measuring the validity of a claim solely by internal methodological rigour, this section of the framework attends to the communicative and justificatory dimensions of the review, that is,

how the claim is constructed, why it is framed in a particular way, and for whom it is intended.

This part of the framework does not seek to determine whether a review is right or wrong in a narrow methodological sense. Instead, it enables a qualitative appraisal of how reviews construct their claims in ways that may either align with or fall short of the standards that Gough outlines, particularly in relation to how clearly the claim is framed, how relevant it is to educational stakeholders, and how coherently its underlying assumptions and values are articulated.

3.3 Categories derived from Toulmin's model

In *The Uses of Argument* (1958), Toulmin proposes a model for analysing the structure of reasoning in practical, context-sensitive fields, distinguishing it from formal deductive logic. He argues that “the concept of validity must be broadened and adapted to suit the variety of arguments which arise in different fields” (Toulmin, 1958, p. 7). Central to his contribution is the notion that arguments have a functional structure, composed of six interrelated elements: Claim, Grounds (Data), Warrant, Backing, Qualifier, and Rebuttal. This model is particularly useful for analysing how claims are justified within the interpretive and value-laden domain of qualitative systematic reviews in education.

In Toulmin's (1958, p. 90) words:

Any argument, of whatever kind, will in practice need to have some claim or conclusion whose merits we are seeking to establish... and also some data, information, or facts we appeal to as a foundation for the claim.

To fully capture the logic and rhetoric of evidence claims in qualitative synthesis, this study retains all six components of Toulmin's model, translating them into coding categories used to systematically analyse the internal argument structure of selected systematic reviews. All six elements are included because each one serves a different purpose; leaving any out would obscure part of the reasoning. The Claim tells us what the review asserts, while the Grounds show the evidence behind it. The Warrant explains why the evidence supports the claim, and the Backing shows where the reasoning gets its authority from, for example, a theory or prior research. The qualifier indicates the strength or generality of the claim, while the rebuttal records any limitations and alternative explanations. Together, these six components provide a comprehensive overview of how a claim is formed, its strength, and its limitations, which is precisely what is required to compare reviews and assess their suitability.

- **Main Claim.** This category refers to the central assertion or key conclusion that the review seeks to establish. In Toulmin's (1958) *Model of argumentation*, the claim occupies a central position, with all other components serving to support, limit, or justify it. Identifying the main claim allows for the systematic reconstruction of the review's argumentative structure and facilitates the assessment of its coherence and evidentiary support. According to Gough (2021), the clarity and articulation of the main claim are also essential for evaluating its communicative fitness, that is, its ability to effectively convey the review's purpose and relevance to its intended audience. In the context of systematic reviews, the main claim often takes the form of a synthesised conclusion about a phenomenon (e.g., "Collaborative learning improves students' sense of belonging"), representing the argumentative endpoint of the synthesis.
- **Other Claims.** In many systematic reviews, multiple insights or layered findings are presented alongside the main conclusion. This category captures secondary or related claims that enrich, qualify, or extend the central argument. Toulmin (1958) acknowledges that practical reasoning is often non-linear and may involve multiple claims supported by diverse strands of evidence. In qualitative syntheses, these additional claims frequently address theoretical, contextual, or methodological dimensions. Coding for such claims enables a more nuanced understanding of the review's overall argumentative structure and the complexity of its reasoning.
- **Claim Framing.** This category refers to the manner in which the main claim is articulated, whether it is presented as a theoretical contribution, a practical implication, or a normative stance. Toulmin's (1958) concept of *field-dependent standards* underscores that the standards for evaluating a claim, and the way it is framed, vary according to disciplinary conventions and the expectations of the intended audience. Claims may be cautiously framed to acknowledge limitations in transferability, or assertively positioned to influence policy or practice. Gough (2021) similarly highlights the communicative intent of evidence claims, focusing on how claims are constructed to engage particular audiences, whether to inform theory, shape policy, or guide professional practice. This category thus captures the orientation and discursive function of the claim within the broader context of its epistemological and practical aims.
- **Grounds.** This category captures the foundational data or reasons upon which the main claim is based. In Toulmin's (1958) model, *grounds* refer to the data, information, or facts appealed to in support of a claim. These are not absolute facts but contextually valid forms of evidence that underpin the argument. Within the context of systematic reviews,

grounds typically include thematic findings, interpretive summaries, and participant quotations extracted from the primary studies. Gough (2021) similarly emphasises the importance of grounding evidence claims in clearly articulated and appropriate forms of support, making this category essential for both structural and evaluative analysis. Coding this allows assessing the extent to which claims are directly and substantively supported by evidence.

- **Synthesis Type.** This category identifies the mode of synthesis employed in the review, whether aggregative, configurative, or a hybrid approach, and examines how this method shapes the construction and justification of claims. The synthesis type significantly influences the logic of the argument: aggregative syntheses tend to summarise patterns across studies, while configurative approaches aim to generate new interpretations or conceptual insights (Gough, 2021). Toulmin's (1958) model accommodates this variability by recognising that different fields apply different standards for warranting claims. In the context of qualitative evidence synthesis, the links between data and claims in configurative reviews are often interpretive rather than empirical in a strict sense. Tracking the synthesis type therefore enables an assessment of how the review's reasoning unfolds and the extent to which claims are shaped by interpretive versus descriptive logics. **Validity/Relevance.** This category assesses the extent to which the review reflects on the quality, credibility, and appropriateness of the evidence used to support its claims. Toulmin (1958) distinguishes between formal validity and contextual credibility, emphasising that arguments must be "well-grounded or firmly-backed" in order to "stand up to criticism" (1958, p. 8). In qualitative synthesis, the credibility of claims often hinges on factors such as methodological transparency, depth and richness of data, and the use of techniques such as triangulation. Gough (2021) likewise argues that the *fitness for purpose* of an evidence claim depends not only on its coherence but also on its clarity, relevance, and appropriateness for its intended audience. This category, therefore, captures whether and how the review engages in evaluative reflection on the trustworthiness and applicability of its evidence base.
- **Warrant.** This category identifies the underlying principle or assumption that authorises the inferential leap from evidence to claim. In Toulmin's (1958) model, the *warrant* serves as the logical bridge between the grounds and the claim, "the general rule or principle that authorises the step from the data to the claim" (1958, p. 91). In the context of qualitative systematic reviews, warrants are frequently interpretive rather than deductive, often grounded in theoretical alignment, contextual reasoning, or assumptions about causality. For example, a warrant might take the form of an interpretive generalisation such as: "when

students feel seen by their teachers, they are more likely to engage.” Although Gough’s (2021) framework does not explicitly adopt Toulmin’s terminology, it similarly emphasises the need to scrutinise the reasoning and assumptions that support evidence claims. Coding for warrants thus helps to uncover the logic that links data to conclusion, shedding light on the inferential structure of the review’s argument.

- **Backing.** This category refers to the additional support provided for the warrant, particularly when the warrant is not self-evident or universally accepted. In Toulmin’s (1958) model, *backing* serves to reinforce the warrant by offering justification from external sources, what he describes as what “buttresses the warrant in case of challenge” (p. 92). In the context of qualitative synthesis, backing may include references to established theoretical traditions, prior empirical findings, or disciplinary norms that lend credibility to the warrant. Gough (2021) similarly calls for transparency regarding the epistemological and theoretical foundations upon which evidence claims rest. This category thus captures the broader conceptual or empirical justifications that support the inferential logic of the review, providing insight into the sources of authority and credibility behind its reasoning.
- **Qualifier.** This category captures the degree of certainty or confidence with which a claim is presented. In Toulmin’s (1958) model, the *qualifier* indicates “the degree of force which our data confer on our claim” (p. 93), typically expressed through modal terms such as “possibly,” “likely,” or “strongly.” Qualifiers are essential in practical reasoning, as they signal the author’s assessment of the robustness or generalisability of a conclusion. In qualitative synthesis, the use of hedged or assertive language reflects the author’s awareness of evidentiary limitations, interpretive ambiguity, or contextual specificity. Gough (2021) likewise emphasises the importance of transparency regarding the strength and scope of claims. Coding for qualifiers allows for an examination of how reviewers calibrate their assertions and communicate the reliability or transferability of their findings, which is especially critical in interpretive research contexts.
- **Rebuttal.** This category examines the extent to which the review anticipates counterarguments, acknowledges limitations, or considers alternative interpretations that may constrain the validity of its claims. Toulmin (1958) describes the *rebuttal* as the component that embeds self-critical reflection into the argument, noting that strong arguments include recognition of exceptions or conditions under which the claim may not hold (p. 94). In the context of systematic reviews, rebuttals often appear in the form of limitations sections or discussions of studies with divergent or contradictory findings. Gough (2021) similarly underscores the importance of reflexivity regarding the scope, relevance,

and limitations of evidence claims. Coding for rebuttals enables the identification of critical self-awareness within the review, enhancing the overall transparency, nuance, and integrity of the argument.

By systematically applying Toulmin's model, this study aims to examine not only what is claimed in qualitative systematic reviews, but also how those claims are constructed, justified, and qualified. Toulmin's framework offers a nuanced way to analyse the logical structure of qualitative syntheses, highlighting the role of claims, grounds, warrants, backing, qualifiers, and rebuttals in shaping the argument. Each element reveals different aspects of the reasoning process, from the type and quality of supporting evidence to the assumptions and limitations that frame the claim. Combined with Gough's (2021) framework, which emphasises the contextual relevance, communicative clarity, and intended use of evidence claims, this integrated approach provides a dual lens for analysis. It enables a rich, multi-dimensional examination of both the internal coherence of arguments and the broader epistemological and practical purposes that inform their justification. Together, the Toulmin-derived categories and Gough's interpretive dimensions offer a robust analytical scaffold for assessing the credibility, structure, and communicative fitness of evidence claims in qualitative research synthesis.

3.4 Protocol development

To operationalise the integrated conceptual framework, the study adopted a protocol that is both systematic and flexible. Systematic refers to a rigorous, transparent search and screening of systematic reviews of qualitative studies in education, guided by predefined inclusion and exclusion criteria and a documented workflow (search strings, databases, screening stages, and audit trail). Flexible refers to the purposive sampling used after the search, enabling iterative, criterion-based selection to ensure conceptual fit with the framework and sufficient variation for analysis. The aim was not to replicate the full procedures of the included reviews or to synthesise primary studies, but to analyse how knowledge claims are constructed and justified within qualitative systematic reviews. Accordingly, a systematic search and eligibility process ensured methodological rigour and transparency, while purposive sampling focused the corpus on reviews aligned with the

study's aims and suitable for in-depth analysis. The detailed procedures for both the systematic search and the flexible sampling are set out in the protocol below.

Following the methodological guidance of Campbell et al. (Campbell et al., 2018), the protocol was designed to enhance transparency, reproducibility, and coherence. It included: (a) clearly articulated inclusion and exclusion criteria; (b) a structured and documented search strategy; and (c) a formalised process for screening and data extraction (see Appendix A). As Campbell et al. emphasise, “a protocol should be seen as both a planning document and a tool to enhance transparency and reproducibility” (p. 5). Moreover, as Zawacki-Richter et al. (2020) argue, the protocol is a foundational document in the review process, crucial for managing the inherent “messiness” of educational research and for maintaining coherence amid iterative or exploratory research questions, especially when dealing with qualitative evidence and complex conceptual constructs.

Petticrew & Roberts (2008) emphasise that a systematic review protocol plays a dual role: it establishes both the technical procedures and the conceptual boundaries of the inquiry. According to them, “a systematic review needs a detailed protocol that describes in advance the process and methods that will be applied... including a description of and rationale for the review question” (p. 44). They also argue that the protocol ensures a review's reliability, comparability, and defensibility by requiring reviewers to “set out their methods in advance, and in detail, as one would for any piece of social research” (p. 9).

Moreover, Petticrew and Roberts (2008) acknowledge that in reviews based on configurative synthesis logics, the protocol must remain adaptable. In these cases, the review's conceptual orientation and even its focus may evolve in response to the literature uncovered. This flexibility is especially relevant when the goal is interpretive understanding rather than measurement of effect. In this study, such adaptability was essential to accommodate the analytical depth required by the Toulmin model and the Gough framework. It allowed for fine-tuning the sampling and coding strategies in light of the emerging body of reviews, ensuring that the analysis remained focused on how evidence claims were framed and justified, rather than on evaluating the effectiveness of interventions.

The full details of the search strategy, screening procedures, and eligibility criteria applied in this study are documented in the protocol provided in [Appendix A](#). What follows are the key steps taken to identify and select a relevant pool of systematic reviews of qualitative studies in education, from which a purposive sample was later drawn for in-depth analysis.

To structure the search and ensure conceptual alignment with the study's aims, the PCC framework (Population–Concept–Context) was adopted, which is particularly well-suited to qualitative and exploratory research questions. As Pollock et al. (2023) note, PCC offers a more flexible alternative to the PICO model, especially in education, where review questions often address processes, experiences, and meaning-making rather than interventions or outcomes. *Population*: primary and secondary students and teachers. *Concept*: subjective experiences (attitudes, beliefs, perceptions, opinions, emotions). *Context*: formal educational settings, from primary to secondary school.

The studies included will be located through an extensive and transparently documented search of the ERIC database (Ebsco platform). Given the conceptual and terminological heterogeneity of qualitative evidence synthesis in education, the search will be approached as a structured but iterative process, with search terms tested and refined to balance sensitivity (capturing relevant records) and specificity (excluding irrelevant material), in line with guidance on systematic searching for complex reviews (Levay & Craven, 2018). The reason for searching ERIC is twofold. Firstly, ERIC is an authoritative indexed and full-text education literature and resources database of indexed and full-text education literature containing records and links to hundreds of thousands of full-text documents. It includes records from various sources, encompassing journal articles, books, conference papers, curriculum guides, dissertations, and policy papers. Secondly, this systematic search strategy aimed to retrieve much literature, such as systematic qualitative reviews in education. The results were exported into Rayyan, a web-based tool for systematic reviews, where titles, abstracts, and, when necessary, full texts were screened. To enhance rigour, a portion of the studies was independently reviewed and discussed with the supervisory team³. The entire selection, search and screening process is documented in Chapter 4.

3.5 Codebook development

The construction of the codebook (Appendix B) proceeded in two linked phases. In the first phase, Toulmin's (1958) model of argumentation and Gough's (2021) framework for assessing the fitness for purpose of an evidence claim were used independently as initial scaffolds. This produced a set of provisional categories capturing the internal structure of

³ The supervisory team consisted of the author, the supervisor, and the co-supervisor

reasoning (claim, grounds, warrant, backing, qualifier and rebuttal) and the contextual and purposive dimensions of claims (topic, importance, justification of method, theoretical framing and justification, audience and scope).

In the second phase, these provisional categories were integrated through the iterative development of the study's conceptual framework. This involved aligning overlaps (e.g. qualifiers with strength/scope and rebuttals with consideration of alternatives), resolving gaps (e.g. distinguishing between warrant and backing and grounds) and establishing the analytic sequence (first applying Gough's categories to determine purpose, audience and scope, and then applying Toulmin's to reconstruct the inferential structure). Operational definitions, inclusion/exclusion rules and indicators were then written for each category, paying attention to both explicit and implicit instances. The integrated framework thus became both the analytical lens and the structural backbone of the codebook. Gough's categories orientated the coding towards relevance and intended use, while Toulmin's categories specified how claims were constructed and justified. Pilot coding and memo-writing informed minor refinements, ensuring that the final codebook was coherent, non-overlapping and practical for comparative analysis across reviews (Glaser et al., 2003).

Each codebook category was developed with reference to these theoretical principles and operationalised to capture specific aspects of how evidence claims are articulated and justified. For example, the categories Main Claim, Warrant, and Backing reflect Toulmin's concern with the internal structure of arguments, while Importance, Justification of Method, and Framework Justification correspond to Gough's emphasis on the value-laden assumptions and communicative aims behind claims. In developing and refining the codebook, additional methodological guidance was drawn from established literature on qualitative synthesis and coding strategies. Thomas and Harden (Thomas & Harden, 2008) outline a clear process for thematic synthesis involving line-by-line coding, the development of descriptive themes, and the generation of interpretive insights, a model particularly suited to reviews aiming to construct rather than aggregate meaning. Similarly, Sandelowski & Barroso (2007) stress the interpretive and integrative nature of synthesis, in which coding must reflect not only recurring patterns but also conceptual linkages across diverse studies. (Finfgeld-Connett, 2014) reinforces this view by showing how content analysis can support both theory-building and knowledge integration in qualitative systematic reviews. To ensure methodological transparency and consistency, the codebook also incorporates principles from (Miles et al., 2014), including the use of operational definitions, analytic memos, and

iterative refinement cycles. These strategies enabled a systematic yet flexible coding approach that accommodated the complexity of the data while remaining aligned with the conceptual aims of the study. Ultimately, the codebook was not merely a tool for data categorisation but a reflexive and theory-informed framework for analysing how qualitative reviews construct their claims. It supported a multidimensional reading of the reviews, examining both what is claimed and how it is justified, why it matters, and to whom the claim is addressed. This aligns precisely with Gough's concern for contextual coherence and Toulmin's commitment to practical reasoning, ensuring that each review was assessed not only on technical merit but also on its argumentative clarity and epistemic integrity.

4 Constructing Claims: A Comparative Analysis of Six Systematic Reviews

This chapter presents the results of a qualitative comparative analysis of six systematic reviews of qualitative educational studies, which were selected through purposive sampling. Each review was analysed using a structured codebook (see [Appendix C](#) for the full codebook analysis), which was developed from the adapted framework introduced in Chapter 3. This integrated conceptual framework combines Toulmin's (1958) Model of Argumentation with Gough's (2021) Fitness for Purpose of an Evidence Claim Framework. It offers a tool with which to investigate how systematic reviews construct, frame and justify their evidence claims, including both explicit elements (e.g. topic, importance, synthesis type) and more implicit ones (e.g. warrant, claim framing, rebuttal).

Sections 4.1 to 4.6 report the detailed findings of each review (SR#1 to SR#6), tracing how claims are shaped by theoretical orientation, methodological strategy and the nature and treatment of the evidence base. This in-depth analysis of the systematic reviews reveals how each review develops its reasoning and the extent to which its claims are explicitly justified, qualified or left implicit and unexplained. Section 4.7 then brings these analyses together in a cross-case synthesis, identifying patterns, gaps and points of divergence across the sample. Particular focus is given to components that were most frequently implicit, particularly warrants, claim framing and rebuttals, and to how the logic and purpose of claims vary depending on the intended contribution of each review. This layered structure provides a thorough understanding of how evidence claims operate in qualitative synthesis by evaluating individual argumentation strategies and the broader norms that influence review writing in educational research.

4.1 Search strategy and study selection.

As outlined in Chapter 3, the included studies were identified through an extensive search of the ERIC database. The PRISMA flow diagram (Figure 4) shows that 335 records were initially retrieved, of which 306 were excluded for failing to meet one or more eligibility criteria. This yielded 28 reviews, which were exported to Excel for initial coding. Following full-text screening, three additional studies were excluded. The final dataset comprised 25

systematic reviews, which were included for general descriptive coding and from which a purposive sample was drawn for deeper analytical work.

The study adopted purposive sampling to address the breadth and heterogeneity of qualitative systematic reviews in education. Purposive sampling is well suited to qualitative research synthesis because it enables the deliberate selection of “information-rich cases from which one can learn a great deal about issues of central importance” (Patton, 2002, as cited in Suri, 2011, p. 64). In this context, the priority was interpretive depth, variation, and conceptual richness rather than statistical generalisability—an orientation Suri (2011) identifies as appropriate when the goal is to understand how knowledge is constructed and justified across texts.

To ensure alignment with the study’s aim—examining how reviews construct and justify evidence claims—a criterion-based approach was applied within the purposive logic (Suri, 2011). Criteria were theoretically and methodologically grounded to maintain fidelity to the integrated framework derived from Toulmin and Gough. Reviews were included when they:

- were systematic reviews or meta-syntheses;
- drew exclusively on qualitative primary studies; and
- focused on the subjective experiences of students and/or teachers in primary or secondary educational settings.

This combination of purposive and criterion-based sampling secured a corpus that was conceptually coherent, methodologically comparable, and sufficiently varied to support close analysis of argumentative strategies and the fitness-for-purpose of resulting claims.

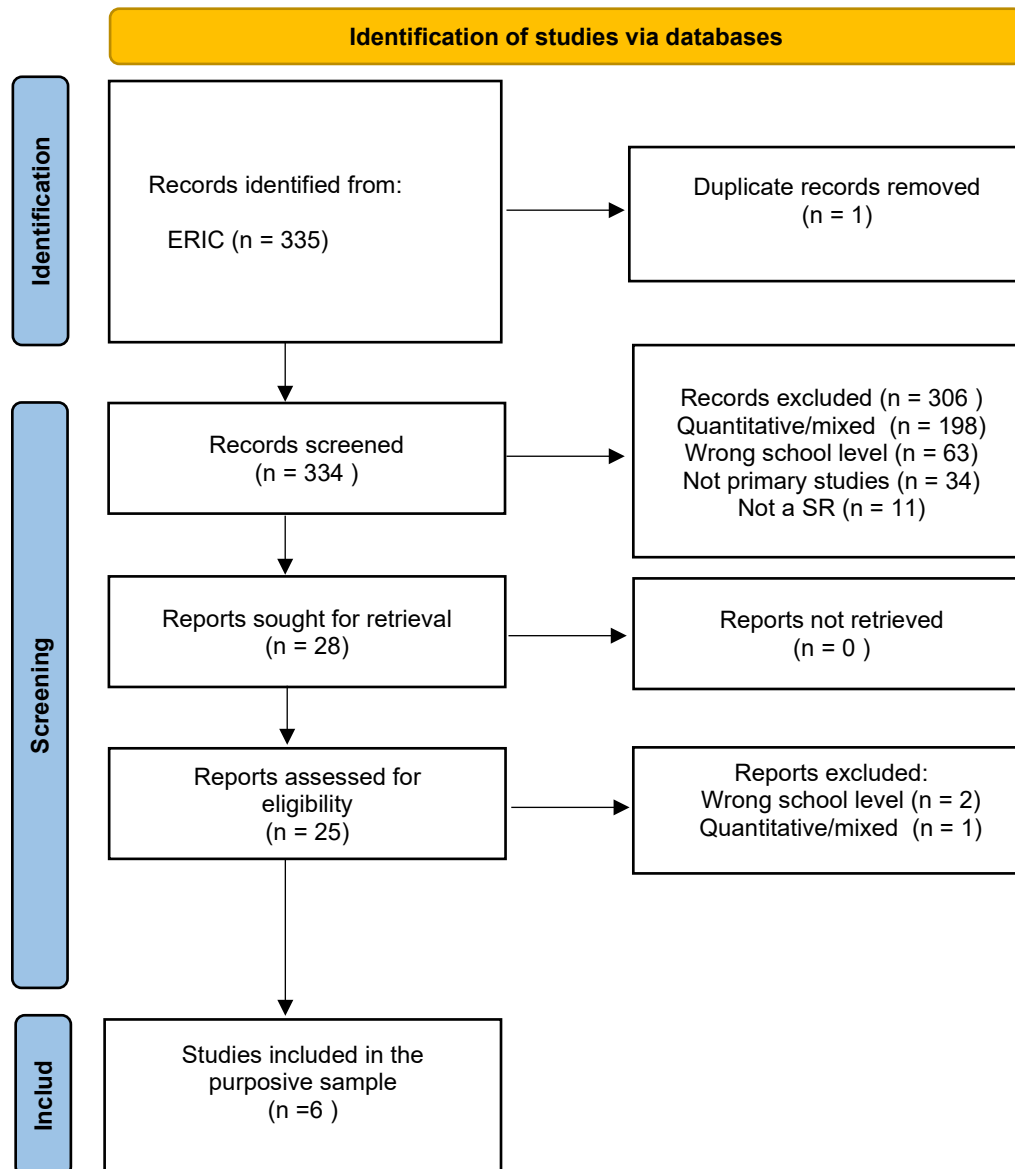
This approach ensured the methodological coherence and conceptual relevance of the sample, allowing the study to focus on reviews most likely to exhibit diverse yet theoretically salient ways of constructing evidence claims. While Patton (2002) proposed 16 purposeful sampling strategies for use in primary qualitative research, he did not explicitly address their application in research synthesis. In response to this gap, Suri (2011) explored how Patton’s strategies might be adapted for synthesising research. She emphasised that, in the context of synthesis, the “participants” are not human subjects but rather textual sources, systematic reviews, studies, or reports. The task thus shifts from eliciting personal narratives to interpreting documents in order to generate new insights about how knowledge is constructed, communicated, and justified across a body of literature. In line with this adaptation, this study also employed elements of maximum variation sampling, a strategy designed to capture diverse manifestations of a phenomenon within a bounded conceptual

frame. As Suri (2011) explains, maximum variation sampling involves identifying cases that share core inclusion characteristics but vary widely across other dimensions. In this project, the selected systematic reviews were all methodologically aligned (qualitative systematic reviews in education) but varied in their population, synthesis methods, and contextual focus. This strategy enhanced the interpretive power of the analysis by enabling comparisons across a broad range of review styles and argumentative strategies, while remaining anchored in a clearly defined analytical purpose. Zawacki-Richter et al. (2020) further support this flexible, iterative approach to protocol development, especially within educational research, where reviews often address complex and "fuzzy" concepts and where configurative synthesis logics are common. Their reflections on systematic reviews in education echo the need for balance between methodological rigour and conceptual flexibility. Ultimately, the protocol and sampling strategies in this study were designed not only to identify a manageable and coherent set of reviews, but also to align methodologically and epistemologically with the study's purpose: to examine how evidence claims are constructed and communicated in qualitative systematic reviews within educational research.

To obtain targeted, in-depth insights while maintaining alignment with the research aim, the sample concentrated on qualitative systematic reviews that directly addressed the study's focal topic and met the core inclusion criteria. Within this bounded corpus, maximum variation sampling was applied in order to preserve relevance while deliberately widening the comparative space needed to examine how evidence claims are framed and warranted across different review traditions (Suri, 2011; see Table 1). The sampling logic assumed that the internal structure and communicative force of evidence claims are shaped not only by topic and population, but also by paradigmatic orientation, synthesis approach, degree of theoretical framing, and contextual positioning; accordingly, the selected reviews were chosen to ensure shared methodological alignment at the level of "qualitative systematic review," while incorporating diversity across populations, synthesis tools, and settings. In practical terms, this meant including reviews oriented toward different knowledge functions - such as practice- and implementation-focused syntheses that foreground actionable implications (e.g., SR A on CLIL teachers' professional development needs; SR C on movement integration in elementary classrooms), alongside a policy-facing review where claims are closely entangled with normative and institutional concerns (SR B on intimate partner violence education policy and practice) as well as ensuring methodological contrast through the inclusion of a configurative synthesis (SR D, a meta-ethnography of qualitative

research on school environment and student health). Paradigmatic diversity was further strengthened by incorporating reviews in which identity, culture, and contextual positioning are central to claim-making, including the review of nonnative-English-speaking teacher candidates' language teacher identity development (SR A1) and the review of curriculum and learning in Aboriginal and Torres Strait Islander education (SR B1), where claims may be shaped by culturally situated and potentially critical interpretive commitments. This strategy also enabled variation within apparently similar review objects: for example, reviews may focus on a comparable population (e.g., teachers) while differing in synthesis approach and geographic context. Bringing together such variation strengthens the analysis by assembling evidence claims produced under diverse yet relevant conditions, enabling systematic comparison of argumentative strategies and of the fitness-for-purpose of resulting claims, and ultimately supporting a more comprehensive understanding of the phenomenon under investigation (Nyimbili & Nyimbili, 2024).

Figure 4
PRISMA 2020 flow diagram



From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: <http://www.prisma-statement.org/>

Table 1

Author	Title	Sample label	Population focus	Review/synthesis approach	Sampling rationale
Kim, Haemin	CLIL Teachers' Needs and Professional Development: A Systematic Review	A	Teachers	Qualitative systematic review	Core case (baseline)
Jamal, Farah	The school environment and student health: a systematic review and meta-ethnography of qualitative research	D	Students	Meta-ethnography (configurative)	Method contrast case (configurative synthesis)
Sheng, Xiaomin	Education Policy and Practice on Intimate Partner Violence among Young People in the UK	B	Students / young people	Qualitative systematic review	Core case (baseline)
Michael, Robert Dan	Facilitators and Barriers to Movement Integration in Elementary Classrooms: A Systematic Review	C	Students and teachers	Qualitative systematic review	Population-scope contrast (multi-actor focus)
Swearingen, Amanda J.	Nonnative-English-Speaking Teacher Candidates' Language Teacher Identity Development in Graduate TESOL Preparation Programs: A Review of the Literature	A1	Nonnative-English-speaking teacher candidates (TESOL)	Review of the literature / qualitative synthesis	Variant of A (same broad population/topic; different tools/geography)
Harrison, N	Curriculum and learning in Aboriginal and Torres Strait Islander education: A systematic review	B1	Students / learners in Aboriginal and Torres Strait Islander education	Systematic review	Variant of B (different geographic/context focus; topic emphasis differs)

4.2 SR#1 – Facilitators and barriers to movement integration in elementary classrooms: a systematic review (Michael et al., 2019)

Michael et al. (2019) conducted a systematic review investigating facilitators and barriers to movement integration in elementary classrooms. The title and introduction explicitly frame the topic, focusing on how physical activity strategies are implemented in instructional contexts and how teacher and context-related variables influence this process. The review responds to pressing educational and public health concerns, particularly the impact of sedentary behaviour on children's development. The authors underscore the importance of movement integration by referencing national physical activity guidelines and citing its cognitive and behavioural benefits. The review is grounded in a systematic review protocol guided by a configurative synthesis approach. The authors justify this choice by emphasising the fragmented nature of existing research and the need to generate conceptual insights across studies. The synthesis is structured using the *Social Ecological Model (SEM)* (McLeroy et al., 1988), which is explicitly named and justified as a framework capable of organising influences at individual, interpersonal, institutional, and policy levels. The main claim of the review is that the most influential facilitators for movement integration are institutional and contextual, such as supportive leadership, teacher autonomy, and school-wide culture, rather than individual teacher motivation alone. The review also asserts that barriers such as time constraints, curriculum pressures, and lack of training are widespread and systemic. These claims are practically framed, with clear implications for policy and teacher professional development. The grounds for these claims are drawn from a thematically coded synthesis of qualitative studies. Evidence is presented through summary tables, themes aligned with SEM layers, and illustrative quotations. Although the logic connecting data to claims is mainly implicit, the coherence of findings across diverse contexts functions as a *de facto* warrant. The review engages with quality and relevance by reporting variation in methodological rigour across studies, although it does not conduct a formal quality appraisal. Claim strength is signalled through cautious modal language (“suggest,” “may influence”), consistent with the interpretive stance. Rebuttals are acknowledged as contextual limitations, particularly regarding generalizability and implementation variability across schools. However, alternative explanations or contradictory findings have not been extensively developed. The review demonstrates a

strong alignment between its aims, methodological framework, and synthesised conclusions. Its structured use of SEM enhances the coherence of the analysis, and its practical orientation makes the findings relevant for researchers and educational stakeholders interested in translating evidence into practice.

4.3 SR#2 – Learning spaces, agency and notions of improvement: an interpretive meta-ethnography (Savin-Baden et al., 2008)

This systematic review, conducted by Savin-Baden, McFarland, and Savin-Baden (2008), centres on how thinking and practices in higher education are shaped by various influences, with a thematic structure grounded in *practice*, *transfer*, and *community*. The review is positioned within the context of pedagogical reform in higher education and problematizes existing literature through an interpretive meta-ethnographic approach. The authors identify a significant conceptual gap in the field, describing it as “a minefield of overlapping concepts, with few clear frameworks for understanding the relationship across practice, transfer and community” (p. 212). The topic’s importance is further emphasised by its ambition to uncover hidden or marginalised subtexts, such as pedagogical stance, disjunction, agency, and learning spaces, that exert considerable influence on academic practices yet often remain unexamined in teaching and learning research. The methodological justification is firmly articulated: the authors critique traditional meta-analysis for its inability to preserve contextual and conceptual depth in qualitative educational research. Instead, they adopt interpretive meta-ethnography, which enables a more nuanced synthesis of diverse literature while retaining interpretive richness. The review is theoretically grounded in the critical pedagogy of (Freire & Torres, 2023), the feminist theory of Hooks (2014), and Barnett’s model of curriculum complexity (Bengtson, 2018). Giddens’ (1984) theory of structuration is also central, especially in framing the interplay between agency and institutional structures. These frameworks are explicitly integrated and justify both the methodological and interpretive choices made in the synthesis. The main claim advanced by the authors is that a set of under-acknowledged factors, particularly pedagogical stance, disjunction, learning spaces, agency, and communities of interest, play a foundational yet overlooked role in shaping educational practice. These influences, they argue, are “sometimes ignored, marginalised or dislocated from the central arguments about teaching and learning thinking and practices in higher education” (Savin-

Baden et al., 2008, p. 225). Additional claims extend this insight by emphasising how individual academic identities, institutional structures, and disciplinary contexts interact in shaping the possibilities for innovation and reform. The review cautions against simplistic assumptions about the efficacy of either top-down or bottom-up improvement models, offering nuanced examples where both approaches succeeded or failed depending on contextual variables. The claims are framed mainly in normative and practical terms, with implications for educational development, policy, and further research. The grounds for these claims are built through a three-tiered meta-ethnographic synthesis (Noblit & Hare, 1988), analysing over 100 qualitative studies. Tables and matrices map methodological characteristics, thematic patterns, and second- and third-order interpretations (Savin-Baden et al., 2008, pp. 216–219). The warrant linking data to claims is primarily implicit, resting on the coherence and recurrence of key concepts across the reviewed literature. The argument is supported by theoretical backing from critical and constructivist traditions, particularly regarding agency, identity, and reflective practice in HE. Claims are qualified through tentative language such as “suggests,” “may,” and “appears to,” although the overall tone conveys interpretive confidence grounded in recurring patterns. The authors explicitly acknowledge limitations in both the methodology and the reviewed literature, including transparency issues, the challenge of synthesising across traditions, and the risk of privileging thematic similarity at the expense of contestation. Nevertheless, the review constructs a compelling interpretive narrative that enriches our understanding of educational reform by illuminating the complex, layered influences on teaching and learning practices in higher education.

4.4 SR#3 – Curriculum and learning in Aboriginal and Torres Strait Islander education: a systematic review (Harrison et al., 2019)

Harrison et al. (2019) offer a conceptually rich and critically engaged systematic review that explores how the Australian curriculum governs learning for Aboriginal and Torres Strait Islander students. Grounded in the broader *Aboriginal Voices* research project, this review seeks to test prevailing historical narratives against empirical research and practice while surfacing alternative ontological and epistemological perspectives on curriculum. Central to the review is the question: *How does the curriculum shape and regulate what counts as learning for Aboriginal and Torres Strait Islander students?* (Harrison et al., 2019, p. 234).

The importance of this investigation lies in its response to how national educational policies, particularly those emphasising standardised testing such as (The National Assessment Program – Literacy and Numeracy – NAPLAN), have marginalised more profound educational questions related to identity, cultural continuity, and social belonging. Harrison et al. critique the dominance of output-driven definitions of success, arguing that such models fail to resonate with the values and aspirations of many Indigenous communities (p. 243–245). They highlight the sharp disjuncture between the knowledge privileged by the Australian Curriculum and the understandings of country, identity, and learning held by Aboriginal and Torres Strait Islander peoples. Methodologically, the review adheres to a rigorous systematic protocol developed for the *Aboriginal Voices* project. A multi-stage appraisal process evaluated 886 studies, ultimately synthesising 29 qualitative studies published between 2006 and 2017 (p. 236–237). All included studies were examined through an interpretive, thematic synthesis framework. The justification for this approach is clear: given the relational, cultural, and contextual dimensions of Indigenous learning, interpretive synthesis is positioned as more appropriate than aggregative or meta-analytic methods (p. 234). Theoretically, the review draws on a range of critical and Indigenous scholars to frame curriculum not as a neutral or static body of knowledge, but as a *representational practice* (Green, 2018) embedded in broader ontological assumptions. The authors cite Verran’s (2013) argument that Indigenous and Western knowledge systems are sometimes “irreconcilable,” and that curriculum design must account for multiple ontologies. They also draw on Barad’s (2007) notion of relational agency and Country-informed pedagogy, in which land, seasons, and local knowledge are seen as active agents in the learning process (Harrison et al., 2019, pp. 241–246). The review’s central claim is that current curriculum structures in Australia are not fit for purpose in Indigenous education. Rather than merely including Aboriginal content or perspectives, the curriculum must be reconceptualised around Indigenous ontologies of knowledge and learning. Harrison et al. argue for curriculum as *Country*: an enacted, relational, and place-based process that includes stories, practices, and relations to land, not just content delivered in a classroom (p. 245–247). They suggest that learning is not an individual pursuit governed by success metrics, but a social practice grounded in relational accountability and local context. Additional claims further support this perspective. Studies such as those by Guenther et al. (2015) and Ewing (2012, 2014) are used to demonstrate how educational success is defined not by test scores or attendance figures, but by community engagement, identity affirmation, and intergenerational transmission of knowledge (p. 244). Guyula (2010) and McKnight (2016a)

show how learning happens through immersion in Country, and how formal schooling often fails to accommodate this mode of knowing. The review’s interpretive synthesis draws on a wide range of empirical studies from diverse settings, urban, rural, and remote, showing thematic consistency around core concerns: identity, representation, and success. Using case examples (e.g., seasonal calendars on Erub Island, Torres Strait; Montessori methods adapted to Indigenous zoology education) serves to ground the analysis in concrete practice (pp. 246–247). While the review’s warrants are often implicit, the strength of convergence across studies and communities makes the argument persuasive. Claims are cautiously framed, using language such as “may,” “tend to,” and “appear to,” but are underpinned by a robust body of theory and evidence. The authors acknowledge the limitations of dominant paradigms, particularly their emphasis on measurable outcomes, and offer a critical rebuttal of these assumptions. They conclude by calling for a reconciled national learning framework that recognises multiple ways of knowing and repositions Indigenous knowledge systems as foundational rather than supplementary (p. 248–250). In sum, Harrison et al. (2019) advance a compelling argument for rethinking curriculum through the lens of relational ontology and Indigenous epistemology. The review critiques current systems’ limitations and provides a roadmap for more equitable, culturally sustaining, and pedagogically grounded approaches to Indigenous education in Australia.

4.5 SR#4 CLIL teachers’ Needs and Professional Development: A Systematic Review (Kim & Graham, 2022)

Kim & Graham (2022) present a comprehensive and well-structured systematic review that synthesises research on the professional development (PD) needs of CLIL (Content and Language Integrated Learning) teachers. The study identifies how teachers’ self-reported needs correspond to or fail to correspond to the literature on CLIL-oriented professional development. Drawing on Cañado (2018) seven-competence framework (linguistic, pedagogical, scientific knowledge, organisational, interpersonal, collaborative, and reflective), the authors organise their thematic synthesis around a structured analytic matrix, revealing both strengths and persistent gaps in teacher support systems (Cañado, 2018, pp. 3–4). The authors make a compelling case for the urgency and relevance of the review. They argue that although CLIL continues to grow globally, teacher development often lags behind policy expectations, with many educators being asked to “use CLIL” (Ting, 2011, p. 314) without adequate training or theoretical grounding. This leads to variability in implementation and undermines the potential of CLIL pedagogies. The review is

methodologically rigorous: it draws on six electronic databases, incorporates forward and backwards searches, and applies eight inclusion criteria to narrow an initial set of 4,227 articles to 43 peer-reviewed empirical studies, all situated in primary and secondary education contexts (Kim & Graham, 2022, pp. 5–6). The theoretical framework is clearly defined and justified. The authors rely on Cañado’s (2018) model to ensure coverage of both technical (e.g., linguistic, pedagogical) and relational (e.g., collaborative, interpersonal) aspects of CLIL teaching. Their central claim is that although CLIL teacher development literature has grown, it does not adequately address the full spectrum of teacher-reported needs. For example, while pedagogical and linguistic competencies are widely discussed, interpersonal and collaborative competencies remain significantly underrepresented (Kim & Graham, 2022, pp. 12–14). Their findings support this critique with empirical data. The review categorises studies into three groups: those addressing teacher needs (n=33), those examining PD (n=9), and one that covered both. Many studies report general dissatisfaction with linguistic proficiency, particularly in subject-specific vocabulary (Yildiz, 2019) or in managing classroom discourse in English (Kewara & Prabjandee, 2018; Lo, 2019). Teachers also express challenges in balancing content and language integration, especially those trained primarily as language teachers, highlighting a strong demand for deeper pedagogical training (Barrios & Milla Lara, 2020; Torres-Rincón & Cuesta-Medina, 2019). Professional development programs tend to address these concerns unevenly. While some PD initiatives include strong theoretical and methodological components (e.g., Cammarata & Haley, 2018; Francomacaro, 2019), others offer limited support in materials development, classroom management, or cross-disciplinary collaboration. The authors note, for example, that reflective and collaborative competencies, though often seen as byproducts of PD, are rarely taught explicitly (p. 13–15). Interpersonal competence, which pertains to responding to students’ affective and identity-based needs, is strikingly absent across the literature, appearing in only one teacher’s needs study (Pladevall-Ballester, 2015).

The review’s claims are grounded in the recurring patterns observed across diverse studies from Europe, Asia, Latin America, and North America. Using Cañado’s framework as a coding tool lends analytic consistency to the synthesis. Warrants are primarily implicit, but the strength of thematic convergence reinforces the credibility of the conclusions. The review’s value also lies in its backing: it draws on a growing body of CLIL theory (Coyle & Meyer, 2021), language pedagogy (Lo, 2020), and teacher identity research (Cammarata & Tedick, 2012). While the authors adopt a cautiously optimistic tone, using qualifiers such as “suggest” or “may indicate” when discussing under-addressed competencies, they also issue

clear calls for improvement. They recommend that future PD programs systematically incorporate all seven competencies and that teacher training be adapted to specific contexts and teacher profiles, rather than relying on generic formats. They also acknowledge limitations, particularly the English-only scope of the review and the possibility of missing studies not indexed in major databases (Kim & Graham, 2022, p. 15–16). In summary, Kim & Graham (2022) contribute a methodologically sound and conceptually rich review that bridges two essential yet often disconnected domains: teacher-reported needs and professional development design. By identifying gaps and good practices, the review lays a strong foundation for improving CLIL teacher training and ultimately, for strengthening the pedagogical integrity of CLIL programs worldwide.

4.6 SR#5 – Education policy and practice on intimate partner violence among young people in the UK: a systematic review (Sheng, 2020)

Sheng (2020) presents a qualitative systematic review that examines educational policies and school-based practices addressing intimate partner violence (IPV) among young people in the UK. The review is motivated by the observed disparity between the growing global awareness of IPV, particularly in the United States, and the relative underdevelopment of this field within the UK educational context. Sheng identifies a clear gap in the evidence base. While some existing systematic reviews focus on interventions from outside the UK or non-educational settings, few directly assess homegrown interventions implemented in UK schools. This review seeks to address that gap by synthesising empirical evidence on how educational policies and interventions have been used to mitigate IPV among young people aged 10 to 19 (Sheng, 2020, pp. 98–100). The study adopts a rigorous systematic review methodology, screening 1,161 ERIC, BEI, and Scopus records, complemented by manual grey literature searches. A total of 11 studies were included based on eight inclusion criteria, such as relevance to educational settings, use of qualitative methods, and UK-based context (pp. 101–103). The studies span a variety of school-based IPV interventions, including drama-based programmes (e.g., Bell & Stanley, 2006), curricular enhancements (e.g., the “Relationship without Fear” programme), and digital resources (e.g., the “Think U Know” campaign). All were thematically analysed using the Thomas and Harden (2008) three-stage approach: line-by-line coding, descriptive theme construction, and analytical synthesis (p. 111). The central claim advanced by Sheng is that UK educational interventions

to address IPV among youth are fragmented, small in scale, and lacking in longitudinal evaluation. While most interventions show promising results in changing attitudes, improving awareness, and promoting help-seeking behaviours, few demonstrate robust evidence of long-term behavioural change (Sheng, 2020, pp. 112–114). For example, studies such as Gadd et al. (2014) and Fox et al. (2016) report that programmes like “Relationship without Fear” successfully reduced acceptance of IPV and improved recognition of abusive patterns. However, the review emphasises that much of this evidence is attitudinal, and longitudinal studies are needed to assess whether increased awareness leads to lasting behavioural shifts (Sheng, 2020, p. 112). Secondary claims highlight further challenges: many interventions adopt a “one-size-fits-all” approach, fail to accommodate gender diversity, and lack content on same-sex relationships or digital forms of abuse (Dobson & Ringrose, 2016; Donovan & Hester, 2008). Students in several studies expressed discomfort with content that presented gender in binary or stereotypical terms, particularly boys who felt blamed or marginalised in sessions framed exclusively around male perpetrators and female victims (Fox et al., 2014; Gadd et al., 2014). Furthermore, teachers were often underprepared or uncomfortable when delivering such material, raising questions about whether external experts or peer-based models might be more effective in specific contexts (Sheng, 2020, p. 113). The grounds for the review’s claims lie in the comparative analysis of the 11 included studies, which draw from diverse UK locations and methodological approaches. These include pre- and post-test surveys, focus groups, and interviews with students and educators. While thematic convergence across studies lends credibility to the claims, particularly around attitude change and help-seeking, warrants remain largely implicit, relying on repetition and coherence across findings rather than formal inferential logic. Sheng’s discussion is supported by relevant UK policy frameworks, including the Home Office’s *Ending Violence Against Women and Girls Strategy* (2016–2020) and the Department for Education’s statutory guidance on Relationships and Sex Education (DfE, 2019), both of which are presented as promising but not yet fully realized (Sheng, 2020, pp. 98–100, 114–115). The review’s conclusions are appropriately cautious. It acknowledges limitations such as language bias (English-only), the small scale of most interventions, and the need for more rigorous evaluation frameworks. Claims are qualified using terms such as “suggests” and “may promote,” and there is consistent attention to the need for more inclusive, long-term, and scalable approaches. Notably, Sheng calls for developing a national curriculum strategy on IPV that is gender-sensitive and reflexive, drawing on student voice and inclusive pedagogies rather than prescriptive delivery models (Sheng, 2020, pp. 114–

115). In sum, Sheng (2020) offers a contribution to the emerging field of IPV prevention through education in the UK. The review advances a strong argument for moving beyond piecemeal interventions and toward systemic, research-informed educational practices that recognise the complex realities of young people’s relationships and the structural dynamics that shape them.

4.7 SR#6 – Non-native-English-speaking teacher identity in TESOL programs: a systematic review (Swearingen, 2019)

Swearingen (2019) presents a highly nuanced systematic review of 17 qualitative studies investigating the development of language teacher identity (LTI) among nonnative-English-speaking teacher candidates (NNES-TCs) in graduate-level TESOL programs in English-dominant countries. The central aim of the review is to understand how institutional discourses, sociocultural experiences, and teacher preparation practices shape and mediate NNES-TCs’ evolving sense of self as legitimate English language teachers. Drawing from sociocultural and poststructuralist theories of identity (e.g., Norton, 1997; Varghese et al., 2005), Swearingen conceptualises identity as dynamic, multiple, relational, and often fraught with contradiction, particularly when negotiated under the ideological weight of native-speakerism, racialization, and gendered expectations (Swearingen, 2019, pp. 2–3). The author argues that NNES-TCs’ LTI development is a complex interplay between systemic marginalisation and agentive identity construction. The dominant discourses of native speaker authority and linguistic deficiency persist in many TESOL programs. However, NNES-TCs resist, reinterpret, and sometimes subvert these through reflective practices, counter-discourses, and localised pedagogical commitments (pp. 5–10). The review categorises findings into four analytical themes: (1) (non)native speaking and the native speaker fallacy, (2) racialised and gendered identities, (3) academic identity clashes, and (4) the emotional “glue” of identity development. The most prevalent theme, the native speaker fallacy, illustrates how NNES-TCs are discursively positioned as “less legitimate” users and teachers of English, often leading to internalised feelings of inferiority (Pavlenko, 2003; Golombek & Jordan, 2005; Kim, 2011). Swearingen (2019) notes that this ideological framing frequently undermines teacher candidates’ confidence, especially when accented speech or perceived nonstandardness is equated with reduced competence (Swearingen, 2019, pp. 5–6). However, programs incorporating critical reflections, exposure to multilingual pedagogy, and recognition of multicompetence (Cook, 1999) can provide symbolic resources supporting NNES-TCs in affirming their legitimacy and developing

more robust professional identities (Aneja, 2016; Park, 2012). Another important theme is the racialised and gendered dimension of identity work. NNES-TCs reported experiencing race-based stereotyping, silence, and cultural essentialism during their programs, especially when their teaching legitimacy was questioned on racialised or linguistic grounds (Johnson, 2001; Park, 2015). Gender also emerged as a significant factor: for some, like the Korean NNES-TC in Park (2009), identity formation involved reconciling patriarchal cultural norms with emerging feminist professional aspirations. Such intersectional identity work was often emotionally taxing but also transformative when institutions offered recognition and support for diverse positionalities (Swearingen, 2019, p. 8). Swearingen highlights that academic identity negotiation was equally crucial. Many NNES-TCs reported dissonance between their expectations of graduate study and the demands of teacher education programs, particularly around participation, scholarly discourse, and language proficiency. These tensions sometimes resulted in a loss of professional confidence or silence in class discussions (Cho, 2013; Pavlenko, 2003). Nevertheless, programs that validated prior knowledge and fostered student voice promoted a more integrative and empowering identity development process (Ilieva & Waterstone, 2013). Emotion, finally, is presented as the “glue” of LTI development (Shahri, 2018). Positive emotional experiences, often rooted in supportive relationships, reflective writing, and classroom practice, helped consolidate NNES-TCs’ teacher identities. Conversely, negative emotions stemming from marginalisation or perceived incompetence constrained identity formation and contributed to professional withdrawal or self-doubt (Pavlenko, 2003; Brutt-Griffler & Samimy, 1999). Emotions thus emerge not as by-products but as central forces in shaping how teacher candidates interpret and perform their identities (Swearingen, 2019, pp. 9–10). The review’s claims are substantiated through careful thematic analysis, but Swearingen acknowledges the implicit nature of many warrants and the risk of essentializing NNES-TCs as a homogeneous group. The review advocates for TESOL programs to embrace critical, anti-oppressive pedagogies and recognise teacher identity’s affective and intersectional dimensions. It recommends a curriculum that includes narrative and multimodal reflections, collaborative inquiry, and explicit instruction in classroom discourse as part of a broader institutional commitment to decolonising language teacher preparation (Swearingen, 2019, pp. 12–14). In conclusion, Swearingen (2019) contributes a timely and theoretically robust synthesis that challenges dominant paradigms in TESOL education. By centring identity as a core construct of teacher preparation, the review calls for structural change in how programs are designed, evaluated, and lived by NNES-TCs. It urges the field to move beyond

static notions of nativeness and toward dynamic, plural, and agentive models of becoming a legitimate and empowered English language teacher.

4.8 Comparative synthesis across systematic reviews (SR#1–SR#6)

This section provides a synthesis of the findings from the six systematic reviews, summarising the detailed analyses presented in Sections 4.1 to 4.6. Each review was coded using the structured tool developed in Chapter 3, which integrates Toulmin's (1958) model of argumentation and Gough's (2021) Fitness for Purpose of an Evidence Claim Framework. As detailed in Appendix C (Codebook_Analysis), the reviews were systematically coded for explicit and implicit features. This allows for a detailed comparative analysis of how evidence claims are constructed, justified and communicated in different contexts.

Toulmin's model provided a vocabulary and structure for identifying the internal logic of claims, such as the presence or absence of grounds, warrants, qualifiers and rebuttals, while Gough's framework enabled the appropriateness of claims to be evaluated in relation to the stated purpose, audience and epistemological stance of each review. Together, these models support a dual lens: one structural and one evaluative. This approach moves beyond standard notions of 'quality' or 'rigour' in review methodology and instead interrogates how the reasoning in each review is shaped by its positioning, intended use and knowledge function.

The comparative analysis is organised thematically across core analytical categories. It begins with clarity of topic and importance, followed by justification of methods and synthesis type, use of theoretical frameworks, and treatment of validity and relevance. Particular focus is given to dimensions that proved analytically complex or were often implicit, especially claim framing, warrants and rebuttals. While these components are rarely formalised in qualitative synthesis, they play a crucial role in how evidence claims are received, evaluated and ultimately used.

Across the corpus, the reviews demonstrate a variety of reasoning strategies, degrees of conceptual clarity and rhetorical choices. Some, such as SR#1 and SR#2, offer relatively transparent and practical claims, while others, such as SR#3 to SR#6, embed their conclusions in layered theoretical narratives that require careful interpretation to understand. While this variation reflects the diversity of aims and traditions within qualitative educational research, it also reveals important tensions, such as those between conceptual richness and argumentative transparency, critique and applicability, and methodological coherence and epistemic openness.

This synthesis does not seek to rank or reduce the reviews, but rather to map how evidence claims are made meaningful, what assumptions they make about their audience and how they balance authority with interpretive humility. In doing so, it provides a reflective perspective on the reasoning norms that influence the practice of qualitative review writing in education, and on the potential of frameworks such as those of Toulmin and Gough to promote greater transparency, coherence and evaluability in the development of claims.

In the following sections, each analytical category—topic, importance, method, framework, synthesis type, claim framing, warrant, qualifier, and rebuttal—is addressed in detail, drawing on all six reviews to explore how these elements manifest, overlap, or remain underdeveloped. This structure allows for a thorough comparison of the reasoning strategies employed in qualitative systematic reviews, emphasising their strengths and limitations in terms of evidentiary contributions.

4.8.1 Topic

Across all six reviews, the topic is presented with clarity and specificity, typically in the title and introduction. Each review articulates a focused object of inquiry grounded in a clearly defined educational context. For instance, Michael et al. (2019) examine how movement is integrated into elementary classrooms, focusing on both instructional and systemic enablers and barriers. Savin-Baden et al. (2008) explore the underlying influences on teaching and learning in higher education, centring on themes like disjunction, transfer, and community. Harrison et al. (2019) examine how curricular design in Australia shapes Indigenous students' learning experiences, while Kim and Graham (2022) focus on CLIL teachers' professional development needs, structured around a competence-based model. Sheng (2020) synthesises qualitative research on IPV education interventions in UK schools, and Swearingen (2019) investigates how nonnative-English-speaking teacher candidates develop professional identities in English-dominant TESOL contexts. These reviews demonstrate a strong alignment between topic and purpose. Each defines not only what is being studied, but also whom the synthesis is for—teachers, policymakers, researchers, or institutional leaders—thereby satisfying Gough's (2021) criterion that evidence claims must be evaluated in light of their relevance to a defined audience and use.

4.8.2 Importance

In addition to presenting focused topics, all six reviews establish the importance of their research question within broader educational, social, or policy contexts. For Michael et al.

(2019), the review is positioned at the intersection of public health and pedagogy, responding to national concerns about sedentary behaviour and physical activity in schools. Savin-Baden et al. (2008) argue that higher education research is conceptually fragmented, highlighting hidden influences that shape pedagogy but often go unnamed in institutional discourse. Harrison et al. (2019) foreground the marginalisation of Aboriginal and Torres Strait Islander perspectives in the Australian curriculum, framing the review as a response to ongoing colonial epistemic injustice. Kim and Graham (2022) highlight the gap between what CLIL teachers need and what current professional development models offer. Sheng (2020) notes the relative absence of UK-based research on IPV education, especially compared to the U.S. Swearingen (2019) highlights the epistemic marginalisation of NNES teacher candidates within TESOL. In each case, authors use the review not only to synthesise but to advocate—to make visible an unmet need, theoretical gap, or structural failure. This mirrors Gough’s (2021) emphasis on fitness for purpose: a claim’s strength depends on internal logic and its resonance with relevant social, professional, or institutional concerns.

4.8.3 Theoretical Framework

Most reviews integrate formal theoretical frameworks into their synthesis design and interpretation. Michael et al. (2019) apply the *Social Ecological Model (SEM)* (McLeroy, 1988) to categorise facilitators and barriers to movement integration at multiple systemic levels, from the individual to the policy domain. Savin-Baden et al. (2008) draw on Freirean pedagogy, constructivism, and meta-ethnographic traditions, while Harrison et al. (2019) employ decolonial theory, Indigenous epistemologies, and Giddens’ structuration theory to reconceptualise curriculum. Kim and Graham (2022) use Cañado’s (2018) seven-competency model to synthesise teacher development needs. Sheng (2020) incorporates feminist and relational policy analysis to critique gendered assumptions in IPV education, and Swearingen (2019) anchors her analysis in poststructural and sociocultural identity theory, particularly Norton’s (1997) work on identity as negotiated and performative.

In these cases, frameworks are not merely cited but serve as structuring devices that shape inclusion, analysis, and interpretation. They guide how data are synthesised and claims are framed, ensuring internal coherence and giving conceptual depth to the reviews’ conclusions.

4.8.4 Justification of Method

All six reviews provide a rationale for conducting a systematic review, though the nature and depth of justification vary. Michael et al. (2019) argue the need to synthesise fragmented evidence to better inform policy and teacher practice on movement integration. Kim and Graham (2022) follow a protocol based on Zawacki-Richter et al. (2020), applying explicit inclusion and coding criteria to analyse PD needs across studies. Sheng (2020) justifies a qualitative systematic review to address a specific policy context (UK education) and includes a focus on student voice and affective impact. More conceptually driven reviews - Savin-Baden et al. (2008), Harrison et al. (2019), and Swearingen (2019) - use their methodology as a way of handling interpretive complexity. Savin-Baden rejects traditional meta-analysis as too reductive, opting for interpretive meta-ethnography to accommodate variation in pedagogical theory. Harrison frames the review as a way to challenge dominant constructions of success in the curriculum. At the same time, Swearingen aligns her methodological approach with poststructuralist concerns about identity, voice, and discursive positioning in TESOL.

What unites these choices is a consistent attention to alignment between method and purpose. Reviews with conceptual or transformative aims select interpretive synthesis approaches suited to meaning-making, while reviews with policy-facing goals select methods that support clear recommendation building. This reflects the central idea in Gough (2021): what counts as a strong method depends on what the review is for.

4.8.5 Framework Justification

Justifying the choice of framework is another strength across the corpus. Michael et al. (2019) justify the use of SEM as a means of capturing the multilevel complexity of movement integration. Savin-Baden et al. (2008) explain that interpretive meta-ethnography enables them to retain contextual nuance and surface under-theorised pedagogy themes. Harrison et al. (2019) argue that decolonial and relational theories are necessary to critique how curricula marginalise Indigenous perspectives and epistemologies. Kim and Graham (2022) adopt Cañado's framework because it encompasses technical and relational competencies essential to CLIL instruction. Sheng (2020) defends her use of feminist policy analysis as particularly suited to the affective and gendered dimensions of IPV education. Swearingen (2019) justifies her framework as allowing attention to the discursive and

emotional labour of teacher identity formation. In all cases, theoretical frameworks are not arbitrarily selected but fit the review's intended conceptual and practical purposes. The coherence between framework, method, and claims strengthens the credibility of each review's synthesis, fulfilling the expectation that strong evidence claims be contextually appropriate, interpretively transparent, and theoretically anchored (Gough, 2021).

4.8.6 Claim Framing

The six reviews vary in how explicitly they frame their claims, that is, how they signal whether their conclusions are meant to be normative, practical, conceptual, or some combination thereof. In SR#1 (Michael et al., 2019) and SR#2 (Savin-Baden et al., 2008), claim framing is relatively clear. Michael et al. present their findings as directly applicable to school-level policy and professional development, recommending supports for movement integration based on a structured summary of barriers and facilitators. Savin-Baden et al. frame their work as a conceptual contribution, calling for greater attention to pedagogical stance, disjunction, and agency in higher education literature and institutional development. In contrast, claim framing becomes more implicit and diffuse in SR#3 through SR#6. Harrison et al. (SR#3) construct a powerful curriculum critique through decolonial and Indigenous lenses. However, they do not explicitly differentiate whether their conclusions are intended as theoretical reframing, pedagogical recommendation, or curriculum policy intervention. Similarly, Kim and Graham (SR#4) identify competence gaps in CLIL professional development but do not signal whether their recommendations are meant for teacher educators, policymakers, or researchers. Sheng (SR#5) critiques fragmented IPV education strategies and calls for reform, but the boundaries between empirical description and normative advocacy are not always distinct. In SR#6, Swearingen synthesises identity development challenges among NNES teacher candidates, with findings that oscillate between conceptual, emotional, and pedagogical claims without explicitly marking their function. This ambiguity is analytically important. As Gough (2021) notes, understanding the purpose of a claim, whether it aims to inform, persuade, or reform, is crucial to judging its appropriateness and strength. Even well-supported conclusions can become difficult to evaluate or apply without precise framing, particularly for external stakeholders.

4.8.7 Warrant

Warrants, the reasoning that connects evidence (grounds) to the claims made, are among the most consistently implicit components across all six reviews. Even in the more procedurally structured reviews such as SR#1 and SR#2, the connection between findings and conclusions is typically assumed rather than articulated. In Michael et al. (2019), for instance, the warrant linking identified barriers to policy recommendations is left unspoken; it is taken for granted that if barriers are frequently reported, they should be addressed. Savin-Baden et al. (2008) similarly infer systemic influence from thematic patterns across studies, without formal reasoning. In SR#3 to SR#6, the warranting logic becomes even more embedded in narrative and theory. Harrison et al. (2019) build their claims about curricular reform from repeated references to the invisibility of Indigenous knowledge systems. However, they do not specify why this pattern necessitates a particular type of reform. Kim and Graham (2022) present a mismatch between teacher-reported needs and available PD structures but do not explicitly explain why this requires structural revision rather than local adaptation. Sheng (2020) makes a case for national IPV education reform based on evidence of program inconsistency, but the inferential steps connecting evidence to recommendations are implicit. Swearingen (2019) grounds her conclusions in affective and discursive themes, suggesting the need for critical pedagogy in TESOL, but without spelling out the logical justification sequence. According to Toulmin's model, these gaps weaken the structural clarity of an argument. According to Gough (2021), they obscure the evaluability of the claim, making it harder for readers to judge whether the conclusions fit *their stated purpose*. While implicit warranting is common in interpretive synthesis, its unacknowledged presence challenges transparency and uptake.

4.8.8 Rebuttal

Rebuttals, acknowledgements of the limitations, exceptions, or counterpositions to the review's claims, are inconsistently handled across the corpus. In some reviews, such as SR#5 (Sheng, 2020) and SR#6 (Swearingen, 2019), rebuttals are explicitly addressed by discussing scope limitations, methodological challenges, or contested interpretations. Sheng notes the absence of long-term outcome data and warns against overinterpreting short-term attitudinal shifts in IPV education. Swearingen reflects on the limitations of the NNES/native-speaker binary and cautions against essentialising identity categories. In contrast, SR#1 and SR#2

acknowledge methodological variation or literature gaps but do not engage directly with alternative perspectives or potential objections to their claims. SR#3 and SR#4 are mostly silent on this dimension. Harrison et al. (SR#3) do not consider potential tensions between Indigenous knowledge systems and existing curricular mandates. In contrast, Kim and Graham (SR#4) do not discuss how constraints on teacher development might affect the feasibility of their recommendations. From Toulmin's perspective, the absence of rebuttals leaves claims open to challenge, particularly when claims are ambitious or contextually bound. From Gough's perspective, rebuttals help establish the boundaries of claim relevance, clarifying what a synthesis can and cannot support, for whom, and under what conditions. Their inclusion strengthens the credibility and applicability of the review, particularly for external readers seeking to use the evidence to inform decisions or policies.

Concluding remarks

This comparative synthesis highlights the varied and often uneven ways qualitative systematic reviews construct their evidence claims. Across the six reviews analysed, all clearly defined their topic and relevance to educational and social concerns. Most also offered a credible synthesis process and a well-justified methodological orientation. Stronger reviews exhibited precise alignment between topic, method, theoretical framework, and synthesis strategy, reinforcing Gough's (2021) argument that the credibility of an evidence claim depends not on abstract standards, but on its coherence and fitness for a specific purpose and audience. At the same time, the analysis revealed persistent gaps in how reviews handle the internal logic of their claims. Key components such as claim framing, warrant, and rebuttal were frequently implicit, particularly in reviews that were theoretically rich or politically situated (SR#3–SR#6). In these cases, the reasoning linking data to conclusions was often assumed rather than made explicit, and the intended function of claims (conceptual, practical, normative) was not always clearly framed. This implicitness may reflect the interpretive nature of qualitative synthesis, where meaning emerges through narrative and theory rather than formal logic. However, without greater transparency, such claims risk becoming opaque to readers, undermining their evaluability and uptake. Applying the *Integrated conceptual framework* adapted from Toulmin's model and Gough's framework in this chapter has surfaced both the rhetorical strengths and epistemic blind spots of the reviews, offering a lens through which to assess not only *what* claims are made, but *how* they are constructed, justified, and bounded. In this way, the chapter contributes not just a comparison of six reviews but also a reflection on the broader discursive and

methodological challenges of building credible evidence claims in qualitative educational research.

5 Discussion

This chapter reflects on the patterns, variations, and methodological implications that emerged from applying the adapted Gough–Toulmin framework to six systematic reviews of qualitative studies in education. It is structured around three core themes: (1) cross-cutting insights into the nature of evidence claims in qualitative synthesis; (2) the role of implicit reasoning in shaping claims; and (3) the contribution, application, and critical reflection on the adapted framework and codebook developed for this thesis.

5.1 The Nature of Evidence Claims in Qualitative Synthesis

To understand the nature of evidence claims in qualitative systematic reviews (SRs), it is necessary to look beyond surface-level assessments of review findings. In this study, evidence claims were not just seen as final conclusions, but as the result of complex processes of interpretation, synthesis, and justification. These processes were analysed using a special framework. The framework is operationalised through analytical categories that are tailored to qualitative synthesis. The aim was to interrogate what was claimed, how these claims were constructed, why they were constructed, and under what conditions these claims were constructed.

The evidence from the six reviews we analysed was found to be very varied in terms of its form and function. Traditional evidence hierarchies, where a claim is usually a generalizable statement supported by statistical probability, are different. Qualitative systematic reviews often present interpretive claims that emerge through layered reading, reflexive interpretation, and thematic synthesis. The claims that are made are influenced by the epistemological orientation, the review's purpose, the synthesis method, and the contextual framing of the primary studies.

According to (Dixon-Woods et al., 2006), it is important to distinguish between the logics of aggregative and interpretive synthesis. The goal of an aggregative synthesis is to summarise data into categories that have been set in advance. These categories are often similar to those used in thematic or framework syntheses. In these cases, the evidence claim tends to focus on patterns of recurrence across studies, what works, for whom, and under

what circumstances. On the other hand, interpretive syntheses (like meta-ethnography or critical interpretive synthesis) focus on coming up with new ideas, with the goal of creating new insights or theoretical explanations that are based on the data. In this context, claims made are considered constructive, suggesting a shift in interpretation rather than mere accumulation.

This distinction is not purely technical; it has deep implications for how claims are made and justified. In interpretive reviews, the primary function of the claim is not to predict or prescribe, but to illuminate and make sense of complex phenomena. (Borgnakke, 2017) points out that this creates a tension within the evidence movement. This is because demands for standardised reporting and quality criteria often conflict with the iterative and emergent nature of qualitative inquiry. This is especially important in education, where research aims often include understanding meaning, context, and lived experience. These are domains that cannot be reduced to separate variables or outcomes.

The conceptual adapted framework applied in this thesis addressed these tensions by attending not only to the structure of the argument (Toulmin) but also to the purposive and contextual dimensions of claims (Gough). Categories such as claim framing, warrant, and relevance to users helped identify whether and how systematic reviews justified the claims they made, and whether these claims aligned with the synthesis method employed. In some cases (e.g. SR#1), there was high congruence between the stated review aim, the synthesis approach and the type of claim. In other cases (e.g., SR#4 and SR#5), there was a misalignment: although interpretive methods were used, the final claims resembled aggregative summaries, with little reflection on the logic of synthesis or the implications of conceptual abstraction.

This misalignment is symptomatic of a broader challenge in qualitative systematic reviews, namely the lack of epistemological reflexivity. As (Carroll & Booth, 2015) argue, the quality assessment of qualitative evidence synthesis must be informed by the synthesis type and the nature of the claim. It is not possible to standardise evaluative criteria across all review types because interpretive and aggregative logics require different forms of justification. For example, while aggregative claims may be evaluated in terms of thematic saturation or recurrence, interpretive claims demand coherence, explanatory power and theoretical insight. This requires criteria that are epistemologically attuned to the type of knowledge being produced, a principle that is core to the conceptual adapted framework developed in this study.

The findings align with recent calls for education research to reconceptualise evidence. This evidence should be understood as situated, value-laden, and oriented toward use. The value of educational research, as argued by (Pigott et al., 2022), lies not only in its methodological rigour but also in its ability to address critical issues of equity, justice, and context. In this light, evidence claims must be judged not just by their internal coherence, but also by their relevance to the lived realities of students, teachers and communities. Ethical and practical considerations are thus embedded in the evaluative process through categories such as importance, review purpose and user relevance.

Qualitative synthesis claims are multi-dimensional in nature. A hierarchy of findings is not applicable here, nor can it be judged solely on technical grounds. Instead, evidence claims must be understood as situated arguments. These are crafted through interpretive labour and framed by theoretical and normative commitments. They are aimed at specific audiences and uses. As the importance of qualitative synthesis grows, especially in complicated fields like education, this adapted framework can support more transparent, reflective, and contextually grounded approaches to reviewing and using evidence.

5.2 Implicit Argumentation: Patterns and Consequences

One of the most significant insights from the cross-case analysis of the six systematic reviews is the pervasiveness and analytical significance of implicit argumentation, particularly concerning the warranting logic, claim framing and rebuttal structures. As has been previously mentioned, while these elements are pivotal to Toulmin's (1958) model of practical reasoning and Gough's (2021) "Fitness for Purpose" framework, their analytical application was operationalised in this study through the adapted conceptual framework and structured codebook. This framework was not just a tool for retrospective analysis. It offered a generative lens. This lens was used to surface and interpret the reasoning processes embedded in qualitative systematic reviews.

The components of this adapted framework, such as claim framing, grounds, warrant, backing, rebuttal and qualifier, were distinguished and coded systematically. However, despite this structured approach, the coding process frequently revealed the absence or implicitness of key argument elements, especially the warrant, which links the data (grounds) to the conclusion (claim). In many reviews, this link was assumed rather than

articulated, which is an issue because the reviewer is not being clear or precise. Likewise, rebuttals were often underdeveloped or missing, which play a crucial role in demonstrating argumentative integrity and reflexivity.

This observation aligns with Toulmin's claim that strong arguments should show what is claimed and how and under what conditions the claim is valid. However, it also supports more recent theoretical developments. For instance, Karbach (1987) described Toulmin's model as a "heuristic" for revealing the silent architecture of argument, namely those steps and assumptions that support conclusions but are rarely made explicit. This confirmed Karbach's insight that even in peer-reviewed SRs, much of the reasoning occurs beneath the surface of the text. This reasoning is concealed within narrative synthesis or thematic summaries.

Furthermore, (Erduran et al., 2004) emphasise in their Toulmin Argument Pattern (TAP) how differentiating between components such as data, warrants, and backings can enhance evaluative judgement, particularly in disciplines that depend on interpretive, as opposed to statistical, inference. The challenges encountered in this study are strongly reflected in their argument: while most reviews presented rich qualitative findings (grounds), the principles justifying the movement from findings to claims (warrants) remained largely unexamined.

This issue was addressed by using a framework that incorporated Toulmin's structural categories and Gough's evaluative lens, as well as insights from (Khambete, 2019) adaptation of Toulmin for design research.

As established in Chapter 2.2, Toulmin's model is applied in the operational form proposed by Khambete (2019) to capture the distributed and frequently implicit reasoning of review-level claims. This adaptation specifically made use of the distinction between micro-claims and macro-claims. This distinction was especially useful when analysing SRs, which frequently present a main claim (e.g. about educational policy or practice) supported by several smaller claims derived from theme-level findings. However, the links between these micro-claims and the overarching macro-claim were often narrative rather than logical, with no explicit synthesis rule or warrant guiding the aggregation.

The concept of Reasoning Mode (e.g., inductive, abductive, analogical) is introduced by Khambete (2019) as a core component in applied reasoning. This is especially the case in educational synthesis, where inferences are often based on "best fit" interpretations rather than general rules. However, none of the reviewed SRs provided a clear explanation of the

reasoning behind linking evidence and claims. The absence of this reasoning mode, what Khambete calls a hidden operator, limits both the transparency and the transferability of the conclusions drawn.

This recurring implication in reasoning is not just a minor stylistic issue; it is a structural challenge in qualitative evidence synthesis. Readers (and end-users such as practitioners or policymakers) are left to infer the credibility and scope of the claims when warrants and rebuttals remain unstated, often without sufficient cues. This means that it becomes more difficult to evaluate the "fitness for purpose" of these claims, as Gough would define it. This, in turn, undermines their epistemic and practical utility.

The framework developed in this study offers a means of addressing this gap. Providing structured categories for identifying the components of an evidence claim enables a more reflexive reading of review conclusions. It does not presume that all reasoning must be made clear in every situation; instead, it encourages reviewers to make considered decisions about what is stated, justified, and qualified, especially when claims are meant to encourage action or inform policy.

The application of the adapted Toulmin–Gough framework revealed that implicit argumentation is both a defining feature and a vulnerability of many qualitative systematic reviews in education. This is not just because authors neglect logic, but also because the conventions of qualitative synthesis have not historically demanded argumentative transparency. The study offers a practical tool for identifying, interpreting and evaluating these hidden structures, contributing to a broader movement towards rigour without reductionism in qualitative research synthesis.

5.3 Reflections on the Adapted Framework and Codebook

The adapted conceptual framework was developed and applied to enable a close reading of each review's argumentative scaffolding, helping to surface what Ravitch & Riggan (2017) describe as the "conceptual logic" that connects empirical findings to larger knowledge claims. The adapted tool reconfigures Toulmin's *Model of argumentation* and Gough's "Fitness for Purpose" framework into an operational codebook. This makes the implicit reasoning within qualitative systematic reviews visible. It considers not only what is said, but also how it is justified, to whom, and for what purpose.

This approach is in line with what Sandelowski and Barroso (2007) call the "craftsmanship" of qualitative synthesis. This methodological sensibility is attentive not just to methodological transparency, but also to interpretive integrity and conceptual alignment. They believe that qualitative synthesis is more about understanding what is absent, unsaid, or subtly implied, rather than about rigid replication. The adapted framework responds to this call by incorporating categories such as claim framing, warrant, and rebuttal, even when these are not explicitly stated in the texts under analysis. Applying the framework proved demanding, particularly in identifying and coding implicit argumentative elements. Yet this difficulty underscores the need for such an analytical tool. Many qualitative reviews operate within interpretive or configurative logics, where conclusions are not directly drawn from data but emerge through acts of conceptual abstraction, contextual judgment, and synthesis. In this sense, the framework did not only function as a coding device but as a lens for interpretive accountability: it helped to detect where argumentation was assumed rather than shown, and where coherence between findings and claims depended more on narrative flow than on articulated logic.

The value of this approach is twofold. First, it supports researchers and reviewers in clarifying their reasoning and in articulating the epistemological and methodological assumptions underpinning their claims. Second, it provides editors, peer reviewers, and users of systematic reviews with a scaffold for evaluating the transparency, coherence, and utility of evidence claims. In doing so, it serves what Petticrew and Roberts (2006) identified as the foundational goal of systematic reviews: to produce trustworthy syntheses through transparency in methods (p. 11).

Importantly, the framework developed in this thesis extends beyond procedural guidance, such as that offered by Booth et al. (2016), by emphasising rhetorical structure and epistemic responsibility rather than just procedural validity. While procedural tools are essential for managing the review process (e.g., search strategy, inclusion/exclusion criteria), they do not fully address how knowledge is constructed through language, interpretation, and argument. The adapted framework addresses this gap by shifting the evaluative lens from what is included to how claims are constructed and justified, and whether they are fit for their intended interpretive, theoretical, or practical purpose (Gough, 2021).

This perspective aligns with that of Wilson & Anagnostopoulos (2022), who contend that qualitative reviews should strike a balance between rigour and reflexivity. They advocate for embracing interpretive principles while providing systematic accounts of how evidence

is evaluated, compiled, and interpreted. Their 'five core tasks' - clarifying purpose, defining quality, situating context, synthesising insights and practising reflexivity - closely align with the categories operationalised in this study's codebook. Like Wilson and Anagnostopoulos, this thesis favours a broader notion of conceptual coherence, where the logic of inclusion is subordinate to the logic of meaning-making.

The adapted framework contributes to this evolving vision by operationalising analytical structure and interpretive reflexivity. This enhances the methodological credibility and conceptual transparency of qualitative synthesis. It supports an approach to evidence that considers not only reliable results, but also reasoned judgements, claims that can be evaluated, contested and adapted by others within the educational research community.

5.4 Synthesis Insights Across Reviews

This subsection summarises the comparative insights that emerged from the cross-case analysis of the six systematic reviews examined in this study. Using the integrated framework (based on the approaches of Toulmin and Gough) allowed us to interpret evidence claims in a layered way, showing how they are constructed, justified and limited across different types of review. Rather than treating each review as an individual object of analysis, this synthesis identifies recurring patterns, conceptual tensions and epistemological blind spots that influence the formation of qualitative evidence claims in educational research.

The first key insight concerns the heterogeneity of synthesis logics and their implications for claim formation. Reviews such as those by Michael et al. (2019) and Kim and Graham (2022) adopted relatively structured methods (e.g. thematic synthesis or framework synthesis), framing claims in pragmatic terms aligned with practical recommendations. In contrast, reviews such as those by Savin-Baden et al. (2008) and Harrison et al. (2019) relied on interpretive meta-ethnography, embedding claims within broader conceptual or political narratives. This reflects the distinction between aggregative and configurative logics proposed by Dixon-Woods et al. (2006), and highlights that the nature of a claim — its ambition, form, and evaluability — must be considered in relation to the synthesis method employed.

Secondly, across all six reviews, warrants were implicit rather than explicit. Even in procedurally transparent reviews, the inferential step linking evidence to claims was usually taken for granted rather than explained. For instance, SR#1 and SR#2 present findings

supported by thematic patterns; however, the reasoning that justifies how these patterns lead to specific policy or conceptual recommendations remains unstated. In SR#4 and SR#5, warrants are embedded in theoretical discourse, but are not analytically disentangled. This reinforces Erduran et al.'s (2004) view that warranting logic is often the least visible but most critical component of educational arguments and validates the need for tools, such as the one developed in this thesis, that make such logic examinable.

A third finding relates to claim framing. Only two reviews (SR#1 and SR#2) clearly indicate whether their conclusions are conceptual, normative or practical. In contrast, SR#3 to SR#6 often oscillate between modes without explicitly positioning the function or audience of their claims. While this ambiguity is partly intrinsic to interpretive synthesis, it complicates the task of evaluating a claim's appropriateness for policy or practitioner use (Gough, 2021). Misalignments between purpose, synthesis type and claim structure, as observed in SR#5 (which uses interpretive methods but makes prescriptive policy recommendations), illustrate the need for greater epistemological reflexivity (Carroll & Booth, 2015).

Fourthly, rebuttals, which in Toulmin's model function as constraints on the generalisability of claims, were often underdeveloped. Where present, as in Swearingen (2019) and Sheng (2020), they served to acknowledge conceptual complexity or contextual limitations. However, in most cases, alternative interpretations or potential objections remained unexplored. This weakens the dialogic potential of the reviews and limits their evaluability by external readers, particularly in contexts where evidence must inform contested decisions (Petticrew & Roberts, 2006).

Finally, the analysis revealed a tension between conceptual richness and argumentative transparency. While reviews based on critical or poststructural frameworks (e.g. SR#3 and SR#6) provide valuable insights into power, identity, and social justice, their claims are frequently conveyed in interpretive prose that hinders systematic analysis. Rather than simplifying these contributions, the adapted framework provides a way to map their argumentative contours, clarifying how claims are constructed without flattening their complexity. As such, it supports what Sandelowski and Barroso (2007) refer to as 'methodological craftsmanship': an approach that values the integrity of interpretation and the clarity of justification equally.

Taken together, these findings suggest that, although qualitative reviews in education are methodologically rich and conceptually ambitious, they often lack argumentative clarity.

The framework developed in this study addresses this issue by supporting both analytical critique and future review practice.

Concluding remarks

This chapter has examined the construction of evidence claims in six qualitative systematic reviews through the lens of the adapted Gough–Toulmin framework. The analysis revealed that, while the reviews often displayed thematic coherence and methodological diligence, their argumentative structures frequently relied on implicit reasoning, especially with respect to warrants, claim framing, and rebuttals. These elements were rarely made visible, leaving readers to infer the logic connecting findings to conclusions. This confirms the recurrent concern in the literature that systematic reviews, particularly in education, may adopt a formal structure without systematically developing or justifying their claims (Roor et al., 2023; Borgnakke, 2017).

By operationalising an adapted framework, this study provides a practical tool for identifying both explicit and latent components of qualitative synthesis. It contributes not only to the diagnostic reading of published reviews, but also to the reflexive practice of conducting them, helping reviewers to articulate claims more transparently, justify inferential steps, and situate conclusions within their intended purpose and audience.

The chapter also highlighted the need to align synthesis type with the nature of the evidence claim, whether aggregative or interpretive, and to account for the epistemological complexity of qualitative research. The adapted framework responds to recent methodological calls (e.g., Dixon-Woods et al., 2006; Wilson & Anagnostopoulos, 2021) for more nuanced, rhetorical, and reflexive approaches to synthesis. It promotes a vision of review work not as a mechanical procedure, but as a form of argumentative craftsmanship (Sandelowski & Barroso, 2007).

In this way, the framework fosters rigour without reductionism, enabling systematic reviews to remain systematic without sacrificing the interpretive depth, conceptual sensitivity, and contextual responsiveness that define the best of qualitative educational research.

6 Conclusion

6.1 Summary of the Research

This thesis set out to examine how evidence claims are formulated, justified, and interpreted in systematic reviews of qualitative studies in education. Drawing on Toulmin's *Model of argumentation* (1958) and Gough's "Fitness for Purpose" framework (2021), an adapted integrated framework was developed to assess the internal reasoning of systematic reviews and the contextual relevance of their claims. The motivation for this work stemmed from the growing demand for systematic reviews to inform educational policy and practice, and the parallel concern that qualitative syntheses, particularly those addressing the subjective experiences of students and teachers, often rely on implicit warrants and loosely articulated purposes.

To investigate this, a systematic review of systematic reviews was conducted, applying a purposive sampling strategy to select six systematic reviews that met defined criteria regarding topic, methodological transparency, and epistemological orientation. These systematic reviews were then analysed using a bespoke coding framework grounded in Toulmin's argument elements (claim, warrant, grounds, backing, rebuttal, qualifier) and Gough's dimensions of review purpose, coherence, and utility. The analysis proceeded iteratively, with close reading of the systematic reviews texts, coding of argument structures, and reflective synthesis.

6.2 Key Findings

The findings reveal substantial variability in how systematic reviews construct and justify their evidence claims. Across the six cases analysed, several patterns emerged:

- **Implicit Reasoning:** In four of the six systematic reviews, the claim framing and warrant were implicit, requiring considerable interpretive work to reconstruct the internal logic of the conclusions. While the grounds (i.e., synthesised findings) were often well documented, the link between those findings and the claim, what Toulmin would call the warrant, was frequently assumed rather than articulated.

- **Review Purpose Alignment:** systematic reviews that clearly stated an interpretive or theory-building purpose showed greater congruence between synthesis method and evidentiary logic. In these reviews, Gough’s “fitness for purpose” categories were more robustly addressed, with greater reflection on epistemological fit and methodological choices.
- **Configurative vs. Aggregative Logic:** As noted by Newman & Gough (2020), qualitative systematic reviews typically follow a configurative synthesis logic, aimed at conceptual exploration rather than testing causal relationships. This configurative orientation was evident in all reviews analysed, but not always acknowledged or aligned with the stated goals. In some cases, there was a slippage toward aggregative framings (e.g., claims of generalizability or policy applicability) that were not fully justified by the methodological base.
- **Rebuttals and Limitations:** Most systematic reviews paid limited attention to alternative explanations, contradictions, or tensions within the evidence base. Rebuttals were either absent or framed narrowly in terms of limitations of primary studies, rather than as challenges to the synthesis itself.
- **Claim Strength and Qualification:** Only two systematic reviews made systematic use of qualifiers to nuance the strength or scope of their claims. Others presented unqualified conclusions despite heterogeneous and interpretively complex source material.

6.3 Contribution to Knowledge

This study makes three key contributions:

- **Conceptual Contribution:** The adapted integrated framework offers a novel lens for analysing how evidence claims are structured, justified, and presented in qualitative systematic reviews. This framework moves beyond checklist-based appraisals and instead supports a relational and purposive reading of reviews.
- **Methodological Contribution:** Through operationalising the framework into a detailed codebook and applying it to a purposive sample, this study demonstrates a transferable process for reviewers and educators seeking to interrogate reasoning within qualitative

syntheses. It also offers a structured yet flexible tool to guide reflexive reviewing practices.

Empirical Contribution: The analysis of six published systematic reviews in the field of education provides an evidence base about current practices of evidence claim construction. These findings shed light on the often-unexamined assumptions that shape qualitative review outputs and how these assumptions influence their perceived credibility, relevance, and use.

6.4 Reflections on the Framework

The process of developing and applying the integrated framework highlighted both its strengths and challenges. On one hand, the framework enabled a detailed reconstruction of the evidentiary reasoning in each systematic review and foregrounded the often-hidden logic of how claims are built from qualitative syntheses. On the other hand, it also required significant interpretive judgment, particularly when dealing with implicit warrants or tacit purposes. As such, the framework is not a neutral tool but a reflective lens, aligning with Ravitch and Riggan's (2021) view of conceptual frameworks as evolving, researcher-dependent constructs.

Moreover, the combination of Toulmin and Gough proved productive precisely because it merged structural (argument-based) and contextual (purpose-driven) approaches. While Toulmin offered analytical granularity, Gough ensured alignment with broader epistemological and pragmatic dimensions. In practice, this synergy encouraged a more robust assessment of not just how claims are made, but why, and for whom.

Implications for researchers

- Encourages reviewers to articulate claims, warrants, and qualifications explicitly.
- Promotes alignment between synthesis method, review purpose, and claim strength.
- Suggests that transparency in reasoning is as important as methodological rigour.

For Reviewers and Editors

- Supports the development of review guidelines and quality appraisal tools that foreground argument structure and epistemological coherence.

- Encourages the peer review process to focus on how claims are justified, not just what is claimed.

For Educators and Doctoral Students

- Offers a pedagogical framework for teaching evidence reasoning in qualitative synthesis.

Provides a practical tool for training early-career researchers to read systematic reviews critically and reflexively.

6.5 Limitations

As with any interpretive endeavour, this study has its own boundaries and limitations, some of which are methodological and some conceptual. Rather than flaws, these are reminders of the context in which this work was carried out and areas where further reflection and development are possible and necessary.

Firstly, the scope of the analysis was deliberately narrow. The six systematic reviews selected for analysis were chosen to allow for close and detailed reading within a thematically coherent field, namely the subjective experiences of students and teachers, rather than to represent the full diversity of qualitative synthesis in education. While this approach enabled depth, it also meant that other domains, topics and traditions were excluded. Applying the framework to a broader or more varied corpus - across disciplines, methodologies, or review cultures - would likely reveal new challenges or push the framework into territories requiring adaptation.

Secondly, identifying claims, warrants and argumentative moves, especially when they are implied, inevitably involves a degree of interpretive subjectivity. Despite having a structured codebook, decisions about what constitutes a warrant, how a claim is framed or where a rebuttal might be hidden in the text always depend on the reader's perspective. This is both the strength and the complexity of working with qualitative data: it invites critical engagement, but it also resists closure.

Finally, the framework is rooted in educational research traditions, particularly those that prioritise conceptual interpretation and practical significance. While I believe the core principles (reasoning, coherence and fitness for purpose) are transferable, other disciplines

may have different assumptions about what constitutes a claim, what counts as evidence or how synthesis should function. Bringing this tool into new contexts would therefore require careful recalibration.

In short, this work offers one framework applied to one set of reviews within one disciplinary context. However, it is precisely from these limitations that the potential for growth, adaptation and further dialogue emerges.

6.6 Directions for Future Research

The integrated framework developed in this thesis offers a promising basis for further research into qualitative synthesis methodology, particularly in the field of educational research where the construction of meaning is influenced by contextual, theoretical and frequently disputed assumptions. Future studies could expand the framework's empirical scope, apply it in new contexts and explore its potential to enhance analytical rigour and reflexive reasoning in the review process.

Firstly, while the current study focused on a purposive sample of six systematic reviews, future applications could test the framework across a larger and more diverse body of work, including reviews from related fields such as social work, public health and higher education. Such cross-disciplinary testing would refine the framework's adaptability and identify category boundaries that may differ across epistemological traditions. This could also inform ongoing efforts to develop quality appraisal tools sensitive to the argumentative structure of claims rather than just procedural features (Carroll & Booth, 2015).

Secondly, the framework could play a formative role in reviewer training and reflexive synthesis practice. As (Wilson & Anagnostopoulos, 2022) suggest, conducting a qualitative review depends not only on methodological choices, but also on the ability to navigate rhetorical, ethical and epistemological tensions. The integrated framework offers a vocabulary and logic that can support early-career researchers in understanding how to synthesise and reason through qualitative studies, foregrounding decisions around claim framing, warranting and justification. It could therefore be applied in doctoral programmes or in reviewer guidelines for journals and editorial boards.

Thirdly, this tool could be useful for both retrospective analysis of completed reviews and real-time use during the construction of reviews. Authors could apply the framework

iteratively while designing their synthesis, asking: What is the main claim I am making? What grounds support it? What assumptions or warrants connect them? Who is this claim for, and how does it fit the purpose? Encouraging such reflection during the writing process could reduce the gap between synthesis and claim and make the process more transparent for knowledge users.

Fourth, and increasingly relevant, future research could examine how the framework engages with emerging creative and metaphor-based forms of qualitative synthesis, in which the mode of representation is not only a way of communicating results but also part of the interpretive work. In poetic meta-ethnography, for example, translation across studies may be crystallised into poetic forms that condense meaning, convey affect, and foreground ethical attention to voice and experience; similarly, metaphor-driven analytical strategies treat figurative language as an analytic resource for generating insight, revealing how experiences are structured and understood rather than merely adding stylistic colour (Toye et al., 2014; Larsen et al., 2025). Such approaches pose specific challenges for evaluative reading because the reasoning that links synthesised material to an overarching claim may not be expressed through explicit propositional statements. Instead, the warrant can be carried through patterns of selection and arrangement (what is included or omitted), the internal coherence of an image or metaphor across cases, and the interpretive force created by juxtaposing fragments from different studies. Future research could therefore explore how Toulmin's categories might be used to reconstruct the inferential architecture of these syntheses by asking: what functions as grounds (e.g., translated metaphors, recurring images, distilled lines), what implicit warrant authorises the move from these grounds to a generalised insight (e.g., that metaphorical convergence signals a shared experiential structure), and how qualifiers and rebuttals are communicated (often through ambiguity, conditionality, or explicit boundarying of context). Gough's categories could then support judgements about fitness for purpose by clarifying the intended audience and use of the synthesis (for example, whether the aim is explanatory understanding, ethical witnessing, or practitioner reflection), so that such work is assessed in relation to its epistemic and ethical goals rather than against inappropriate expectations of standardised reporting.

Finally, further research could examine how users of reviews, such as educators, policymakers and institutional stakeholders, perceive and evaluate the strength of arguments in qualitative syntheses. Do more explicitly reasoned claims enhance perceived credibility or usefulness? How do implicit versus explicit argumentative strategies influence uptake? Investigating these questions would shed light on the practical consequences of reasoning

visibility and could lead to evidence-based recommendations for improving review design and reporting standards.

In fields such as education, where research often intersects with lived experience, cultural values and contested goals, evidence cannot be separated from interpretation. For this reason, there is an urgent need for tools that foreground reasoning as a visible, reflexive and purposeful practice. This thesis presents the integrated framework as a conceptual and ethical contribution to the future of qualitative synthesis in education, as well as a technical aid.

6.7 Final Thoughts

This thesis has been both a research project and a learning journey. At its core was a simple yet persistent question: how can we tell if a claim made in a qualitative systematic review is convincing? As I worked through six very different reviews, each with its own logic, tone and aims, I realised that the challenge was not just about quality assessment or methodology. It was also about understanding how authors reason with evidence, how they build their arguments and how much of their reasoning remains invisible.

To explore this, I developed an integrated framework that combined Toulmin's model of argumentation with Gough's concept of 'fitness for purpose'. The result was not a checklist or a formula, but rather a tool that slows down the process and makes the logic connecting findings to claims visible. Through this process, it became clear that evidence claims in education are never merely technical statements; they are positioned, purposeful and often deeply interpretive. Yet the reasoning behind them is often implicit, particularly in qualitative synthesis.

Applying the framework was not always easy. At times, it felt like reading between the lines of the reviews and searching for connections that were only hinted at. However, this was also the most intellectually rewarding part of the work. It enabled me to view reviews as more than just containers of knowledge; they are also acts of reasoning, crafted by authors for specific audiences and shaped by their choices, assumptions and values.

I don't see this framework as a definitive solution. Rather, I hope it can provide a foundation for researchers, reviewers and students to consider the structure and strength of their claims more carefully. In a field such as education, where research often informs practice, policy

and lived experience, the way in which we justify our claims is important. We owe it to our readers - and to ourselves - to be as clear, honest and reflective as possible.

This thesis is one attempt to support that effort. I hope it will encourage others to read reviews more critically, to write them more consciously and to engage with synthesis not just as a method, but as a form of thoughtful argument.

Appendix A - The Protocol

The evidence claims in the Systematic Reviews of Qualitative Studies in education: a protocol.

Background

The field of educational research is a dynamic and complex realm that defies easy definition and containment. It encompasses a vast array of disciplines, methodologies, and perspectives, making it challenging to establish clear boundaries or limits.

Assuming that this study will develop within the frame of a systematic review, the aim is not to repeat the searches and assessment of study eligibility, but rather to investigate the knowledge claims made by qualitative systematic reviews and how the researchers who conduct these reviews construct the ‘support or warrant’ for their claim.

Starting from the assumption that an appraisal needs to be done of whether the claims are justified and whether they are useful to the decisions being made, the aim is to develop a comprehensive theoretical framework of core issues relevant to assist the appraisal of the “fitness for purpose” of evidence claims by integrating Toulmin's Argumentation Model and Gough's Claim Appraisal Framework.

In consideration of the extensiveness of the area of interest, that is qualitative systematic reviews in education, this review will be preceded by an *a priori* protocol including detailed inclusion and exclusion criteria, a structured search process to locate and select relevant existing reviews and a formal process of data extraction (Cooper et al., 2022).

For the purpose of this study, only systematic reviews focused on qualitative studies involving primary, secondary, or postsecondary students’ and teachers’ subjective experiences will be considered.

The PCC framework will be used to construct clear and meaningful objectives and eligibility criteria (Pollock et al., 2023).

Population: primary, secondary, or postsecondary students and teachers.

Concept: students’ and teachers subjective experiences.

Context: school environment, from primary to post-secondary education within the formal educational system where students engage in structured learning activities.

Since systematic reviews have emerged as a prominent and widely embraced methodology in the field of healthcare and health education, there is a substantial overlap between studies conducted within the realms of health education and education research more broadly. For this purpose, the systematic reviews under consideration need not focus on health education unless they are specifically conducted within the school context.

Systematic review designates both a process and a product (Alexander, 2020). How is the method of qualitative systematic review adapted in the educational research field when subjective experience is investigated? Constructs like attitudes, beliefs, opinions, and feelings are studied mainly in psychology and are crucial for understanding human behaviour, decision-making, and social interactions. Psychological measurement is often referred to as psychometrics, and although many variables studied by psychologists are straightforward to measure (age, height, weight), others are more complex. These constructs include personality traits, emotional states, attitudes, and abilities. We are talking about some scores that somebody can assign to individuals in a way that communicates more or less quantitative information about the variable of interest (Jhangiani et al., 2019). This measurement is mainly done through surveys, questionnaires, and experimental studies.

Qualitative research, a multifaceted and intricate field, encompasses diverse approaches to comprehending the intricacies of human experiences, behaviours, and social phenomena (Patton, 2015). It involves deep investigation to understand the meaning of phenomena and their relationships (Gough et al., 2017).

This review aims to map the body of literature on systematic reviews of qualitative studies in education. I will consider only systematic reviews of qualitative studies involving primary and secondary students and teachers' subjective experiences investigated through qualitative methods and tools (interviews, focus groups, questionnaires).

Considering the extensiveness of the area of interest, this review will follow a protocol, including a structured search process to locate and select relevant existing reviews, detailed inclusion and exclusion criteria, and a formal data extraction process (Aromataris et al., 2015).

Defining Pedagogical Beliefs (Construct)

The construct of attitudes has been broadly defined as "not directly observable, inferred aspects, consisting of beliefs, feelings, and behavioural predispositions towards the object to which they are directed" (Auzmendi, 1992, p. 17, cited in Mondéjar-Jiménez & Vargas-Vargas, 2010). This understanding is consistent with the view of attitudes as latent psychosocial constructs that are inferred from patterns of evaluation and action rather than directly observed (Aiken, 2002). Attitudes encompass cognitive (beliefs, knowledge, expectations), affective (feelings, moods, emotions), and behavioural (actions, both intended and actual) components (Eagly, 1993). These components collectively shape how individuals evaluate and respond to various objects and events in their environment (McGuire, 1985. In Lindzey, G. and Aronson, E., Eds., *The Handbook of Social Psychology*).

Understanding teachers' attitudes is critical in educational research, as these attitudes significantly influence classroom practices. Borg (2017) explores this relationship in depth, emphasizing the strong connection between teachers' beliefs and their classroom practices. This interplay is vital in influencing the learning environment and impacting student outcomes.

The importance of attitudes extends beyond the classroom to the broader context of home-school partnerships. Christenson and Sheridan (2001) conceptualise optimal home-school partnerships through the "four A's": approach, attitude, atmosphere, and action. They emphasise that a positive attitude, characterised by a non-blaming atmosphere, understanding multiple perspectives, and focusing on student strengths, is essential for effective collaboration between families and schools. This holistic approach fosters a respectful and inclusive climate, which is crucial for the success of such partnerships.

Attitudes are part of a broader set of constructs, including values, opinions, and beliefs, which are fundamental in psychology and sociology for evaluating and interacting with the environment. These constructs are of interest to social and behavioural scientists and educators, administrators, and policymakers, as they significantly influence perception and behaviour, thereby shaping the environment individuals navigate and create.

Historically, the term "attitude" has evolved since its introduction in psychology in the 1860s by philosophers Herbert Spencer and Alexander Bain (Cacioppo, Petty, Losch, &

Crites, 1994). Despite variations in definitions among researchers, attitudes remain a central and indispensable concept in social psychology, as emphasised by Allport (1935). Gagne and Briggs (1974, p. 62) describe attitudes as an "internal state which affects an individual's choice of action toward some object, person, or event", reinforcing the idea that attitudes are inferred from behaviour rather than directly observed.

The Research question.

What are the characteristics of the knowledge claims made in qualitative systematic reviews in education, and how do the authors of the reviews support them?

In consideration of the extensiveness of the area of interest, namely qualitative systematic reviews in education, for this study, I will focus on systematic reviews that include only qualitative studies involving primary and secondary students, as well as teachers' subjective experiences.

The aim of this protocol is to conduct a systematic search to find the literature (systematic reviews) that I will then analyse for an in-depth reading of the eligible reviews to choose a purposive valuable sample for the successive development and application of the integrated framework that will help answering the research question.

I will proceed working through the following phases:

- A systematic search in ERIC database (Education Resources Information Center)
- Find eligibility criteria
- Screening in Rayyan
- Choose a purposive sample

Search strategy

The initial search strategy follows the PCC framework to construct clear and meaningful objectives and eligibility criteria (Pollock et al., 2023).

Population: primary and secondary students and teachers.

Concept: Students and teachers' subjective experiences (attitudes, experience, beliefs, opinions).

Context: school environment, from primary to secondary education, within the formal educational system where students engage in structured learning activities.

The studies included will be identified through an extensive, transparently documented search of the ERIC database (Ebsco platform). Given the conceptual and terminological heterogeneity of qualitative evidence synthesis in education, the search will be approached as a structured but iterative process, with search terms tested and refined to balance sensitivity (capturing relevant records) and specificity (excluding irrelevant material), in line with guidance on systematic searching for complex reviews (Levay & Craven, 2018).

The reason for searching ERIC is twofold. Firstly, ERIC is an authoritative indexed and full-text education literature and resources database of indexed and full-text education literature containing records and links to hundreds of thousands of full-text documents. It includes records from various sources, including journal articles, books, conference papers, curriculum guides, dissertations, and policy papers. Secondly, this systematic search strategy aims to retrieve much literature, such as systematic qualitative reviews in education. However, my aim is not to gather all the existing systematic reviews as it is not the aim of this search. The ultimate goal is to achieve a purposive sample whose characteristics will be explained in the paragraph on purposive sample.

The following search strings will be entered in the database ERIC,

S1= (DE Student Attitudes) OR (DE Student Experience) OR (DE Teacher Attitudes) OR (DE Student Teacher Attitudes) OR (DE Teaching Experience)

S2= ((DE Beliefs) OR (DE Opinions)) AND (DE teachers OR DE students)

S3= S1 OR S2

S4= (TI systematic review OR AB systematic review)

S5= S3 AND S4

Eligibility criteria

All the records will be exported into Rayyan. Rayyan is a web-based tool designed to streamline the process of conducting systematic reviews by enabling efficient collaboration, blind screening, and organising references, thereby enhancing productivity and reducing bias.

Inclusion criteria.

The included studies must be:

1. Systematic reviews (a review and synthesis of existing primary research studies with reported methods)
2. systematic reviews in which the primary studies included use qualitative methods (i.e. investigate the views/ perceptions/ experiences of participants using text/narrative/ speech as data)
3. systematic reviews which use a qualitative method of synthesis
4. systematic reviews in which the participants are students or teachers in primary or secondary school settings (from grades 1 to 13)
5. The research topic should be education or learning broadly conceived.

Exclusion criteria.

Reviews in which:

1. there are no precise methods and/ or primary research studies
2. the primary studies included the use of quantitative methods (data is in the form of numbers)
3. the method of synthesis is a statistical
4. the participants are not students or teachers in primary or secondary school settings (from grades 1 to 13)
5. research focuses on health (not to be confused with educational health programs in schools).

Screening

In Rayyan, titles, abstracts and, when necessary, the full text will be screened by me, and a random sample will be double-checked with my tutor and co-tutor for assessment against the inclusion criteria for the review. The search results and the study inclusion process will be reported in the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) flow diagram (Page et al., 2021).

The final eligible records will be included for coding in the data extraction table.

From these records I will find a purposive sample to work on for the in-depth analysis for the construction of the integrated conceptual framework derived from Toulmin's model (2008) and Gough's framework (2021).

Purposive sample

Qualitative research is a type of empirical investigation that generally involves: (a) selecting cases deliberately to gather information-rich insights; (b) conducting thorough, open-ended interviews, extended participant or field observations, and/or examining documents or artifacts; and (c) employing analysis and interpretation methods that delve deeper than the initial data and their superficial aspects. (Sandelowski & Barroso, 2006)

The primary benefit of using a purposeful sample in constructing a conceptual framework lies in its ability to enhance the relevance and specificity of the research.

The inherent selectivity of a non-probability sample stems from the researcher's intentional selection of a particular group, with the understanding that this group does not represent the broader population but only itself (Cohen, Louis et al., 2018).

To develop a framework adapted from integrating the *Fitness for Purpose of an Evidence Claim Framework* (Gough, 2021) and Toulmin's *Model of Argumentation* (1958), I will work on a purposeful sample of SRs chosen from the final records that will be included.

The effectiveness of purposeful sampling comes from choosing cases that are rich in information for in-depth study. These cases offer significant learning opportunities about key issues relevant to the research purpose, hence the name purposeful sampling. Focusing on information-rich cases provides deep insights and understanding, rather than broad empirical generalisations. (Patton, 2002).

Patton (2002) proposed 16 strategies for purposeful sampling in qualitative research, each designed to achieve a specific goal, but he does not mention using purposeful sampling for research syntheses. His writings focus solely on its application in primary research. In Suri (2011) article, she investigates how Patton's concept of purposeful sampling might be adapted for research synthesis by employing three techniques.

In primary qualitative research, the subjects are human participants. The focus is gathering in-depth data about individual experiences, behaviours, and perceptions. In research synthesis, the subjects are documents, such as research studies, reports, or other textual materials. The aim is to analyse and integrate findings from existing studies to create a comprehensive understanding of the existing body of knowledge on a particular topic.

In this research project, I can gain more targeted and in-depth insights by intentionally selecting reviews that address this study's particular topic of interest. This approach ensures that the analysis remains concentrated on the most pertinent data, enhancing the relevance and applicability of the findings.

For this reason, I will opt for a maximum variation sampling. In this type of sampling, the researcher targets participants, or in this case, systematic reviews, with the same characteristics but different unique attributes across a great deal of variation (Suri, 2011).

This approach enhances the quality of the data, as it incorporates characteristics from diverse contexts that are all relevant to the study. The main advantage of this technique is that it allows us to gain insights from multiple perspectives, providing a comprehensive understanding of the research topic. (Nyimbili & Nyimbili, 2024)

Coding

The codebook and coding procedure will be developed and implemented in a linked, iterative sequence connecting initial descriptive extraction, purposive sampling, and analytical coding. The coding process will follow established qualitative coding principles, moving from initial descriptive coding toward more analytical categorisation through iterative refinement of code definitions and memo-writing (Saldana, 2021). The codebook will be constructed in two phases. In the first phase, Toulmin's (1958) model of argumentation and Gough's (2021) Fitness for Purpose of an Evidence Claim Framework will be used independently as scaffolds to generate provisional categories capturing, respectively, the internal structure of reasoning (claim, grounds, warrant, backing, qualifier, rebuttal) and the contextual and purposive dimensions of claims (topic, importance, justification of method, theoretical framing and justification, audience and scope). In the second phase, these categories will be integrated through iterative framework development: overlaps will be aligned (e.g., qualifiers with strength/scope; rebuttals with consideration of alternatives), gaps will be resolved (including clearer distinctions between grounds, warrant, and backing), and an analytic sequence will be established (first applying Gough's categories to determine purpose, audience, and scope, then applying Toulmin's to reconstruct inferential structure). Operational definitions, inclusion/exclusion rules, and indicators will be specified for each category, with explicit attention to both explicit and implicit instances; pilot coding and memo-writing will inform minor refinements to ensure the final codebook is coherent, non-overlapping, and workable for comparative analysis. Methodological guidance from qualitative synthesis and coding literature will inform this process, particularly line-by-line

and descriptive coding as a foundation for interpretive work (Thomas & Harden, 2008), the integrative logic of qualitative synthesis (Sandelowski & Barroso, 2007; Finfgeld-Connett, 2014), and systematic procedures for operationalisation, analytic memos, and iterative refinement (Miles et al., 2014). In practice, an initial descriptive coding/extraction stage will be conducted across the full set of candidate reviews to document key review characteristics (e.g., topic, population, synthesis approach, and context), and these descriptive codes will inform the purposive sampling strategy.

Appendix B *Integrated conceptual framework* - Codebook

Category*	Definition	Coding Question	Search helpers		Interpretive Notes
			Location in text structure	Textual marker	
Topic	Focus and scope of the systematic review.	What is the topic of the SR? Is it clearly defined?	Title, objectives, opening paragraphs, explicit statements of focus.		Only code the primary focus, not all peripheral issues.
Importance	Significance or relevance of the topic.	Do authors mention why this topic matters?	Often found in intro or justification	Statements on societal, educational, policy importance. “Urgent need”, “critical issue”, “addressing a gap.”	.
Justification of Method	Rationale for using a systematic review.	Do authors justify the use of a systematic review?		“A systematic review was needed to...”, “synthesis across studies”, “no single study sufficient.”	Focus on reasons for using SR, not reasons for topic importance.
Theoretical Framework	Use of guiding theory or conceptual lens.	What framework or theory do authors say they use?		“Guided by X framework”, references to models or paradigms.	Only code what is explicitly identified as a framework.
Framework Justification	Rationale for the chosen framework.	Why did they choose this framework?		“This model is suitable because...”, “captures nuances of...”	Look for explicit reasoning. If none, note as absent.
Main Claim	Central conclusion or assertion from the synthesis.	What is the main conclusion?	Conclusions, results.	“This review shows that...”, “Findings indicate...”, “Overall, we conclude...”	The key “takeaway.” (the central conclusion)
Other Claims	Subordinate or related claims.	Are there additional conclusions or insights?	Secondary conclusions in results/discussion.		Code if they present multiple layered conclusions.
Claim Framing	How is the claim presented (theoretical, practical, normative)?	How do authors position the claim?		“Implications for practice/policy”, “contributes to theory.”	Captures the angle or orientation of the main claim.
Grounds	Data supporting the claim.	What evidence underpins the claim?	Summaries of included studies, quotations, thematic results.		Only the data or evidence base itself, not the reasoning.

Synthesis Type	Nature of synthesis approach.	What type of synthesis is used?		“Thematic synthesis”, “meta-ethnography”, “configurative review.”	Important for understanding how claims are built.
Validity/Relevance	Appraisal of evidence quality or appropriateness.	Do authors discuss validity or relevance?		“Limitations include..”, “most studies were robust”, “contextual constraints.”	Not about claim limits (rebuttals) but about data quality.
Warrant	Principle linking data to claim.	What reasoning connects evidence to conclusion?		“Because X, therefore Y”, “this suggests that...”, “the pattern indicates...”	May be unstated, inferred from flow of argument.
Backing	Theoretical or empirical support for the warrant.	What supports the underlying logic of the warrant?	References to prior theories, models, or corroborative studies.		Distinct from data (grounds); backing justifies the warrant.
Qualifier	Degree of certainty or strength of claim.	How strong or tentative is the claim?		“Probably”, “likely”, “possibly”, or bold statements “clearly”, “definitely”.	Signals level of certainty or generalizability.
Rebuttal	Acknowledged limitations or exceptions.	Do authors note alternative explanations or limits?		“However..”, “this may not apply where..”, “limitations include...”	Different from validity of data; here the focus is on claim’s reach.

*specify if explicit (E) or implicit (I)

Appendix C Codebook analysis

Categories	(Michael et al., 2019) - Facilitators and Barriers to Movement Integration in Elementary Classrooms: A Systematic Review (SR #1)	Explicit	Implicit
Topic	The title explicitly states the focus: “Facilitators and Barriers to Elementary Classroom Teachers’ Use of Movement Integration Products and Programs: A Systematic Review.” In the introduction, the aim is clearly framed: “The purpose of this study was to conduct a systematic review of the literature to identify facilitators and barriers to elementary classroom teachers’ (CTs’) use of MI in the classroom.” (p. 2) Additionally, they state the guiding question: “What factors enable or hinder elementary CTs’ use of MI implementation?”	E	
Importance	The authors emphasise the public health relevance of movement integration (MI) by stating that “Participation in regular physical activity benefits children and adolescents by reducing risk factors for diseases...” (p. 1). They also highlight that “most children are not meeting the recommended PA guidelines,” and position schools as critical spaces for PA: “schools provide an existing infrastructure for providing PA before, during, and after school.” (p. 1) These statements establish both the urgency and contextual relevance of the review.	E	
Justification of Method	The authors justify the need for a systematic review by noting that no prior efforts had synthesised the literature on movement integration in this context: “To date, no efforts have been made to systematically review the research on MI...” (p. 2). They further clarify that the review is intended to “support multiple stakeholders... by informing their PA promotion efforts based on the existing evidence base” (p. 2), justifying synthesis as a method for accumulating relevant, applicable knowledge.	E	
Theoretical Framework	Michael et al. (2019) explicitly apply the Social Ecological Model (SEM) to categorise findings on MI use: “A social-ecological perspective was used to categorise themes...” (p. 3). The SEM is used as a conceptual map to organize findings at different levels (institutional, intrapersonal, etc.).	E	
Framework Justification	SEM is justified as allowing consideration of multi-level influences on teacher behaviour and implementation: “This framework provides a meaningful way to conceptualise the different levels of factors that can influence behaviour...” (p. 3). The framework helps categorise facilitators and barriers at different levels (institutional, interpersonal, intrapersonal) to understand MI implementation comprehensively. [McLeroy et al. (1988) suggested five levels of factors that relate to health promotion interventions: (a) intrapersonal, (b) interpersonal, (c) institutional, (d) community, and (e) public policy... For the purposes of the present study, McLeroy et al.’s (1988) model provided a way to align the study’s findings with an ecological perspective.’ (p. 4)].	E	

<p>Main Claim</p>	<p>The review concludes that the most influential factors affecting teachers' implementation of movement integration (MI) fall within two primary domains: institutional and intrapersonal. This dual-level framework is grounded in the authors' identification of 12 thematic categories, which they explicitly organize using the Social Ecological Model: "A total of 12 themes of MI facilitators and barriers were identified and categorized into two social-ecological levels: institutional factors (e.g., administrative support, resources) and intrapersonal factors (e.g., teacher confidence, ease of implementation)" (p. 5). Among institutional barriers, the most frequently reported challenge was lack of time: "CTs' biggest institutional barrier to MI was not having enough time" (p. 9). Additional institutional-level impediments included limited physical space and insufficient material resources, while key enablers were administrative support and school-level prioritization of physical activity: "At the institutional level, administrative support and resources for MI emerged as key factors" (p. 9). On the intrapersonal level, teacher confidence, beliefs about MI's benefits, and the perceived ease of implementation played central roles. Summarizing these findings, the authors state: "Our findings show that facilitators and barriers to MI exist at two levels of influence: institutional and intrapersonal" (p. 9). Overall, the review highlights how both structural conditions and individual dispositions shape the success of MI initiatives in elementary classrooms.</p>	<p>E</p>	
<p>Other Claims</p>	<p>While the review focuses primarily on institutional and intrapersonal factors, the authors acknowledge a significant gap in the literature concerning interpersonal, community, and policy-level influences on movement integration (MI), noting that "few studies addressed factors beyond the institutional and intrapersonal levels" and identifying this as a direction for future research (p. 9). Among the intrapersonal facilitators, three key elements emerge: (a) teachers' perception that physical activity (PA) is valuable, (b) their beliefs about the benefits of MI, and (c) their confidence in implementing MI strategies. Confidence is consistently highlighted as one of the most impactful enablers: "Teacher confidence was one of the most significant facilitators to teachers' MI, and teachers' playing and coaching experiences in PA (e.g., playing sports, earning coaching credentials) were important factors underpinning their confidence in using MI" (Usher & Anderton, 2014, p. 7). Additional intrapersonal factors include the perceived ease of implementation, as teachers were more likely to adopt MI when they viewed it as simple to integrate into daily routines. On the institutional level, the most frequently cited barrier was lack of time: "CTs' biggest institutional barrier to MI was not having enough time" (p. 9). However, the authors argue that this barrier could be reframed through targeted professional development: "Efforts to increase teachers' use of MI should therefore focus on helping teachers learn to view MI as a noncompetitor to other school priorities and even take advantage of MI as a time-saving strategy" (p. 9). Moreover, access to professional development and administrative support emerged as critical facilitators. As the review explains: "Teachers would likely face fewer MI implementation challenges and feel more motivated to use MI if their administrators provided increased access to MI trainings that focused on using a wide range of MI strategies and the conditions needed for successful MI implementation" (p. 9). These findings underscore that both personal beliefs and structural support mechanisms must</p>	<p>E</p>	

	be addressed to enhance MI uptake in schools, while also pointing to important underexplored areas at broader ecological levels.		
Claim Framing	The authors frame their central claim as practical and policy-relevant, with a clear orientation toward informing professional development, intervention design, and school-level implementation strategies (pp. 2, 9). Rather than positioning the findings as theoretical contributions, the review consistently emphasizes their utility for shaping classroom practices and guiding decision-making at the institutional level. The findings are presented as actionable insights to improve school-based movement integration (MI) by addressing both structural and attitudinal barriers. For instance, the authors suggest that “sharing research on the academic benefits of MI (e.g., improved on-task behavior, increased standardized test scores) [...] may be an important step in convincing teachers that MI will reduce the time needed to gain students’ attention and establish/reinforce a classroom management system” (p. 9). This statement directly connects the review’s conclusions to practical teacher concerns, such as time management and classroom control. Similarly, they highlight that “conditions for success could include enacting appropriate classroom management protocols,” listing specific examples such as establishing ground rules and using calming activities to transition back to seatwork (p. 9). These examples reinforce the pragmatic orientation of the review’s recommendations and position the findings as immediately applicable within the realities of elementary school teaching. Overall, the claim is framed in a way that supports educator empowerment, institutional change, and evidence-informed policy decisions, making it a clear example of a practically framed evidence claim.	E	
Grounds	The grounds are clearly articulated through the systematic selection and thematic analysis of 28 peer-reviewed studies. These studies were retrieved from four major databases—Google Scholar, ERIC, PubMed, and PsycINFO—using a comprehensive search strategy: “A literature search was conducted to identify all published research, in English, that reported facilitators and/or barriers to using MI in elementary school settings” (p. 3). The included studies span a range of methodological designs and were subjected to content analysis, with themes identified by coding the results sections. The synthesis process is transparently reported, including a PRISMA flow diagram that documents the screening and inclusion criteria (p. 4). Thematically derived findings—such as institutional barriers (e.g., lack of time) and facilitators (e.g., teacher confidence)—are grounded in empirical evidence across these 28 studies. As such, the grounds for the review’s claims are both systematically retrieved and analytically processed.	E	

Synthesis Type	The synthesis is best understood as a configurative, thematic content analysis. The authors state that they applied content analysis to code and synthesize findings across studies: “Relevant data were extracted from each article and organized using an author-developed coding form... categorized using a social-ecological framework” (p. 3). This approach aligns with a configurative logic, as it aims not to summarize predefined outcomes but to construct thematic interpretations across diverse studies. While they do not use the term “thematic synthesis” explicitly, their approach shares core features of that methodology, particularly the inductive organization of findings into higher-order categories. These categories (e.g., institutional and intrapersonal barriers/facilitators) were not aggregated outcomes, but interpretive constructions aimed at building conceptual understanding of what shapes MI implementation.	E	
Validity/Relevance	Attention to validity and relevance is demonstrated through multiple procedural safeguards. First, the review adheres to PRISMA guidelines (p. 4), providing a clear and replicable account of search, screening, and selection processes. Second, the study describes a dual-coding process used during data extraction, enhancing inter-rater reliability and minimizing interpretive bias (p. 3). Third, the authors explicitly acknowledge the limitations of their synthesis, noting that their findings are restricted to institutional and intrapersonal levels of the Social Ecological Model (SEM), while interpersonal, community, and policy-level influences were underexplored due to gaps in the available literature (p. 9). This reflexivity about the scope and limitations of the evidence base strengthens the transparency and relevance of their claims.	E	
Warrant	The warrant is not explicitly stated but can be clearly inferred from the structure and intent of the argument: namely, that removing institutional barriers (e.g., lack of time, insufficient training) and strengthening intrapersonal facilitators (e.g., teacher confidence, positive beliefs) will increase teachers’ use of movement integration (MI) strategies. The review’s organization and discussion sections suggest a causal or at least a strongly conditional logic: If teachers are given administrative support, time-saving strategies, and adequate training, then they are more likely to adopt MI in the classroom. This interpretive assumption is consistent with the use of the Social Ecological Model (SEM), which embeds human behavior within layered systems of influence and suggests that changes at the institutional or intrapersonal level can trigger behavioral shifts.		I
Backing	The warrant, that institutional and intrapersonal conditions influence teachers’ use of movement integration (MI), is supported by explicit theoretical and empirical backing. The review is grounded in the Social Ecological Model (SEM), which is presented as a theoretical framework suited for capturing multi-level factors that influence teacher behavior (p. 3). The authors justify its use by explaining that SEM “provides a meaningful way to conceptualize the different levels of factors that can influence behavior” (p. 3), thereby giving epistemological credibility to the way they structure and interpret their findings. In addition, the introduction draws on prior studies that establish the academic and health benefits of MI for children—for example, improvements in on-task behavior, test scores, and classroom management (p. 2). These references serve as empirical backing, showing that MI is not only feasible but desirable, thereby lending legitimacy to the practical recommendations made later in the review.	E	

Qualifier	<p>The qualifier is consistently expressed through hedging language that modulates the strength of their claims. For example, in the discussion of implications, the authors use phrases such as “may enhance,” “could support,” and “appears important” (pp. 8–9), indicating moderate confidence in the practical applicability of their findings. These phrases function to convey epistemic caution, acknowledging the variability of school contexts and the limitations of the included evidence. This is typical of qualitative and configurative reviews, where conclusions are interpretively derived and require nuance in how strongly they are framed. Thus, the authors’ use of qualifiers reflects a responsible and field-appropriate approach to evidence communication. It aligns with Toulmin’s call for arguments to specify their modal strength, and with Gough’s insistence that evidence claims must be transparent about their certainty and attuned to user needs. By employing cautious but meaningful language, the review signals the conditional nature of its recommendations, while still maintaining their relevance and usefulness for practice and policy.</p>	E	
Rebuttal	<p>The authors incorporate rebuttals in several ways. They clearly state that the review findings are limited to institutional and intrapersonal levels of the Social Ecological Model, due to the absence of studies exploring community, interpersonal, or policy-level factors: “Few studies addressed factors beyond the institutional and intrapersonal levels” (p. 9). This functions as both a limitation of the evidence base and a pointer to underexplored domains, qualifying the scope of the review’s conclusions. Additionally, they suggest that implementation challenges may persist even with supports in place, particularly if training or time constraints are not adequately addressed, further indicating an awareness of the complexity of real-world application.</p>	E	

Categories	(Savin-Baden et al., 2008) – Learning spaces, agency and notions of improvement: what influences thinking and practices about teaching and learning in higher education? An interpretive meta-ethnography (SR #2)	Explicit	Implicit
Topic	The review focuses on 'thinking and practices about teaching and learning in higher education', structured around three themes: practice, transfer, and community (p. 211).	E	
Importance	Describes the field as a 'minefield of overlapping concepts, with few clear frameworks' (p. 212), and highlights underexplored areas such as pedagogical stance, disjunction, and learning spaces (p. 211–212).	E	
Justification of Method	Authors argue meta-analysis is ill-suited to HE literature because it 'decontextualises material' (p. 214). They choose interpretive meta-ethnography for its ability to handle conceptual richness.	E	
Theoretical Framework	The review explicitly draws on the interpretive meta-ethnography framework by Noblit & Hare (1988) and extends it using Savin-Baden and Major's (2007) staged method. It also references critical theory and constructivism, including Freire, hooks, Barnett, and Giddens (pp. 212–213, 221–222).	E	
Framework Justification	Authors justify the use of interpretive meta-ethnography as suitable for capturing social and theoretical contexts and for dealing with the interpretive complexity of teaching and learning literature in HE (p. 214–215).	E	
Main Claim	The review concludes that the most influential factors affecting teachers' implementation of movement integration (MI) fall within two primary domains: institutional and intrapersonal. This dual-level framework is grounded in the authors' identification of 12 thematic categories, which they explicitly organize using the Social Ecological Model: "A total of 12 themes of MI facilitators and barriers were identified and categorized into two social-ecological levels: institutional factors (e.g., administrative support, resources) and intrapersonal factors (e.g., teacher confidence, ease of implementation)" (p. 5). Among institutional barriers, the most frequently reported challenge was lack of time: "CTs' biggest institutional barrier to MI was not having enough time" (p. 9). Additional institutional-level impediments included limited physical space and insufficient material resources, while key enablers were administrative support and school-level prioritization of physical activity: "At the institutional level, administrative support and resources for MI emerged as key factors" (p. 9). On the intrapersonal level, teacher confidence, beliefs about MI's benefits, and the perceived ease of implementation played central roles. Summarizing these findings, the authors state: "Our findings show that facilitators and barriers to MI exist at two levels of influence: institutional and intrapersonal" (p. 9). Overall, the review highlights how both structural conditions and individual dispositions shape the success of MI initiatives in elementary classrooms.	E	

<p>Other Claims</p>	<p>While the review focuses primarily on institutional and intrapersonal factors, the authors acknowledge a significant gap in the literature concerning interpersonal, community, and policy-level influences on movement integration (MI), noting that “few studies addressed factors beyond the institutional and intrapersonal levels” and identifying this as a direction for future research (p. 9). Among the intrapersonal facilitators, three key elements emerge: (a) teachers’ perception that physical activity (PA) is valuable, (b) their beliefs about the benefits of MI, and (c) their confidence in implementing MI strategies. Confidence is consistently highlighted as one of the most impactful enablers: “Teacher confidence was one of the most significant facilitators to teachers’ MI, and teachers’ playing and coaching experiences in PA (e.g., playing sports, earning coaching credentials) were important factors underpinning their confidence in using MI” (Usher & Anderton, 2014, p. 7). Additional intrapersonal factors include the perceived ease of implementation, as teachers were more likely to adopt MI when they viewed it as simple to integrate into daily routines. On the institutional level, the most frequently cited barrier was lack of time: “CTs’ biggest institutional barrier to MI was not having enough time” (p. 9). However, the authors argue that this barrier could be reframed through targeted professional development: “Efforts to increase teachers’ use of MI should therefore focus on helping teachers learn to view MI as a noncompetitor to other school priorities and even take advantage of MI as a time-saving strategy” (p. 9). Moreover, access to professional development and administrative support emerged as critical facilitators. As the review explains: “Teachers would likely face fewer MI implementation challenges and feel more motivated to use MI if their administrators provided increased access to MI trainings that focused on using a wide range of MI strategies and the conditions needed for successful MI implementation” (p. 9). These findings underscore that both personal beliefs and structural support mechanisms must be addressed to enhance MI uptake in schools, while also pointing to important underexplored areas at broader ecological levels.</p>	<p>E</p>	
<p>Claim Framing</p>	<p>The authors frame their central claim as practical and policy-relevant, with a clear orientation toward informing professional development, intervention design, and school-level implementation strategies (pp. 2, 9). Rather than positioning the findings as theoretical contributions, the review consistently emphasises their utility for shaping classroom practices and guiding decision-making at the institutional level. The findings are presented as actionable insights to improve school-based movement integration (MI) by addressing both structural and attitudinal barriers. For instance, the authors suggest that “sharing research on the academic benefits of MI (e.g., improved on-task behavior, increased standardized test scores) [...] may be an important step in convincing teachers that MI will reduce the time needed to gain students’ attention and establish/reinforce a classroom management system” (p. 9). This statement directly connects the review’s conclusions to practical teacher concerns, such as time management and classroom control. Similarly, they highlight that “conditions for success could include enacting appropriate classroom management protocols,” listing specific examples such as establishing ground rules and using calming activities to transition back to seatwork (p. 9). These examples reinforce the pragmatic orientation of the review’s recommendations and position the findings as immediately applicable within the realities of elementary school teaching. Overall, the claim is framed in a way that supports educator empowerment, institutional change, and evidence-informed policy decisions, making it a clear example of a practically framed evidence claim.</p>	<p>E</p>	

Grounds	The grounds are clearly articulated through the systematic selection and thematic analysis of 28 peer-reviewed studies. These studies were retrieved from four major databases—Google Scholar, ERIC, PubMed, and PsycINFO—using a comprehensive search strategy: “A literature search was conducted to identify all published research, in English, that reported facilitators and/or barriers to using MI in elementary school settings” (p. 3). The included studies span a range of methodological designs and were subjected to content analysis, with themes identified by coding the results sections. The synthesis process is transparently reported, including a PRISMA flow diagram that documents the screening and inclusion criteria (p. 4). Thematically derived findings—such as institutional barriers (e.g., lack of time) and facilitators (e.g., teacher confidence)—are grounded in empirical evidence across these 28 studies. As such, the grounds for the review’s claims are both systematically retrieved and analytically processed.	E	
Synthesis Type	The synthesis is best understood as a configurative, thematic content analysis. The authors state that they applied content analysis to code and synthesize findings across studies: “Relevant data were extracted from each article and organized using an author-developed coding form... categorized using a social-ecological framework” (p. 3). This approach aligns with a configurative logic, as it aims not to summarize predefined outcomes but to construct thematic interpretations across diverse studies. While they do not use the term “thematic synthesis” explicitly, their approach shares core features of that methodology, particularly the inductive organization of findings into higher-order categories. These categories (e.g., institutional and intrapersonal barriers/facilitators) were not aggregated outcomes, but interpretive constructions aimed at building conceptual understanding of what shapes MI implementation.	E	
Validity/Relevance	Attention to validity and relevance is demonstrated through multiple procedural safeguards. First, the review adheres to PRISMA guidelines (p. 4), providing a clear and replicable account of search, screening, and selection processes. Second, the study describes a dual-coding process used during data extraction, enhancing inter-rater reliability and minimizing interpretive bias (p. 3). Third, the authors explicitly acknowledge the limitations of their synthesis, noting that their findings are restricted to institutional and intrapersonal levels of the Social Ecological Model (SEM), while interpersonal, community, and policy-level influences were underexplored due to gaps in the available literature (p. 9). This reflexivity about the scope and limitations of the evidence base strengthens the transparency and relevance of their claims.	E	
Warrant	The warrant is not explicitly stated but can be clearly inferred from the structure and intent of the argument: namely, that removing institutional barriers (e.g., lack of time, insufficient training) and strengthening intrapersonal facilitators (e.g., teacher confidence, positive beliefs) will increase teachers’ use of movement integration (MI) strategies. The review's organization and discussion sections suggest a causal or at least a strongly conditional logic: If teachers are given administrative support, time-saving strategies, and adequate training, then they are more likely to adopt MI in the classroom. This interpretive assumption is consistent with the use of the Social Ecological Model (SEM), which embeds human behavior within layered systems of influence and suggests that changes at the institutional or intrapersonal level can trigger behavioral shifts.		I

Backing	The warrant—that institutional and intrapersonal conditions influence teachers’ use of movement integration (MI)—is supported by explicit theoretical and empirical backing. The review is grounded in the Social Ecological Model (SEM), which is presented as a theoretical framework suited for capturing multi-level factors that influence teacher behavior (p. 3). The authors justify its use by explaining that SEM “provides a meaningful way to conceptualize the different levels of factors that can influence behavior” (p. 3), thereby giving epistemological credibility to the way they structure and interpret their findings. In addition, the introduction draws on prior studies that establish the academic and health benefits of MI for children—for example, improvements in on-task behavior, test scores, and classroom management (p. 2). These references serve as empirical backing, showing that MI is not only feasible but desirable, thereby lending legitimacy to the practical recommendations made later in the review.	E	
Qualifier	The qualifier is consistently expressed through hedging language that modulates the strength of their claims. For example, in the discussion of implications, the authors use phrases such as “may enhance,” “could support,” and “appears important” (pp. 8–9), indicating moderate confidence in the practical applicability of their findings. These phrases function to convey epistemic caution, acknowledging the variability of school contexts and the limitations of the included evidence. This is typical of qualitative and configurative reviews, where conclusions are interpretively derived and require nuance in how strongly they are framed. Thus, the authors’ use of qualifiers reflects a responsible and field-appropriate approach to evidence communication. It aligns with Toulmin’s call for arguments to specify their modal strength, and with Gough’s insistence that evidence claims must be transparent about their certainty and attuned to user needs. By employing cautious but meaningful language, the review signals the conditional nature of its recommendations, while still maintaining their relevance and usefulness for practice and policy.	E	
Rebuttal	The authors incorporate rebuttals in several ways. They clearly state that the review findings are limited to institutional and intrapersonal levels of the Social Ecological Model, due to the absence of studies exploring community, interpersonal, or policy-level factors: “Few studies addressed factors beyond the institutional and intrapersonal levels” (p. 9). This functions as both a limitation of the evidence base and a pointer to underexplored domains, qualifying the scope of the review’s conclusions. Additionally, they suggest that implementation challenges may persist even with supports in place, particularly if training or time constraints are not adequately addressed—further indicating an awareness of the complexity of real-world application.	E	

Categories	(Harrison et al., 2019) – Curriculum and Learning in Aboriginal and Torres Strait Islander Education: A Systematic Review (SR #3)	Explicit	Implicit
Topic	The review focuses on how curriculum governs learning for Aboriginal and Torres Strait Islander students in Australia, with emphasis on representation, epistemology, and cultural identity (p. 234).	E	
Importance	It critiques how national priorities (e.g. NAPLAN, attendance) overshadow meaningful educational goals such as identity, belonging, and cultural knowledge. The review contrasts mainstream definitions of success with Indigenous community values (p. 234–235, 243).	E	
Justification of Method	The authors justify the SLR as a way to test historical assumptions against empirical evidence and to identify alternative ways of knowing and teaching (p. 234). It follows the Aboriginal Voices project's structured protocol and critical appraisal framework (p. 236–237).	E	
Theoretical Framework	The review draws on decolonial theory, Indigenous epistemologies (e.g. Country as teacher), and critiques of representation in curriculum (e.g. Verran 2013, Green 2018, Barad 2007). Authors emphasize ontological plurality and the cultural production of knowledge (p. 235–246).	E	
Framework Justification	Frameworks like both-ways education, Red Dirt Thinking, and land-based pedagogies are justified as better suited to Indigenous learners' realities. The authors argue that mainstream epistemologies marginalize relational, place-based knowledge (p. 241–246).	E	
Main Claim	The review claims that mainstream curriculum structures misalign with Indigenous concepts of learning, success, and knowledge. It proposes a reconceptualization of curriculum that recognizes Country, identity, and community as central to pedagogy (p. 243–245).	E	
Other Claims	Other claims include the importance of teacher engagement with local communities, the legitimacy of non-western ontologies, and a critique of deficit discourses. Parents' involvement and intergenerational learning are framed as pedagogical strengths (p. 244).	E	
Claim Framing	While Harrison et al. (2019) make strong normative and political assertions, such as the misalignment between mainstream curriculum structures and Indigenous ways of knowing, these claims are more embedded in the narrative and theoretical discourse than explicitly signposted as 'claims' in the formal argumentative sense. Drawing from concepts such as Country as curriculum and ontological plurality, the authors critique epistemic injustice and deficit models of education. However, they do not formally foreground these positions as contestable claims with clearly articulated stakes or counterpositions. Instead, their critique is woven through conceptual elaboration and theoretical synthesis (e.g. discussions of representation, identity and relational learning), relying on the reader's interpretive engagement to discern the main argument threads. Thus, the framing of the claims is implicit in terms of tone, structure and organisation rather than being stated overtly in argumentative terms (e.g. 'We argue that X because Y').		I

Grounds	The review synthesizes 29 qualitative studies using a structured appraisal process. Key themes—such as Country as curriculum, knowledge representation, and the purpose of learning—are supported by detailed case evidence (p. 236–249).	E	
Synthesis Type	The synthesis is interpretive and thematic, following the Aboriginal Voices protocol. It emphasizes concept-led aggregation of themes across context-rich studies (p. 236, 241–246).	E	
Validity/Relevance	Quality appraisal used a 6-point framework. Studies scoring below 3/6 were excluded. Authors highlight limits of generalizability but argue for conceptual transferability (p. 236–237, 241).	E	
Warrant	The review draws on converging themes from a wide range of qualitative studies, including empirical findings, community narratives and critical theory, to support its claims. However, the inferential logic linking data to conclusions is largely taken for granted rather than being justified step by step. For instance, the authors contend that standardised outcomes (such as NAPLAN scores) inadequately reflect Indigenous conceptions of success and advocate relational, place-based curriculum models as alternatives. This position is supported by rich thematic synthesis and theory-informed examples (e.g. Guenther et al., Verran, Ewing), but the warrant — the logical bridge that legitimises the leap from evidence to generalised claim — is never fully explained. Instead, credibility is derived from coherence and resonance rather than from the explicit justification of why these data necessarily warrant the broader claims made. Rather than providing formal warrants, the authors trust the reader’s sense of plausibility and critical alignment.		I
Backing	The review is supported by theoretical references to Barad (2007), Verran (2010), Green (2018), and empirical studies by Guenther et al., Disbray, and Guyula that exemplify Indigenous learning frameworks (p. 241–248).	E	
Qualifier	The authors use tentative language like 'may', 'tend to', and 'can be' while also emphasizing the strength of convergence among findings (p. 245–247).	E	
Rebuttal	The review acknowledges that dominant paradigms define success and inclusion, and argues these constructs must be critically re-evaluated. It critiques measurement-based outcomes and offers alternatives grounded in Indigenous perspectives (p. 243–247).	E	

Categories	(Kim & Graham, 2022) – CLIL Teachers’ Needs and Professional Development: A Systematic Review (SR #4)	Explicit	Implicit
Topic	The review focuses on CLIL (Content and Language Integrated Learning) teachers’ self-reported needs and professional development across seven core competencies: linguistic, pedagogical, scientific knowledge, organizational, interpersonal, collaborative, and reflective. It aims to identify how reported challenges align with or diverge from professional development literature (p. 2–3).	E	
Importance	The review responds to a critical gap in CLIL literature: the mismatch between what teachers report needing and what PD actually addresses. The authors argue that without structured, needs-based PD, teachers cannot deliver effective CLIL instruction, thus undermining CLIL’s educational promise (p. 3–4).	E	
Justification of Method	A systematic literature review was employed following Zawacki-Richter et al. (2020). The search covered six databases, applying eight inclusion criteria to ensure relevance and quality. The goal was to synthesize evidence regarding both needs and PD experiences (p. 5–6).	E	
Theoretical Framework	The review adopts Pérez Cañado’s (2018) framework of seven CLIL teacher competencies: linguistic, pedagogical, scientific knowledge, organizational, interpersonal, collaborative, and reflective. This model structures both coding and interpretation (p. 3–4).	E	
Framework Justification	The authors justify their framework choice by its comprehensiveness and its explicit focus on linking teacher needs to pedagogical competence development—both theoretical and practical. It facilitates gap analysis across domains (p. 3–4, 13–14).	E	
Main Claim	The main claim is that professional development does not consistently address the full range of competencies CLIL teachers report needing. While some areas (e.g. pedagogical, linguistic) are well-covered, others (e.g. interpersonal, collaborative) are significantly underrepresented (p. 12–14).	E	
Other Claims	Other claims include that even commonly addressed competencies are often treated superficially; that time constraints, resource shortages, and collaboration barriers persist across contexts; and that future PD must better reflect specific teacher profiles (p. 13–15).	E	
Claim Framing	This review makes claims through a normative, decolonial and justice-oriented lens, but this framing is not explicitly stated as a rhetorical device. Instead, readers must infer the review’s position by interpreting the repeated critique of mainstream curricula and the celebration of Indigenous knowledge systems. The authors do not explicitly indicate that they are contrasting competing educational paradigms or framing claims for a specific audience or purpose. Rather, the framing is embedded in the review’s choice of theoretical references (e.g. Barad and Verran), its critique of		I

	standardised assessment and its conceptualisation of curriculum as 'Country'. This makes the framing visible through interpretive cues, but it is neither explicitly declared nor reflexively positioned as an argumentative strategy.		
Grounds	The review synthesizes 43 empirical studies from diverse regions, with coding structured by competence area. Most studies concern teacher-reported needs (n=33), with fewer on PD (n=9) or both (n=1). Tables and thematic summaries support analytic conclusions (p. 7–9).	E	
Synthesis Type	The authors conduct an interpretive thematic synthesis using a researcher-designed coding matrix based on the Pérez Cañado framework. Each included study was coded for which competencies it addressed (p. 6–7).	E	
Validity/Relevance	Studies were appraised using eight inclusion criteria focusing on empirical quality, language, educational level, and thematic fit. The review acknowledges its linguistic scope limitation (English-only) (p. 5–6, 15).	E	
Warrant	The framing of claims in this review is also implicit. The central argument — that professional development does not adequately address the range of needs reported by teachers — is presented as a finding rather than as an issue of educational design, policy misalignment or systemic oversight. The framing is analytical, but not overtly strategic; it emerges through the structure of the comparison between needs and professional development (PD) responses. There is no meta-commentary explaining why this misalignment is important and to whom (e.g. policy-makers, teacher trainers or curriculum designers). Consequently, the review's framing remains descriptive and analytical rather than explicitly persuasive or directive.		I
Backing	The review is supported by theoretical work on CLIL (e.g., Coyle & Meyer 2021; Lo 2020), language acquisition, and teacher identity formation. It also draws on prior systematic reviews and emerging frameworks for PD design (p. 2–4, 13–15).	E	
Qualifier	Claims are tentatively phrased (e.g. 'suggest', 'may indicate'), especially when referring to underdeveloped competence areas or possible future research needs (p. 13–16).	E	
Rebuttal	The authors acknowledge the limitations of their review, including linguistic scope, potential omission of non-indexed studies, and the early stage of CLIL PD research. These caveats temper their conclusions (p. 15–16).	E	

Categories	Sheng (2020) – Education Policy and Practice on IPV among Young People in the UK (SR #5)	Explicit	Implicit
Topic	The review focuses on educational policy and practice in the UK addressing intimate partner violence (IPV) among young people. It specifically reviews school-based interventions and educational strategies, aiming to assess their impact and propose recommendations for IPV prevention within formal education (p. 98–99).	E	
Importance	The study addresses a major policy and research gap: while IPV is well-studied in the US, UK-based school interventions have been relatively neglected. Sheng (2020) argues that UK-specific evidence is needed due to cultural specificity in IPV experiences and responses (p. 99–100).	E	
Justification of Method	A qualitative systematic review approach was adopted, covering ERIC, BEI, Scopus, and grey literature. 1,161 articles were screened using PICO and Boolean searches, with 11 studies meeting strict inclusion criteria focused on qualitative educational IPV interventions in the UK (p. 101–102).	E	
Theoretical Framework	The study is informed by feminist and policy analysis traditions, with a focus on how IPV educational policy engages with gender, agency, and relational dynamics. The review uses thematic synthesis (Thomas & Harden, 2008) to guide coding and interpretation (p. 111).	E	
Framework Justification	The choice of thematic synthesis is justified by the need for both descriptive and analytical interpretation of diverse qualitative findings. The method supports the generation of both policy insights and intervention critiques (p. 111).	E	
Main Claim	The review finds that UK school-based IPV interventions have positive impacts on awareness and attitudes but remain fragmented, under-evaluated, and inconsistently implemented. Longitudinal data and national-level strategies are lacking (p. 98, 112–114).	E	
Other Claims	Additional claims include the need for gender-sensitive content, inclusion of same-sex and digital abuse contexts, improved teacher training, and country-wide curriculum reform. Students' discomfort, gendered assumptions, and ineffective delivery methods are also reported (p. 109–114).	E	
Claim Framing	Although Sheng (2020) clearly states the central claims — that educational IPV interventions in the UK are fragmented, small in scale and insufficiently evaluated — the framing of these claims remains largely implicit. They are not situated within a broader epistemological, theoretical or policy critique (e.g. the role of schools in achieving gender justice or the limitations of attitude-based interventions). Nor are they systematically introduced and developed through an argumentative structure. Instead, the claims emerge inductively from the		I

	thematic findings and are presented as observations rather than propositions based on a normative or theoretical standpoint. For instance, the conclusion states that 'more studies are needed' and 'IPV education is still in its infancy' (p. 115), yet it fails to clearly articulate what kind of knowledge or change the review is advocating for and what assumptions guide its interpretation of 'effectiveness' or 'impact'. This makes the framing of the claims more inferential than articulated.		
Grounds	The synthesis includes 11 studies representing a variety of school-based interventions, including drama-based workshops, curricular projects, and digital campaigns. Methods included interviews, focus groups, and pre/post evaluations (p. 102–104).	E	
Synthesis Type	Three-stage thematic synthesis (Thomas & Harden, 2008): line-by-line coding, descriptive themes, and analytical themes. Themes included impacts, evaluations, and recommendations (p. 111).	E	
Validity/Relevance	Eight inclusion criteria ensured methodological rigor: peer-reviewed, English language, UK-based, ages 10–19, educational setting, IPV-related focus, and qualitative methods. Scope limitations are acknowledged (p. 101–103, 115).	E	
Warrant	The connection between evidence and claims in Sheng’s review is not accompanied by an explicit warrant—that is, a statement explaining why the evidence supports the conclusions. The review presents converging findings from the 11 included studies—such as improvements in awareness, limited long-term behavioral outcomes, and student preferences for support—but does not explain why these patterns justify the broader conclusion that IPV interventions need to be more systemic, inclusive, and longitudinal. For instance, claims about the inadequacy of current efforts are built on repetition across studies rather than a formal reasoning process or articulated evaluative criterion (such as a theory of educational change or policy implementation). In Toulmin’s terms, the warrant is assumed rather than stated—readers are expected to infer that convergence equals justification, but this step is not made explicit in the argumentation.		I
Backing	The review is supported by policy frameworks (Home Office 2016; DfE 2019), and empirical studies on UK-based IPV interventions (Fox et al., 2016; Bell & Stanley, 2006; Dobson & Ringrose, 2016). It also refers to feminist and youth development literature (p. 98–114).	E	
Qualifier	Claims use cautious language (e.g. 'suggests', 'may promote') and call for more rigorous, long-term evaluations before definitive outcomes can be asserted (p. 114–115).	E	

Rebuttal	The author acknowledges mixed findings regarding delivery (teacher vs. external facilitator), challenges of gender representation, and student discomfort. The review critiques the 'one-size-fits-all' approach and recommends inclusive, adaptive models (p. 112–114).	E	
-----------------	--	---	--

Categories	(Swearingen, 2019) – Nonnative-English-Speaking Teacher Identity in TESOL (SR #6)	Explicit	Implicit
Topic	This review examines how non-native English-speaking teacher candidates (NNES-TCs) develop their language teacher identities (LTIs) in graduate TESOL programs in English-dominant countries. It synthesises 17 studies to examine how teacher preparation experiences influence identity formation (p. 1).	E	
Importance	The study addresses how NNES-TCs often navigate marginalising discourses (e.g., native-speakerism, racism, linguistic deficit) while trying to claim legitimacy as language teachers. It argues that identity is crucial to teacher preparation, yet under-theorised in LTP design (pp. 2–4).	E	
Justification of Method	A systematic review was conducted using criteria including identity-focused research questions, NNES participants, TESOL MA-level programs in English-dominant countries, and qualitative methods. Databases and journals were searched and studies were peer-reviewed (pp. 3–4).	E	
Theoretical Framework	The review draws on sociocultural and poststructuralist theories of identity (Norton, 1997; Varghese et al., 2005), emphasising identity as dynamic, multiple, socially constructed, and agentively negotiated. Teacher identity is framed as both situated and performative (pp. 2–3).	E	
Framework Justification	The chosen framework enables analysis of identity-in-practice and identity-in-discourse, foregrounding how NNES-TCs both inhabit and resist socially ascribed roles. The review situates itself in the tradition of the 'social turn' in SLA (Firth & Wagner, 1997) (p. 2).	E	
Main Claim	LTI development for NNES-TCs is shaped by a tension between deficit-based discourses and emerging counter-discourses that allow for agentive identity construction. Programs that acknowledge identity work can empower NNES-TCs and improve preparation outcomes (pp. 10–12).	E	
Other Claims	Subclaims include the emotional labour of identity work, the racialization and gendering of NNES-TCs, the clash of academic and professional identities, and the risk of programs perpetuating epistemological racism and the native speaker fallacy (pp. 5–10, 12–14).	E	
Claim Framing	In Swearingen (2019), the main claims are not presented as assertive statements linking data directly to conclusions. Instead, they are framed narratively and dispersed across the thematic findings. The author discusses how non-native English speaker (NNES) teacher candidates struggle with legitimacy, internalise and resist native-speaker ideologies, and negotiate their professional identities within Teaching English to Speakers of Other Languages (TESOL) programmes. These points come together to form an implicit claim: that teacher training programmes must change their approach to better support the identity development of NNES candidates. However, this claim is not presented as a clear, standalone		I

	statement; rather, it is built cumulatively through descriptive synthesis. The normative stance — advocating empowerment and equity — is expressed through the thematic structure and critical framing of literature, without explicitly naming the review's overarching argument as a 'claim'. This interpretive and value-oriented approach aligns with a critical-constructive stance on qualitative synthesis. However, it makes the reasoning structure less transparent from an argumentative perspective (Toulmin, 1958).		
Grounds	The synthesis includes 17 qualitative studies from the US, Canada, and Australia. Data sources include narratives, interviews, journals, course reflections, and class observations. Thematic analysis was applied to construct four main categories of findings (p. 4–5).	E	
Synthesis Type	Thematic synthesis following Suri & Clarke (2009) and Thomas & Harden (2008). Themes emerged from open and axial coding: native speaker fallacy, racialized/gendered identities, academic identity clashes, and affective dimensions (p. 4–5).	E	
Validity/Relevance	Selection criteria included empirical focus on identity, clear participant description, relevance to MA-level TESOL in English-dominant settings, and qualitative method use. Studies were peer-reviewed and contextually diverse (p. 4).	E	
Warrant	The justifications connecting the synthesised data to the review's claims, or 'warrants', are largely implicit and inferred through the cumulative convergence of findings. Swearingen justifies claims about NNES identity development and the need for programmatic change by relying on thematic coherence across 17 studies. However, she does not explicitly state how or why the evidence base supports the broader conclusions. For instance, the findings concerning the emotional and racialised challenges faced by NNES candidates are not accompanied by a formal explanation of why these findings justify altering TESOL programme structures. Theoretical backing from Norton (1997), Varghese et al. (2005) and Barkhuizen (2017) serves as implicit justification: it lends epistemological legitimacy to the interpretation of identity as fluid, contested and discursively constructed. However, the link between this framework and the claims is never explicitly explained in Toulmin's terms. Consequently, while the review demonstrates conceptual richness, it leaves the logic of its inferential reasoning largely to the reader.		I
Backing	The review draws on key identity theorists (e.g., Norton, Varghese, Pavlenko, Park) and supports findings with case studies from multiple cultural and institutional settings. It also references postcolonial critiques of TESOL (e.g., Phillipson, Selvi) (pp. 5–14).	E	
Qualifier	Cautionary language is used consistently (e.g. 'suggest', 'may', 'appears') and the author acknowledges methodological limitations, especially the essentialization risk in labeling participants NNES-TCs (pp. 12–13).	E	

Rebuttal	The review critiques native speakerism, White normativity, and neoliberal models of teacher preparation. It also addresses conflicting findings on NNES-TCs' agency, and institutional silence around race and gender (pp. 11–14).	E	
-----------------	--	---	--

References

- Aguinaldo, J. P. (2004). Rethinking validity in qualitative research from a social constructionist perspective: From "Is this valid research?" to "What is this research valid for?". *The Qualitative Report*, 9(1), 127–137.
- Aiken, L. R. (2002). *Attitudes and Related Psychosocial Constructs: Theories, Assessment, and Research*. SAGE Publications.
- Alexander, P. A. (2020). Methodological guidance paper: The art and science of quality systematic reviews. *Review of Educational Research*, 90(1), 6–23.
- Andrews, R. (2005). The Place of Systematic Reviews in Education Research. *British Journal of Educational Studies*, 53(4), 399–416.
- Andrews, R. (2010). *Argumentation in higher education: Improving practice through theory and research*. Routledge.
- Anfara, V. A., & Mertz, N. T. (2015). *Theoretical frameworks in qualitative research* (Second edition). Sage.
- Aromataris, E., Fernandez, R., Godfrey, C. M., Holly, C., Khalil, H., & Tungpunkom, P. (2015). Summarizing systematic reviews: Methodological development, conduct and reporting of an umbrella review approach. *International Journal of Evidence-Based Healthcare*, 13(3), 132–140.
- Bengtsen, S. S. E. (2018). Supercomplexity and the university: Ronald Barnett and the social philosophy of Higher Education. *Higher Education Quarterly*, 72(1), 65–74.
- Biesta, G. (2007). Why "What Works" won't work: evidence-based practice and the democratic deficit in educational research. *Educational Theory*, 57(1), 1–22.
- Biesta, G. (2020). *Educational research: An unorthodox introduction*. Bloomsbury Academic.

- Booth, A., Noyes, J., Flemming, K., Gerhardus, A., Wahlster, P., Wilt, G., Mozygemba, K., Refolo, P., Sacchini, D., Tummers, M., & Rehfuess, E. (2016). *Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions*. Integrate-HTA.
- Borenstein, M. (2009). Effect sizes for continuous data. In L. V. H. H. Cooper & J. C. Valentine (Eds.). (Ed.), *The handbook of research synthesis and meta-analysis* (Book, Section, pp. 221–235). Russell Sage Foundation.
- Borg, S. (2017). Teachers' Beliefs and Classroom Practices. In *The Routledge Handbook of Language Awareness*. Routledge.
- Borgnakke, K. (2017). Meta-ethnography and systematic reviews – linked to the evidence movement and caught in a dilemma. *Ethnography and Education*, 12(2), 194–210.
- Bourke, B. (2014). Positionality: Reflecting on the Research Process. *The Qualitative Report*, 19(33), 1–9.
- Campbell, A., Taylor, B., Bates, J., & O'Connor-Bones, U. (2018). Developing and Applying a Protocol for a Systematic Review in the Social Sciences. *New Review of Academic Librarianship*, 24(1), 1–22.
- Cañado, M. L. P. (2018). Innovations and Challenges in CLIL Teacher Training. *Theory Into Practice*, 57(3), 212–221.
- Carroll, C., & Booth, A. (2015). Quality assessment of qualitative evidence for systematic review and synthesis: Is it meaningful, and if so, how should it be performed? *Research Synthesis Methods*, 6(2), 149–154.
- Cohen, Louis, Manion, Lawrence, & Morrison, Keith. (2018). *Research Methods in Education: Eighth edition*. Routledge.
- Cooper, C., Booth, A., Husk, K., Lovell, R., Frost, J., Schauburger, U., Britten, N., & Garside, R. (2022). A Tailored Approach: A model for literature searching in complex systematic reviews. *Journal of Information Science*, 1030-1062

- Denzin, N. K., & Lincoln, Y. S. (2018). *The SAGE handbook of qualitative research* (Fifth edition.). SAGE.
- Dixon-Woods, M., Bonas, S., Booth, A., Jones, D. R., Miller, T., Sutton, A. J., Shaw, R. L., Smith, J. A., & Young, B. (2006). How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research*, 6(1), 27–44.
- Dixon-Woods, M., Booth, A., & Sutton, A. J. (2007). Synthesizing qualitative research: A review of published reports. *Qualitative Research*, 7(3), 375–422.
- Dixon-Woods, M., Fitzpatrick, R., & Roberts, K. (2001). Including qualitative research in systematic reviews: Opportunities and problems. *Journal of Evaluation in Clinical Practice*, 7(2), 125–133.
- Eagly, A. H. (with Internet Archive). (1993). *The psychology of attitudes*. Fort Worth, TX : Harcourt Brace Jovanovich College Publishers.
- Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin’s Argument Pattern for studying science discourse. *Science Education*, 88(6), 915–933.
- Evans, J., & Benefield, P. (2001). Systematic reviews of educational research: Does the medical model fit? *British Educational Research Journal*, 27(5), 527–541.
- Finfgeld-Connett, D. (2014). Use of content analysis to conduct knowledge-building and theory-generating qualitative systematic reviews. *Qualitative Research*, 14(3), 341–352.
- Foucault, M. (1980). *Power / knowledge: Selected interviews an*. The Harvester press.
- Freire, P., & Torres, C. A. (2023). *Education for Critical Consciousness*. Bloomsbury Academic.
- Garside, R. (2014). Should we appraise the quality of qualitative research reports for systematic reviews, and if so, how? *Innovation: The European Journal of Social Science Research*, 27(1), 67–79.

- Gessler, M., & Siemer, C. (2020). Umbrella review: Methodological review of reviews published in peer-reviewed journals with a substantial focus on vocational education and training research. *International Journal for Research in Vocational Education and Training*, 7(1), 91–125.
- Ghirotto, L. (2020). *La systematic review nella ricerca qualitativa*. Rome: Carocci.
- Giddens, A. (1984) *The Constitution of Society. Outline of the Theory of Structuration*. University of California Press, Berkeley.
- Glaser, B. G., Strauss, A. Leonard., & Strati, Antonio. (2003). *La scoperta della grounded theory: Strategie per la ricerca qualitativa*. Armando.
- Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, 22(2), 213–228.
- Gough, D. (2021). Appraising Evidence Claims. *Review of Research in Education*, 45(1), 1–26.
- Gough, D. (2022). Appraising Evidence Claims. *Review of Research in Education*, 45(1), 1–26.
- Gough, D., Davies, P., Jamtvedt, G., Langlois, E., Littell, J., Lotfi, T., Masset, E., Merlin, T., Pullin, A. S., Ritskes-Hoitinga, M., Røttingen, J.-A., Sena, E., Stewart, R., Tovey, D., White, H., Yost, J., Lund, H., & Grimshaw, J. (2020). Evidence Synthesis International (ESI): Position Statement. *Systematic Reviews*, 9(1), 155.
- Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews* (2nd edition.). Sage.
- Gough, D., Thomas, J., & Oliver, S. (2019). Clarifying differences between reviews within evidence ecosystems. *Systematic Reviews*, 8(1), 170.
- 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.

- Harrison, N., Tennent, C., Vass, G., Guenther, J., Lowe, K., & Moodie, N. (2019). Curriculum and learning in Aboriginal and Torres Strait Islander education: A systematic review. *The Australian Educational Researcher*, 46(2), 233–251.
- Hooks, B. (2014). *Teaching To Transgress*. Routledge.
- Jhangiani, R. S., Cuttler, C., Leighton, D. C., & Open Textbook Library. (2019). *Research Methods in Psychology* (4th Edition). Kwantlen Polytechnic University.
- Karbach, J. (1987). Using Toulmin’s Model of Argumentation. *Journal of Teaching Writing*, 6(1), 81–91.
- Khambete, P. (2019). Adaptation of Toulmin’s Model of Argumentation for Establishing Rigour and Relevance in Design Research. In A. Chakrabarti (Ed.), *Research into Design for a Connected World* (pp. 3–13). Springer.
- Kim, H., & Graham, K. M. (2022). CLIL Teachers’ Needs and Professional Development: A Systematic Review. *Latin American Journal of Content & Language Integrated Learning*, 15(1), e1515.
- La Torre, G., Backhaus, I., & Mannocci, A. (2015). Rating for narrative reviews: Concept and development of the International Narrative Systematic Assessment tool. *Senses and Sciences*, 2(2), 31–35.
- Larsen, M. K., Lehmkuhl, L., & Missel, M. (2025). Identifying Metaphors as a Way to Strengthen Qualitative Findings. *International Journal of Qualitative Methods*, 24,
- Lather, P. (1986). Research as Praxis. *Harvard Educational Review*, 56(3), 257–278.
- Levay, P., & Craven, J. (2018). Introduction: Where are we now? In P. Levay & J. Craven (Eds), *Systematic Searching* (Book, Section, pp. 1–24). Facet.
- Lewin, S., Booth, A., Glenton, C., Munthe-Kaas, H., Rashidian, A., Wainwright, M., Bohren, M. A., Tunçalp, Ö., Colvin, C. J., Garside, R., Carlsen, B., Langlois, E. V., & Noyes, J. (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings: Introduction to the series. *Implementation Science*, 13(1), 2.

- MacLure, M. (2005). 'Clarity bordering on stupidity': Where's the quality in systematic review? *Journal of Education Policy*, 20(4), 393–416.
- Maeda, Y., Caskurlu, S., Kenney, R. H., Kozan, K., & Richardson, J. C. (2022). Moving qualitative synthesis research forward in education: A methodological systematic review. *Educational Research Review*, 35, 100424.
- Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3. ed). Sage.
- McGuire, W.J. (1985) *Attitudes and Attitude Change*. In Lindzey, G. and Aronson, E., Eds., *The Handbook of Social Psychology, Vol. 2, Special Fields and Applications, 3rd Edition*, Random House Inc., 233-346. - References—Scientific Research Publishing.
- McLeroy, K. R., Bibeau, D., Steckler, A., & Glanz, K. (1988). An ecological perspective on health promotion programs. *Health Education Quarterly*, 15(4), 351–377.
- Melendez-Torres, G. J., O'Mara-Eves, A., Thomas, J., Brunton, G., Caird, J., & Petticrew, M. (2017). Interpretive analysis of 85 systematic reviews suggests that narrative syntheses and meta-analyses are incommensurate in argumentation. *Research Synthesis Methods*, 8(1), 109–118.
- Merriam, S. B., & Tisdell, E. J. 1955-. (2017). *Qualitative research: A guide to design and implementation* (4th ed.). San Francisco, CA: Jossey-Bass.
- Michael, R. D., Webster, C. A., Egan, C. A., Nilges, L., Brian, A., Johnson, R., & Carson, R. L. (2019). Facilitators and Barriers to Movement Integration in Elementary Classrooms: A Systematic Review. *Research Quarterly for Exercise and Sport*, 90(2), 151–162.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (Edition 3). Sage.
- Ming, N. C., & Goldenberg, L. B. (2021). Research Worth Using: (Re)Framing Research Evidence Quality for Educational Policymaking and Practice. *Review of Research in Education*, 45(1), 129–169.

- Mondéjar-Jiménez, J., & Vargas-Vargas, M. (2010). Determinant factors of attitude towards quantitative subjects: Differences between sexes. *Teaching and Teacher Education*, 26(3), 688–693.
- Munthe-Kaas, H. M., Glenton, C., Booth, A., Noyes, J., & Lewin, S. (2019). Systematic mapping of existing tools to appraise methodological strengths and limitations of qualitative research: First stage in the development of the CAMELOT tool. *BMC Medical Research Methodology*, 19(1), 113.
- Newman, M., & Elbourne, D. (2004). Improving the Usability of Educational Research: Guidelines for the REPOrting of Primary Empirical Research Studies in Education (The REPOSE Guidelines). *Evaluation & Research in Education*, 18(4), 201–212.
- Newman, M., & Gough, D. (2020). Systematic Reviews in Educational Research: Methodology, Perspectives and Application. In O. Zawacki-Richter, M. Kerres, S. Bedenlier, M. Bond, & K. Buntins (Eds), *Systematic Reviews in Educational Research: Methodology, Perspectives and Application* (pp. 3–22). Springer.
- Noblit, G. W., & Hare, R. Dwight. (1988). *Meta-ethnography: Synthesizing qualitative studies*. Sage.
- Nyimbili, F., & Nyimbili, L. (2024). Types of Purposive Sampling Techniques with Their Examples and Application in Qualitative Research Studies. *British Journal of Multidisciplinary and Advanced Studies*, 5(1), 90–99.
- Nylander, E., Fejes, A., & Milana, M. (2024). Exploring the themes of the territory: A topic modelling approach to 40 years of publications in International Journal of Lifelong Education (1982–2021). In *Forty Years of the International Journal of Lifelong Education, Volume I*. Routledge.
- Oakley, A. (2002). Social science and evidence-based everything: The case of education. *Educational Review*, 54(3), 277–286.

- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372.
- Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (Fourth edition.). SAGE Publications, Inc.
- Patton, M. Quinn. (2002). *Qualitative research & evaluation methods* (3. ed.). Sage.
- Pellegrini, M., & Vivanet, G. (2020). Evidence-Based Policies in Education: Initiatives and Challenges in Europe. *ECNU Review of Education*, 4, 25–45.
- Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.
- Pigott, T. D., Tocci, C., Ann, M. R., & Galliher, A. (2022). Quality of Research Evidence in Education: How Do We Know? *Review of Research in Education*, 45(1), vii–xii.
- Pollock, D., Peters, M. D. J., Khalil, H., McInerney, P., Alexander, L., Tricco, A. C., Evans, C., de Moraes, É. B., Godfrey, C. M., Pieper, D., Saran, A., Stern, C., & Munn, Z. (2023). Recommendations for the extraction, analysis, and presentation of results in scoping reviews. *JBIEvidence Synthesis*, 21(3), 520–532.
- Popay, J., Roberts, H. M., Sowden, A. J., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews A Product from the ESRC Methods Programme*.
- Ravitch, S. M., & Riggan, M. (2017). *Reason & rigor: How conceptual frameworks guide research* (2nd edition). Sage.
- Reinharz, Shulamit., Davidman, L. (1955-), & Oxford University Press. (2006). *Feminist methods in social research*. Oxford University Press.

- Roor, T., Mulliez, G., & Bonn, E. (2023). *Between Anything Goes and Methodical Rigor. An Empirical Analysis of Systematic Literature Reviews in Adult Education Research*. 46, 141–160.
- Sandelowski, M., & Barroso, J. (2006). *Handbook for synthesizing qualitative research*. Springer Publishing Company.
- Sandelowski, Margarete., & Barroso, Julie. (2007). *Handbook for synthesizing qualitative research*. Springer.
- Savin-Baden, M., McFarland, L., & Savin-Baden, J. (2008). Learning spaces, agency and notions of improvement: What influences thinking and practices about teaching and learning in higher education? An interpretive meta-ethnography. *London Review of Education*, 6(3), 211–227.
- Sheng, X. (2020). Education Policy and Practice on Intimate Partner Violence among Young People in the UK. *European Journal of Education (EJE)*, 3(2), 98–119.
- Slavin, R. E. (2002). Evidence-Based Education Policies: Transforming Educational Practice and Research. *Educational Researcher*, 31(7), 15–21.
- Stake, R. E. (1995). *The art of case study research*. Sage.
- Suri, H. (2011). Purposeful Sampling in Qualitative Research Synthesis. *Qualitative Research Journal*, 11(2), 63–75.
- Swearingen, A. J. (2019). Nonnative-English-speaking teacher candidates' language teacher identity development in graduate TESOL preparation programs: A review of the literature. *TESOL Journal*, 10(4), e494.
- Talbott, E., Maggin, D. M., Van Acker, E. Y., & Kumm, S. (2018). Quality Indicators for Reviews of Research in Special Education. *Exceptionality*, 26(4), 245–265.
- Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8(1), 45.

- Tod, D., Booth, A., & Smith, B. (2022). Critical appraisal. *International Review of Sport and Exercise Psychology*, 15(1), 52–72.
- Tong, A., Flemming, K., McInnes, E., Oliver, S., & Craig, J. (2012). Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology*, 12(1), 181.
- Toulmin, S. 1922-2009. (1958). *The uses of argument*. University Press.
- Toulmin, S. E. (2008). *The uses of argument* (Updated ed.). Cambridge University Press.
- Toye, F., Seers, K., Allcock, N., Briggs, M., Carr, E., & Barker, K. (2014). Meta-ethnography 25 years on: Challenges and insights for synthesising a large number of qualitative studies. *BMC Medical Research Methodology*, 14, 80.
- Wilson, S. M., & Anagnostopoulos, D. (2022). Methodological Guidance Paper: The Craft of Conducting a Qualitative Review. *Review of Educational Research*, 91(5), 651–670.
- Zawacki-Richter, O., Kerres, M., Bedenlier, S., Bond, M., & Buntins, K. (2020). *Systematic reviews in educational research: Methodology, perspectives and application*. Springer Nature.

Disclaimer on the Use of AI-Assisted Tools

Throughout the process of writing this thesis, I used an AI language model (ChatGPT, developed by OpenAI) to assist me with writing and editing. Specifically, I used the AI to help me:

- clarify and rephrase complex passages;
- receive suggestions on academic tone and structure;
- improve coherence and flow between sections;
- revise drafts in line with feedback while maintaining the integrity of the content.

At no point did I use the AI to generate original research ideas, fabricate data or replace my own analytical reasoning. All content, interpretations and arguments presented in this thesis reflect my own critical thinking. Where sources or citations were mentioned by the AI, I personally verified and contextualised them before including them in the final text.

I acknowledge the use of AI here out of a commitment to transparency and responsible research practices. Ultimately, I take full intellectual responsibility for the thesis — its design, content and conclusions.