What systematic review methodology can be employed to synthesize case studies and develop a comprehensive framework of best practices for successful STEM outreach initiatives across different educational contexts?

An optimal systematic review methodology for developing STEM outreach best practice frameworks combines PRISMA's structured search and screening procedures with specialized theoretical frameworks for synthesis, organizes analysis into sequential phases progressing from identification through thematic analysis to critical synthesis, employs validated quality assessment tools and independent coding for rigor, and explicitly details procedures for synthesizing qualitative case studies.

Abstract

This systematic review of five STEM education reviews reveals that methodology selection depends critically on review objectives. Reviews evaluating intervention effectiveness employed standardized PRISMA guidelines , while those developing conceptual frameworks utilized specialized approaches including Cronbach's UTOS framework and qualitative desktop research with multiple theoretical lenses . The most sophisticated framework development emerged from a three-level analysis approach that progressed through identification, thematic analysis, and critical synthesis phases , resulting in a comprehensive STEM Education Conceptions Framework with eight distinct themes . However, significant methodological gaps persist: only one review employed a validated quality assessment instrument , and only two provided explicit procedures for synthesizing qualitative data and case studies despite most primary STEM research employing qualitative or mixed methods . For developing frameworks of best practices across diverse educational contexts, the evidence suggests an optimal methodology would integrate PRISMA's systematic search procedures with specialized theoretical frameworks for synthesis, employ validated quality assessment tools , organize analysis into sequential phases moving from description through thematic identification to critical synthesis , ensure validity through independent coding , and explicitly detail procedures for handling case studies . The success of multi-framework approaches in producing comprehensive conceptual models indicates that theoretical pluralism may be essential for capturing the complexity of STEM education practices across contexts.

Paper search

We performed a semantic search using the query ”What systematic review methodology can be employed to synthesize case studies and develop a comprehensive framework of best practices for successful STEM outreach initiatives across different educational contexts?” across over 138 million academic papers from the Elicit search engine, which includes all of Semantic Scholar and OpenAlex.

We retrieved the 50 papers most relevant to the query.

Screening

We screened in sources based on their abstracts that met these criteria:

- **STEM Outreach Focus**: Does this study examine programs specifically designed to promote Science, Technology, Engineering, and/or Mathematics education or engagement outside traditional classroom settings?
- **Case Study Methodology**: Does this study employ case study research design (including single case studies, multiple case studies, or comparative case studies) that provides in-depth examination of STEM outreach programs?

- **Outcome Evaluation**: Does this study include assessment of program outcomes, effectiveness measures, participant feedback, or evaluation of program implementation and impact?
- **Peer-Reviewed Publication**: Is this study published in an academic journal, conference proceedings, or other peer-reviewed venue?
- **Beyond Regular Classroom Instruction**: Does this study focus on more than just traditional in-class STEM teaching methods by including outreach or community engagement components?
- **Empirical Content**: Does this study present empirical case study data or systematic synthesis of existing cases (rather than being purely theoretical or conceptual)?
- **STEM Discipline Focus**: Does this study focus on STEM disciplines (Science, Technology, Engineering, Mathematics) rather than exclusively examining humanities, social sciences, arts, or other non-STEM disciplines?

We considered all screening questions together and made a holistic judgement about whether to screen in each paper.

Data extraction

We asked a large language model to extract each data column below from each paper. We gave the model the extraction instructions shown below for each column.

- **Review Methodology**:

  Extract the systematic review methodology employed, including:

- Specific review approach (e.g., PRISMA, Cochrane, narrative review, scoping review)
- Search strategy and databases used
- Inclusion/exclusion criteria
- Quality assessment methods
- Any methodological frameworks applied (e.g., UTOS, triangle framing)
- How case studies or qualitative data were handled in the synthesis

- **Synthesis Approach**:

  Document how data was synthesized and analyzed, including:

- Type of analysis conducted (thematic analysis, meta-analysis, narrative synthesis, framework synthesis)
- Specific synthesis methods for combining case studies
- How qualitative and quantitative data were integrated (if applicable)
- Process for developing frameworks or models from the data
- Steps taken to ensure rigor in synthesis

- **STEM Outreach Practices**:

  Extract all STEM outreach practices, strategies, or interventions identified, including:

- Specific practices or approaches described
- Target populations (age groups, demographics, underrepresented groups)
- Types of STEM activities or programs
- Pedagogical approaches (problem-based learning, inquiry-based learning, project-based learning)
- Implementation strategies
- Any innovative or novel practices highlighted

- **Educational Contexts**:

Document the educational contexts and settings covered, including:

- Educational levels (elementary, secondary, undergraduate, etc.)
- Geographic contexts (national, international, specific countries/regions)
- Setting types (formal education, informal education, community programs)
- Institutional contexts (schools, universities, community centers, museums)
- Cultural or socioeconomic contexts mentioned
- Any context-specific adaptations noted

- **Frameworks Developed**:

  Extract details about any frameworks, models, or conceptual structures developed, including:

- Name and description of framework/model
- Key components or dimensions
- Theoretical underpinnings
- How the framework was derived from the data
- Intended applications or uses
- Visual representations or diagrams provided

- **Effectiveness Evidence**:

  Document evidence of success or effectiveness of STEM outreach practices, including:

- Outcome measures used to assess effectiveness
- Quantitative results (effect sizes, statistical significance, trends)
- Qualitative indicators of success
- Factors that contributed to success or failure
- Barriers and facilitators identified
- Long-term vs. short-term impacts reported

- **Study Characteristics**:

  Extract basic study characteristics including:

- Number of studies included in the review
- Types of primary studies synthesized (case studies, experiments, surveys, etc.)
- Publication date range of included studies
- Geographic distribution of primary studies
- Quality of included studies (if assessed)
- Any notable limitations in the evidence base

- **Methodological Quality**:

  Assess and extract information about methodological rigor, including:

- Strengths and limitations of the review methodology
- Risk of bias assessment (if conducted)
- Transparency of reporting
- Completeness of methodology description
- Author acknowledgments of limitations
- Recommendations for future methodological improvements
- Any concerns about study selection or synthesis bias

Results

Characteristics of Included Studies

Study	Full text retrieved?	Review approach	Studies included	Date range	Primary databases	Quality assessment
Andrew S. Leavitt et al., 2022	Yes	Systematic review using UTOS framework	80	2013-2020	ERIC, EBSCO	LEBIE scale
Hairunnisa Hussim et al., 2024	No	PRISMA	18	2007-2023	Not mentioned	PRISMA procedures
Jessica Tran et al. (n.d.)	Yes	PRISMA guidelines	104 (58 quantitative, 13 qualitative, 33 mixed methods)	1993 onward	ERIC, Education Source, Australian Education Index	Not explicitly mentioned
Anwar Rumjaun et al., 2024	Yes	Qualitative desktop research	50	2012-2022	Google Scholar, EBSCO, Emerald, Web of Science	Not explicitly mentioned
Noor Anita et al., 2021	No	PRISMA	Not mentioned	2016-2020	Scopus, Web of Science	Not mentioned

The five systematic reviews employed diverse methodological approaches to synthesize research on STEM education and outreach. Three studies explicitly utilized PRISMA guidelines , while one employed Cronbach's UTOS framework and another used qualitative desktop research methodology . The scope varied considerably, from 18 studies in the smallest review to 104 studies in the largest, with date ranges spanning from as recent as 2016-2020 to as comprehensive as 1993 onward . Database selection showed variation, with some reviews targeting education-specific databases like ERIC while others incorporated broader academic databases including Web of Science and Google Scholar . Notably, only one review explicitly employed a validated quality assessment tool (the LEBIE scale) , while two others relied on PRISMA procedures for quality assurance .

Thematic Analysis

Systematic Review Frameworks and Approaches

The reviews demonstrated three distinct methodological orientations. PRISMA guidelines emerged as the most common approach, employed in three of the five studies . One review enhanced this standard approach by applying Cronbach's UTOS framework specifically for evaluating research on undergraduate mentors in STEM , which provided a structured lens for assessing research quality beyond basic systematic review standards. This review also incorporated Nora and Crisp's (2007) conceptualization of core functional roles and Gershenfeld's (2014) key mentoring program components , demonstrating how supplementary frameworks can strengthen synthesis rigor.

The qualitative desktop research methodology employed by one review organized analysis into three sequential phases: identification and retrieval of studies, thematic analysis, and critical analysis of synthesized findings . This approach utilized the Framing Triangle theoretical framework alongside two additional frameworks to guide analysis , illustrating how multiple theoretical lenses can enhance the depth of systematic reviews focused on conceptual development rather than outcome evaluation.

Search Strategies and Inclusion Criteria

Search strategies ranged from focused to comprehensive in scope. The most targeted approach searched three education-specific databases (ERIC, Education Source, and Australian Education Index) , while the most expansive included Google Scholar, EBSCO, Emerald, and Web of Science . Inclusion criteria consistently emphasized peer-reviewed scholarly articles , though one review explicitly focused on empirical articles aligned with research questions and another specified original research .

One review implemented particularly detailed inclusion criteria, requiring studies to target pre-college participants, introduce STEM content, focus on participant evaluations, and include marginalized groups . This specificity in selection criteria demonstrates how systematic reviews can be tailored to address particular equity concerns in STEM education. The most restrictive temporal scope limited inclusion to 2016-2020 , while the most inclusive accepted studies from 1993 onward , with the broader range potentially capturing methodological evolution over time.

Quality Assessment and Methodological Rigor

Quality assessment approaches varied substantially across reviews. Only one study employed a standardized instrument - the LEBIE scale - to evaluate methodological rigor and effectiveness . Two reviews utilized PRISMA procedures for quality assessment , though the specific application of these procedures was not detailed in available abstracts. Two reviews did not explicitly mention quality assessment methods , representing a notable gap in methodological transparency.

The review employing the UTOS framework conducted independent evaluation by multiple authors to ensure consistency and address discrepancies , demonstrating a systematic approach to inter-rater reliability. Another review enhanced validity through iterative coding and independent coding by multiple researchers , ensuring robustness in thematic identification. These approaches to ensuring rigor in synthesis represent critical methodological strengths .

Data Synthesis and Analysis Methods

Synthesis approaches divided into three primary categories: thematic analysis, narrative synthesis, and framework synthesis. Thematic analysis was the most commonly reported method, explicitly used in three reviews . One review conducted iterative coding and categorization to identify common themes, followed by critical analysis for new insights , while another used thematic analysis to find themes within STEM program data connected with environmental elements .

Narrative synthesis was employed to summarize and interpret findings across multiple studies , with one review using the UTOS framework to structure this narrative approach . Framework synthesis appeared in one review that categorized studies by methodology (quantitative, qualitative, mixed methods) and integrated different data types to develop frameworks specific to demographic groups . This integration of qualitative and quantitative data through framework synthesis represents an advanced methodological approach for systematic reviews examining diverse evidence types .

The most sophisticated synthesis involved a three-level analysis framework that identified common conceptual themes, compiled narratives to produce syntheses, and critically analyzed findings to uncover new ideas . This multi-level approach demonstrates how systematic reviews can move beyond simple aggregation to generate novel conceptual insights.

Framework Development from Synthesized Data

Four of the five reviews resulted in framework development, though the processes varied in explicitness. The most detailed framework development process involved three sequential phases of analysis culminating in the STEM Education Conceptions Framework (SECF) . This framework included themes such as multidisciplinary approaches, dominance of disciplines, policies and practices, integrated STEM education, epistemological considerations, STEM for employability, and realigning assessment . The framework was supported by multiple theoretical underpinnings including the Framing Triangle, Kelley and Knowles' framework for integrated STEM education, and Ortiz-Revilla et al.'s framework focusing on epistemological, psychological, and didactical dimensions .

One review developed a framework organized into three learning categories: problem-based learning, inquiry-based learning, and project-based learning , derived from thematic analysis of STEM program data connected with environmental elements . Another review identified frameworks developed by individual primary studies rather than generating a new overarching framework, including Wade-Jaimes et al.'s framework combining communities of practice, Activity Theory, and Critical Race Theory, and Garcia et al.'s three-stage intersectional sociopolitical development framework .

The mentoring-focused review applied existing frameworks (Jacobi's four theoretical frameworks, Nora and Crisp's four major components, and Gershenfeld's LEBIE scale) rather than developing new ones , demonstrating how systematic reviews can validate and extend existing conceptual structures. One review identified three thematic categories - core competencies, instructional designs, and requisite STEM execution - though it did not explicitly develop these into a formal framework .

Handling of Qualitative and Case Study Data

The integration of qualitative data and case studies varied across reviews. One study analyzed qualitative data to identify trends and methodological rigor , focusing primarily on qualitative measurements since most included studies employed non-experimental designs with qualitative methodologies . Another review categorized case studies by construct and analyzed them separately from quantitative studies , acknowledging that qualitative and mixed methods studies were generally of higher quality and depth .

The qualitative desktop research approach systematically handled case studies through iterative coding and categorization of key ideas and concepts , with validity ensured through independent coding by multiple researchers . This iterative process allowed for both pattern identification and the emergence of novel conceptual insights . One review noted using thematic analysis but did not specify particular methods for combining case studies , while another followed PRISMA guidelines but similarly lacked detail on case study synthesis .

Synthesis

The five reviews reveal substantial methodological heterogeneity in approaches to synthesizing STEM education research, reflecting both the diversity of research questions and the maturity of systematic review methods in the field. The apparent contradiction between reviews that emphasize standardized PRISMA protocols and those employing

specialized frameworks can be explained by the distinction between reviews focused on intervention effectiveness versus those aimed at conceptual or theoretical development.

Reviews addressing ”what works” questions about specific interventions benefit from PRISMA's structured approach to study selection, bias assessment, and outcome synthesis . However, reviews examining conceptual questions about STEM education frameworks or program characteristics require more flexible, theory-driven approaches . The UTOS framework application demonstrates this principle: when the research goal is to evaluate research quality and identify methodological gaps in studies of mentoring programs , a specialized evaluative framework provides more relevant insights than generic PRISMA procedures alone.

The divergence in quality assessment rigor - ranging from validated instruments like the LEBIE scale to unspecified methods - reflects a broader tension in systematic review methodology. Reviews with full-text access to primary studies (Leavitt et al., Tran et al., and Rumjaun et al.) could potentially conduct more rigorous quality assessment than those working from abstracts alone (Hussim et al. and Anita et al.) , though this advantage was not consistently leveraged. The absence of explicit quality assessment in 40% of reviews represents a significant methodological limitation that may affect the reliability of synthesized findings.

Synthesis approaches align predictably with review scope and objectives. The largest review (104 studies) employed framework synthesis to manage heterogeneity by categorizing studies by method type and developing demographic-specific frameworks , while smaller, more focused reviews (18-50 studies) could employ intensive thematic analysis with iterative coding . This suggests a threshold effect: beyond approximately 50 studies, purely qualitative thematic approaches may become impractical without structured categorization frameworks to organize the evidence base.

The most sophisticated methodological approach emerged from the review employing three-level analysis with multiple theoretical frameworks . By moving through identification, thematic analysis, and critical analysis phases , this review achieved both comprehensive coverage and conceptual innovation - developing a new STEM Education Conceptions Framework with eight distinct themes . This contrasts with reviews that primarily aggregated existing findings without generating novel theoretical contributions . The difference likely reflects both review scope (this review spanned a decade of literature) and methodological investment in iterative synthesis .

Database selection patterns reveal strategic choices about evidence scope. Reviews targeting education-specific databases (ERIC, Education Source) prioritized pedagogical research, while those including multidisciplinary databases like Web of Science and Google Scholar captured broader interdisciplinary perspectives. The most comprehensive search strategy used four databases including Google Scholar , though this breadth came with acknowledged challenges in managing heterogeneous literature . These patterns suggest that reviews focused on educational practice benefit from targeted database selection, while those examining broader conceptualizations require multidisciplinary coverage.

The handling of qualitative data and case studies represents the most underdeveloped methodological area. Only two reviews provided explicit detail on case study synthesis approaches: categorical analysis by construct and iterative coding with independent verification . Three reviews acknowledged using primarily qualitative data but did not specify synthesis methods beyond general ”thematic analysis.” This gap is particularly significant given that most primary STEM education research employs qualitative or mixed methods designs , suggesting a critical need for more rigorous and transparent approaches to qualitative evidence synthesis in this field.

For reviews aimed at developing frameworks of best practices, the evidence suggests an optimal approach would combine PRISMA's systematic search and screening procedures with specialized theoretical frameworks for synthesis (such as UTOS) , employ validated quality assessment tools , organize analysis into sequential phases moving from description through thematic identification to critical synthesis , ensure validity through independent coding , and explicitly detail procedures for handling qualitative data and case studies . The success of the multi-framework

approach in producing a comprehensive conceptual model suggests that theoretical pluralism - applying multiple analytical lenses - may be essential for capturing the complexity of STEM education practices across diverse contexts.

References

Andrew S. Leavitt, Kari L. Nelson, and Christine E. Cutucache. "The Effect of Mentoring on Undergraduate Mentors: A Systematic Review of the Literature." *Frontiers in Education*, 2022.

Anwar Rumjaun, S. Atchia, Ravhee Bholah, and Leonard Molefe. "Broadening Perspectives of STEM Education: A New Conceptual Framework." *Journal of Research in Stem Education*, 2024.

Hairunnisa Hussim, Roslinda Rosli, and Niyazi Erdogan. "Incorporating Environmental Elements in Learning STEM: A Systematic Literature Review." *International Journal of Social Science Research*, 2024.

Jessica Tran, Jessica Wolf, Shouka Farrokh, Katherine Lyon, Robyn Newell, Jenna Usprech, Karen Cheung, and Agnes d'Entremont. "Methodologies for Evaluating the Impact of STEM Outreach on Historically Marginalized Groups in Engineering: A Systematic Literature Review (Other, Diversity)." *2024 ASEE Annual Conference &Amp; Exposition Proceedings*, n.d.

Noor Anita, R. Rosli, Azmin Sham, and L. Halim. "Mathematics Teachers' Practices of STEM Education: A Systematic Literature Review." *European Journal of Educational Research*, 2021.