Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Clinical Education in the 21st Century: A Bibliometric-Systematic Review of Transformation Models Across Health Professions

Mohammad Javad Hajivand
School of Medicine, Iran University of Medical Sciences
Omid Gheisavandi
Students' Scientific Research Center, Tehran University of Medical Sciences
Amir Hossein Dehghan
Students' Scientific Research Center, Tehran University of Medical Sciences
Yaser Davodi
yaserdavodi@proton.me

Students' Scientific Research Center, Tehran University of Medical Sciences

Systematic Review

Keywords: Clinical Competence, Clinical Clerkship, Medical Education, Nursing Education, Interprofessional Education, Curriculum, Organizational Innovation, Bibliometrics, Systematic Review

Posted Date: September 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10897781/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background: Clinical education across health professions has shifted considerably, yet evidence on transformation models remains fragmented. Existing reviews tend to examine single interventions or professions, and no prior work paired bibliometric mapping with qualitative synthesis. This review aimed to map the evidence and classify models by their underlying mechanisms.

Methods: We searched Scopus, Web of Science, and PubMed in August 2026. Records qualified if they described transformation, reform, redesign, or innovation in clinical training for any health profession at pre-licensure or postgraduate level. Conference proceedings, editorials, letters, and studies limited to preclinical or classroom teaching were excluded. Two DeepSeek-R1 variants did initial screening triage, then two authors resolved disagreements by consensus. VOSviewer was used for keyword co-occurrence and bibliographic coupling. Within each coupling cluster, papers were ranked by normalized citations and a composite bibliometric influence score; top-ranked papers were read in full. Risk of bias was not assessed because the design combined bibliometric and qualitative synthesis rather than pooled effect estimates. The review was not registered.

Results: Of 6,125 records identified, 774 met eligibility criteria; original research articles made up 89.8%. Keyword co-occurrence produced 80 terms across seven clusters (11.3% to 18.8%). Bibliographic coupling connected 414 records into six clusters (10.4% to 26.6%); 360 records lacked enough shared references but stayed in qualitative synthesis. Concordance between mapping approaches ranged from 16.0% to 45.7%. Close reading of 48 documents yielded five transformation types: continuity-based longitudinal, academic-practice partnership, instructional and technology redesign, immersive experiential, and structured transition or whole-curriculum reform. Explicit theoretical framing was absent in 18 of 48 documents (37.5%). Eleven journals supplied 30.0% of the sample; 47.7% of records had at least one US-affiliated author.

Conclusions: Clinical education transformation clusters around a small set of recurring mechanisms rather than a random collection of projects. Theory, when named, usually describes findings instead of testing causal pathways. Weak overlap between vocabulary and citation clusters signals conceptual fragmentation. Most evidence comes from high-income settings, and nearly half the records could not enter coupling analysis, so conclusions about less connected work stay tentative. Coordinated, theory-driven research across more countries would help identify which mechanisms travel.

Introduction

Clinical education anchors health professions training. It is where learners take theoretical knowledge and turn it into the judgment and skill that clinical practice demands. Supervised clinical experiences build more than technical competence: learners develop clinical reasoning and a professional identity through real encounters with patients and care teams [1–3]. This differs from classroom learning in one basic way. Clinical education happens inside a workplace, and what gets learned there depends on who else is present; learners, supervisors, patients, and the wider clinical environment shape it together [4].

Because of this, how clinical education is organized has direct consequences: it shapes whether future health professionals are ready for practice, and whether health systems can keep pace with population and healthcare needs [5, 6].

Healthcare delivery has changed substantially. Clinical work has grown more complex, technology has reshaped how care is delivered, and collaborative, patient-centered models have become the norm rather than the exception. Training approaches built for an earlier era have struggled to keep up [7]. Many of the field's long-standing models still rely on time-based progression through discipline-specific rotations, an apprenticeship structure inherited from an earlier era of medicine [8, 9]. These models produced generations of capable clinicians, but their limits have become harder to ignore: learning experiences that fragment across rotations, supervision that varies by site and by supervisor, and continuity of patient care that breaks down between placements [4, 7]. The response has not been a string of small fixes. Programs have increasingly pursued transformation efforts that rethink how clinical learning is organized and supported from the ground up [5, 7].

Transformation, in this context, means something more than adjusting a single course or session. It means redesigning the structures that hold clinical learning together [7]. Recent years have brought a range of such redesigns: longitudinal integrated clerkships, competency-based frameworks, technology-enhanced teaching, community-based placements, and full curriculum overhauls [10–14]. These approaches take different routes, but they converge on similar goals: better continuity, stronger learner support, closer ties between academic programs and clinical sites, and readiness for increasingly complex care [7, 15]. The difficulty is that these models have been tried in very different settings, evaluated in different ways, with different outcomes reported. That variation makes the literature hard to synthesize [4].

Yet the literature stays scattered. Interest in transformation keeps growing, but the evidence sits spread across professions, contexts, and research traditions that rarely engage with one another. Existing reviews tend to zoom in on one intervention or one model at a time; that gives depth but leaves the bigger picture unclear, how these approaches relate to each other across professions [4, 16]. Bibliometric work in health professions education adds another layer, mapping publication trends and thematic shifts, but rarely connecting those maps to a qualitative account of why transformation happens the way it does [17]. What is missing is a framework that sorts transformation models by the mechanism driving them and shows how different approaches solve different problems.

Closing that gap calls for a method that can hold both the shape of the field's knowledge base and the messiness of educational change on the ground. Bibliometric-systematic literature review (B-SLR) does both: it pairs quantitative mapping of the scientific landscape with systematic qualitative synthesis, making it possible to trace thematic development, intellectual connections, and conceptual patterns across a large body of work [18]. For a topic like clinical education transformation, where evidence spans professions, settings, and theoretical traditions, this combination fits better than a conventional single-method review.

This study set out to map that landscape. Specifically, it aimed to (1) identify the major thematic and citation-based structures in the literature, (2) examine how conceptual themes relate to bibliographic patterns, (3) build a taxonomy of transformation models based on their underlying mechanisms, and (4) point to priorities for future research and practice.

Methods

Study design and reporting framework

This review followed the Bibliometric-Systematic Literature Review (B-SLR) protocol, which combines bibliometric mapping with a qualitative systematic literature review to support theory development [18]. Reporting followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement, adapted for a bibliometric-systematic design [19]. This bibliometric-systematic review used publicly available bibliographic data and did not involve human participants, so institutional ethics approval was not required.

Research question and eligibility criteria

One question guided the review: what models, drivers, and outcomes characterize the transformation of clinical education across health professions programs, and what patterns emerge across disciplines? Eligibility followed three dimensions. Population: learners in any health profession with a clinical training component, at pre-licensure or postgraduate level. Concept: transformation, reform, redesign, restructuring, or innovation in clinical education; a single adjustment to one teaching session did not qualify on its own. Setting: clinical education broadly defined, spanning clerkships, clinical placements, and bedside, ambulatory, inpatient, or outpatient teaching. Eligible records included journal articles, reviews, and other relevant publication types, with no restrictions on health profession or country. Excluded were conference proceedings, editorials, letters, corrigenda, retraction notices, reports issued outside the peer-reviewed journal literature, and studies focused on preclinical or classroom-based education rather than clinical training.

Information sources and search strategy

Three databases were searched in August 2026: Scopus, Web of Science Core Collection, and PubMed (MEDLINE). Scopus served as the primary source given its journal coverage across the health and social sciences; Web of Science and PubMed supported cross-validation, following recommended practice when a single database serves as the primary source. Search strings combined three concept blocks, covering population, transformation, and clinical setting terms, joined with Boolean operators; terms were truncated with wildcards to capture morphological variants. Field scope followed each platform's native syntax: title and abstract fields in PubMed, TITLE-ABS-KEY in Scopus, and Topic (TS) in Web of

Science. The complete strategy, including the exact strings used in each database, appears in Supplementary Text A1.

Study selection

Records from the three databases were imported into Rayyan for reference management and screening [20]. Duplicates were identified through DOI matching, with title comparison applied to records lacking a DOI.Document type restrictions were applied at the screening stage rather than built into the search string itself, so that borderline formats such as short innovation reports would not be lost to an overly narrow query; ineligible formats, including conference proceedings, errata, corrigenda, retraction notices, and non-journal reports, were identified and removed at this point.

The remaining records underwent title and abstract screening using the A/B/C classification scheme described by Pittaway and colleagues, in which each record was coded as particularly relevant (A), potentially relevant (B), or of little relevance (C) [21]. Two variants of DeepSeek-R1 were used independently for preliminary classification of the full set of records against the predefined eligibility criteria. These AI-generated classifications were used as an initial screening aid rather than as final eligibility decisions. Records with discordant classifications between the two AI models, as well as records classified as potentially relevant (B) by either model, were subsequently reviewed by M.J.H. and Y.D. The final screening decisions were made by the two authors through cross-review and consensus, with disagreements resolved through discussion.

Data extraction and management

For each retained record, the following fields were extracted: title, abstract, authors, year, journal, DOI, and citation count. Records were checked for export artifacts, including character-encoding errors and column misalignment, before analysis began.

Cited-reference data required for bibliographic coupling were retrieved directly from Scopus and Web of Science, the two databases that support reference-list export; PubMed does not export cited references, so records retrieved only through PubMed were included in all other stages of the review but excluded from the bibliographic coupling analysis specifically. In Scopus, eligible records were queried through the Advanced Search interface using the DOI() field code, with the TITLE() field code used for the small number of records lacking a valid DOI; because Scopus enforces a character limit on query length, the identifier list was divided into batches of approximately 100–120 items per query. Results were exported in .CSV format with full citation information, including cited references. In Web of Science, records were queried through Advanced Search using the DO= field code, or TI= combined with publication year where no DOI was available; Web of Science permits longer query strings than Scopus, allowing larger batches per query. Results were exported in plain-text format, selecting the "Full Record and Cited References" option to ensure reference lists were included.

Bibliometric mapping

Science mapping was carried out in VOSviewer version 1.6.21 [22]. Two complementary co-occurrence techniques were applied, both using full counting. Keyword co-occurrence analysis used author keywords. Before analysis, the full candidate keyword list was inspected for abbreviations, spelling variants, and conceptually equivalent terms, and a thesaurus file was constructed to merge these variants; only conceptually justified merges were made. Bibliographic coupling, which groups documents according to the references they share, was conducted using the Scopus and Web of Science export files described above, restricted to records for which reference data were available [23]. For both techniques, the resolution parameter began at VOSviewer's default setting and was adjusted iteratively to balance thematic coherence against fragmentation. Minimum cluster size followed the threshold recommended by Marzi and colleagues, set at approximately 10% of the total items included in each network. The minimum cluster size parameter in VOSviewer was set to correspond to this threshold when generating the networks. For bibliographic coupling, documents that were not mutually connected through shared references were excluded from the coupling map but retained in the qualitative synthesis. For bibliographic coupling, documents that were not mutually connected through shared references were excluded from the coupling map but retained in the qualitative synthesis.

Sample ordering and selection

Within each bibliographic coupling cluster, documents were ranked by normalized citation count, which adjusts raw citation figures for publication year so that older studies are not favored by default. Where a finer-grained ranking was needed, records were additionally ranked using a composite bibliometric influence score (CBIS), computed by min-max scaling normalized citations, total links, and total link strength to a common range and combining the three scaled values multiplicatively, following the procedure described by Marzi and colleagues. Scaling was performed within each cluster rather than across the full sample, so that clusters with generally higher citation levels did not disproportionately inflate scores for their member documents. For each cluster, a minimum of the top 10% of ranked documents, with an absolute floor of seven documents per cluster, was selected for close reading during synthesis; abstracts of documents outside this subsample were re-read as a check, beyond the initial title-and-abstract screening at the inclusion/exclusion stage, to confirm that their exclusion from close reading did not compromise the overall quality of the review.

Systematic synthesis and theory development

The qualitative synthesis paired a holistic analysis of the dataset as a whole with a cluster-specific analysis of each identified theme, following the approach set out by Petticrew and Roberts [24]. Studies within each cluster were tabulated by research question, theoretical framing, methodology, and key findings, then synthesized narratively to surface patterns, discrepancies, and gaps. Findings across clusters were compared to identify drivers of transformation that cut across disciplines and points

where the disciplines diverged. Following Torraco's framework for literature-review synthesis, the resulting theoretical contribution drew on two synthesis types: an integrative synthesis, classifying the models of clinical education transformation identified across clusters into a taxonomy, and an agenda-setting synthesis, outlining directions for future inquiry based on gaps and mixed findings surfaced during the review [25].

Results

Study selection

A total of 6,125 records were identified across the three databases: PubMed (n = 1,704), Scopus (n = 2,545), and Web of Science Core Collection (n = 1,876). Removing 3,132 duplicates left 2,993 records for title and abstract screening, and that step excluded a further 2,219. The final sample held 774 records (Fig. 1).

Sample characteristics

The included studies span 2000 to 2026, and output rose sharply across that window. Annual counts sat in the single digits early on, seven records in 2000, then climbed past fifty a year from 2021 onward (61 in 2021, 53 in 2022, 52 in 2023, 57 in 2024, 66 in 2025). Nearly half the sample, 342 of 774 records, appeared in just the last five years, a pattern consistent with the growing attention to clinical education transformation documented elsewhere in this review.

Records came from 314 distinct journals. A small cluster of titles accounted for a disproportionate share: eleven journals together held 30% of the sample, led by *Academic Medicine* (n = 49) and *BMC Medical Education* (n = 46), with *Medical Teacher* (n = 24) and *Clinical Teacher* (n = 20) some distance behind (Table 1). The remaining 542 records were spread across 303 other titles, mostly specialty and regional journals.

Table 1
Characteristics of included studies (n = 774)

Category	n (%)
Document type	
Original research article	695 (89.8%)
Review	50 (6.5%)
Other (editorial/letter, book chapter, conference paper/abstract, note, short survey)	29 (3.7%)
Language	
English	739 (95.5%)
Chinese	15 (1.9%)
German	11 (1.4%)
Other (Spanish, Norwegian, Korean, Japanese, Hebrew, Portuguese)	9 (1.2%)
Top 10 contributing journals	
Academic Medicine	49
BMC Medical Education	46
Medical Teacher	24
Clinical Teacher	20
Medical Education Online	15
Nurse Education in Practice	15
Teaching and Learning in Medicine	13
Journal of Nursing Education	13
Journal of Dental Education	13
Medical Education	12
Advances in Medical Education and Practice	12
All other journals (303 titles)	542 (70.0%)
Top 10 contributing countries (full counting, all author affiliations)	
United States	369 (47.7%)

Note: Geographic counts under full counting reflect all author affiliations per record and sum to more than 774 for records with multi-country author teams. Full journal and country breakdowns for all 314 journals and 69 (full counting) / 55 (first-author) countries are provided in Supplementary Data D1.

Category	n (%)
Document type	
Australia	67 (8.7%)
China	59 (7.6%)
United Kingdom	54 (7.0%)
Canada	49 (6.3%)
South Africa	25 (3.2%)
Germany	19 (2.5%)
Netherlands	15 (1.9%)
Iran	15 (1.9%)
Japan	11 (1.4%)
Top 4 contributing countries (first-author affiliation only)	
United States	40.6%
Australia	6.1%
United Kingdom	5.8%
Canada	5.2%
Note: Geographic counts under full counting reflect all author affiliations per record and sum to more than 774 for records with multi-country author teams. Full journal and country breakdowns for all 314 journals and 69 (full counting) / 55 (first-author) countries are provided in Supplementary Data D1.	

Original research articles made up the large majority of the sample (n = 695, 89.8%), with reviews a distant second (n = 50, 6.5%); editorials, letters, book chapters, conference material, notes, and short surveys together filled the remaining gap. English dominated the language distribution (n = 739, 95.5%), and Chinese and German were the only other languages to clear ten records each.

Contributing institutions spanned 69 countries (Fig. 2). Counted by all author affiliations, the United States alone accounted for nearly half the sample (n = 369, 47.7%), well ahead of Australia (n = 67, 8.7%), China (n = 59, 7.6%), the United Kingdom (n = 54, 7.0%), and Canada (n = 49, 6.3%); South Africa, Germany, the Netherlands, Iran, and Japan rounded out the top ten, none exceeding 4% individually (Table 1). Restricting the count to first-author affiliation only softened this concentration somewhat but did not remove it: the United States still led at 40.6%, followed by Australia, the United Kingdom, and Canada in single digits each. Even so, 55 different countries appeared as a first-author location at least once, so while output concentrates at the top, it is not confined to a narrow band of high-income settings.

Keyword co-occurrence

Thesaurus merging and a minimum occurrence threshold of six left eighty author keywords for analysis, forming seven clusters. Cluster sizes ranged from nine terms (11.3%) to fifteen (18.8%), and every cluster cleared the 10% floor recommended by Marzi and colleagues [18]. *Medical education* appeared most often (n = 88), followed by *medical student* (n = 50), *curriculum* (n = 46), *undergraduate medical education* (n = 38), and three terms tied at thirty-six apiece: *clinical education*, *nursing student*, and *covid-19*. *Medical education* also carried the strongest total link strength (133), a pattern consistent with its role as a bridge across clusters rather than a term confined to one theme.

The seven clusters spanned competency-oriented clinical skills training paired with active learning methods; technology-mediated clinical teaching accelerated by the pandemic; nursing clinical education built around placement-based partnerships; interprofessional education alongside graduate and system-level training; undergraduate medical curriculum and clerkship experiences; reform-oriented models built on longitudinal and community-based approaches; and innovative formats centered on blended and flipped instruction. Table 2 lists each cluster with its representative terms; the full term-to-cluster mapping, together with the underlying VOSviewer map and thesaurus files, appears in Supplementary File S1.

Table 2
Keyword co-occurrence clusters and representative terms

Cluster	Label	Representative author keywords
1	Competency-oriented clinical skills and active learning methods	clinical competence; problem-based learning; clinical skills; case-based learning; teaching method; physical examination
2	Technology-mediated clinical teaching and pandemic-driven innovation	covid-19; clinical teaching; learning; innovation; simulation; e-learning
3	Nursing clinical education and placement-based learning partnerships	clinical education; nursing student; nursing education; clinical placement; nursing; academic-practice partnerships
4	Interprofessional education and graduate/system-level medical training	interprofessional education; assessment; graduate medical education; teaching; residency; competency-based medical education
5	Undergraduate medical curriculum and clerkship experiences	medical student; curriculum; education; clerkship; primary care; competency
6	Reform-oriented medical education and longitudinal or community-based models	medical education; longitudinal integrated clerkship; undergraduate; curriculum reform; program evaluation; community-based education
7	Innovative clinical training formats and blended learning	undergraduate medical education; clinical clerkship; clinical training; dental education; flipped classroom; blended learning

The overlay visualization, in which node color reflects the average publication year of each keyword's occurrences (2016, blue, to 2022+, red; Fig. 3b), showed a clear temporal stratification layered on top of this thematic structure. The oldest, most established terms clustered toward the network's structural core: evaluation (mean year 2012.7), physical examination (2013.4), problem-based learning and undergraduate medical education (2017.0 and 2017.2, respectively), and curriculum / teaching method (2016.5 each), reflecting the foundational, long-running nature of curriculum-level and problem-based-learning discourse in this literature. Interprofessional education (2018.4) sat closer to the established side of the network. Competency-based education (2020.8), by contrast, occupied a middle position, newer than the foundational curriculum-reform terms but still measurably older than the pandemic-era cluster described below.

The most recent terms concentrated in two areas of the network: a technology/pedagogy-innovation region (clinical reasoning, mean 2023.9; blended learning, 2023.4; telemedicine, 2023.4; self-efficacy, 2023.3; experiential learning, 2023.0) and an emerging cross-cutting region anchored by health professions education (2024.2), the single most recent keyword in the network. Several COVID-era-adjacent terms, online learning (2022.9), case-based learning (2022.9), and flipped classroom (2021.9), also skewed recent, consistent with the acceleration in remote/blended clinical teaching literature noted in the sample's publication-year distribution above. Covid-19 itself (mean year 2021.9) sat at the visual center of this recent shift, bridging the older curriculum-reform core and the newer pedagogy-innovation periphery, suggesting the pandemic functioned as a connective driver between established curricular themes and newer instructional-technology themes rather than constituting an isolated topic.

Bibliographic coupling

The coupling network held 414 connected documents across six clusters, with a minimum cluster size of forty, close to 10% of the connected sample; all six cleared that mark. Cluster sizes ran from 110 documents (26.6%) down to 43 (10.4%). Of the 774 records screened overall, 360 (46.5%) lacked mutual reference links dense enough to enter this map and were excluded from it while remaining in the qualitative synthesis (Supplementary File S2). Figure 4a shows the network by cluster; Table 3 lists cluster composition and representative documents.

Table 3
Bibliographic coupling clusters: composition and representative documents (N = 414 connected documents)

Cluster	n (%)	Label	Representative documents (year) [NC]
1	110 (26.6%)	Longitudinal integrated clerkships and rural or community-based longitudinal placements	Norris et al. (2009) [7.03]; Poncelet et al. (2011) [4.69]; Ogur and Hirsh (2009) [3.67]; Hudson et al. (2012) [2.29]; Padiko et al. (2025) [1.00]
2	80 (19.3%)	Technology-enhanced and active learning formats in clinical teaching	Belfi et al. (2015) [6.80]; O'Connor et al. (2016) [5.43]; Lin et al. (2017) [4.78]; Coates et al. (2008) [2.88]; Keeley et al. (2019) [1.01]
3	71 (17.1%)	Nursing clinical education models, dedicated education units, and academic-practice partnerships	Rhodes et al. (2012) [2.52]; Mulready-Shick et al. (2013) [2.22]; Johnsson et al. (2023) [1.75]; Plemmons (2018) [1.40]; Masters et al. (2016) [1.27]
4	62 (15.0%)	Interprofessional, system-based, and community-oriented clinical learning models	Samuriwo et al. (2020) [3.31]; Edwards and Nash (2023) [1.97]; Taoube (2023) [1.75]; Ottenheijm (2008) [1.27]; Miller-Rosser (2019) [0.19]
5	48 (11.6%)	Curricular and clerkship redesign in specialty and traditional clinical rotations	Ogur et al. (2007) [5.75]; Gay et al. (2013) [3.53]; Patterson et al. (2016) [1.81]; Prideaux et al. (2007) [1.57]; Botha (2023) [0.44]
6	43 (10.4%)	Transition-to-clerkship, integrated curricula, and system-level transformation	Atherley et al. (2019) [8.40]; Müller et al. (2025) [3.00]; Talib (2017) [1.44]; Taylor (2014) [0.77]; Austin (2020) [0.58]
Total	**414 (100.0%)**	—	—

Table 4
Bibliographic coupling clusters: size and temporal profile (N = 414 connected documents)

Cluster (theme)	n (%)	Mean year (range)
1. Longitudinal integrated clerkships and rural or community-based longitudinal placements	110 (26.6%)	2018.8 (2000–2026)
2. Technology-enhanced and active learning formats in clinical teaching	80 (19.3%)	2015.2 (2001–2026)
3. Nursing clinical education models, dedicated education units, and academic-practice partnerships	71 (17.1%)	2017.6 (2000–2026)
4. Interprofessional, system-based, and community-oriented clinical learning models	62 (15.0%)	2017.7 (2002–2026)
5. Curricular and clerkship redesign in specialty and traditional clinical rotations	48 (11.6%)	2015.8 (2005–2026)
6. Transition-to-clerkship, integrated curricula, and system-level transformation	43 (10.4%)	2016.6 (2003–2026)
Total in coupling network	**414 (100.0%)**	—
Screened records excluded from map (no mutual reference links)	360 (46.5% of 774)	—

Note. Cluster sizes and themes derive from VOSviewer bibliographic coupling (minimum cluster size = 40, ≈ 10% of the connected network). Percentages in Table 4 are of the 414-document connected network, not the full 774-record sample.

Within each cluster, documents were ranked first by normalized citations, then, where finer distinctions mattered, by a composite bibliometric influence score built from normalized citations, total links, and total link strength, each scaled to a common range and multiplied together following Marzi and colleagues' procedure [18]. This kept early, heavily cited work from crowding out newer papers that had not yet had time to accumulate citations. The top 10% of each cluster, with a floor of seven documents, went forward for close reading; abstracts of the rest were reread as a check rather than skipped outright.

Cluster 1: longitudinal integrated clerkships and rural or community-based placements. This cluster held the largest body of work, and continuity ran through nearly all of it as the organizing mechanism. Norris and colleagues [16] documented fifteen active programs across four countries, built around a median forty-week placement with comparable exam performance and an apparent pull toward primary care. Poncelet and colleagues [26] and Ogur and Hirsh [27] traced how sustained contact with the same patients-built relationship-based learning, deepened responsibility, and occasionally produced role confusion for students still learning where their boundaries sat. Several studies grounded this in explicit theory: Hudson and colleagues [28] drew on situated learning and communities of practice, Nielsen and colleagues [29] applied Benner's novice-to-expert progression and Tanner's clinical judgment model, and Padiko and colleagues [30] combined Bandura's social cognitive theory with Kolb and transformative

learning to explain how early community exposure builds self-efficacy. Latessa and colleagues [31] reported that longitudinal students outperformed matched traditional students on Step 2 and most shelf exams and leaned further toward primary care. Mazotti and colleagues [32], reading growth through Rogers' diffusion of innovation, argued that the roughly a third of programs offering six months or more of continuity signaled the field approaching a critical mass rather than a niche experiment.

Cluster 2: technology-enhanced and active learning formats. Flipped classrooms, blended designs, and structured active learning defined this cluster. O'Connor and colleagues [33] found a flipped format produced greater post-test gains and higher task value with less reported boredom, and Belfi and colleagues [34] reported a significant pre-post improvement under a blended design, with the interactive classroom component the most preferred element. Lin and colleagues [35], working from Bloom's taxonomy, found no difference in exam scores but higher satisfaction, and noted the approach worked better for ocular trauma than for glaucoma content specifically. Morgan and colleagues [36] applied Mayer's cognitive theory of multimedia learning to cut didactic time from three hours to one without moving NBME scores, and Keeley and colleagues [37] used competency-based frameworks, Core EPAs, and Dreyfus's model to structure entrustment decisions and individualized fourth-year planning. Not every redesign landed cleanly: Patel and colleagues [38] found that full horizontal integration was experienced as disintegration until the curriculum was reorganized into seven-week blocks, after which ratings improved. Coates and colleagues [39] linked a redesigned format to a greater likelihood of identifying a mentor or role model, alongside higher satisfaction overall.

Cluster 3: nursing clinical education models, dedicated education units, and academic-practice partnerships. Dedicated education units and formal academic-practice partnerships dominated this cluster. Mulready-Shick and colleagues [40], applying a QSEN measurement framework, found dedicated-education-unit students scored higher on every item of a standardized clinical education evaluation, though grades and exam scores did not differ from traditional placements. Masters and colleagues [41] reported similarly higher QSEN scores under this model, alongside a theme of evolving confidence. Rhodes and colleagues [42], framing their work through Benner and Tanner, described high satisfaction and a shift in faculty role toward coordination rather than direct instruction. Several studies leaned on situated and social learning traditions: Christiansen and Bell [43] applied social constructionism and communities of practice to show how peer support among students reduced isolation and anxiety, Johnsson and colleagues [44] used Vygotsky and work-integrated learning to trace scaffolding that faded as independence grew, and Plemmons and colleagues [45] drew on Bandura's social learning theory to show self-efficacy gains across models, strongest under dedicated-unit and blended arrangements. Vosoughi and colleagues [46] reported gains across clinical learning environment, nurse job satisfaction, and patient satisfaction simultaneously, without a named theoretical frame.

Cluster 4: interprofessional, system-based, and community-oriented clinical learning. This cluster centered on preparing students for collaborative practice and system-level understanding rather than a single skill or rotation. Samuriwo and colleagues [47], working from an interpretivist and constructivist stance informed by GMC standards, found that an integrated curriculum redesign had cost some

programs their nursing placements, closing off interprofessional learning opportunities that earlier structures had provided. Taoube and colleagues [14] went further theoretically, combining systems thinking with a 3P-6C curriculum design framework and communities of practice to build an integrated task-based micro-curriculum, identifying specific tension points between structure and student agency along the way. Edwards and Nash [48] and Ottenheijm and colleagues [49] both drew on Kolb's experiential cycle, the former reporting a shift in perspective and new advocacy motivation tied to disability-focused placements, the latter finding that giving students an active, independent role with patients was feasible and well received in most participating practices. Head and colleagues [50], also working from Kolb, described intimate emotional learning experiences tied to palliative care placements that addressed several ACGME competencies at once. Miller-Rosser and colleagues [51] reported broadened understanding of chronic illness and increased confidence without a named framework, and Gray and colleagues [52] found that adding situational scoring and moderation to assessment increased transparency, though shortened placements limited the evidence students could generate.

Cluster 5: curricular and clerkship redesign in specialty and traditional rotations. Psychiatry, surgery, internal medicine, and other established rotations made up this cluster's focus. Ogur and colleagues [53] described a redesign that held clinical performance steady or improved it while raising patient-centeredness and faculty satisfaction. Prideaux and colleagues [54] framed their approach through the PRISMS model and an explicitly symbiotic, relationship-centered curriculum, identifying four key relationships with the student positioned at the center and reciprocity treated as a requirement rather than a byproduct. Botha and colleagues [55], applying a formal framework for distributed health professions training, found rural placements more likely to build independence and procedural confidence, with most students preferring the distributed format once they had experienced it. Patterson and colleagues [56] used Dewey's and Gibbs' reflective models to show reduced stigma-related anxiety and stronger recovery-oriented skills following a psychiatry redesign, and Gay and colleagues [57], drawing on dual-process theory and evidence-based medicine, reported enthusiasm for a hypothetico-deductive approach to clinical reasoning. Negrete Manriquez and colleagues [58] applied spacing and testing-effect principles alongside a flipped design, finding that a virtual format eased some stressors and improved flexibility without sacrificing engagement. Jacobs and colleagues [59] found daily practice logs valued in principle but underused in practice, a reminder that not every structural addition gets taken up as intended.

Cluster 6: transition to clerkship and system-level curriculum change. Atherley and colleagues [60], working from Mezirow's transformative learning theory, reviewed the transition-to-clerkship literature and identified three lenses at play: an educational perspective that dominated the field, a social perspective, and a developmental perspective that remained comparatively underrepresented. Müller and colleagues [61] extended this theoretical line, combining Mezirow with communities of practice and mentalization to argue that transformative learning can curdle into what they call transformative trauma without deliberate facilitation, and proposed an explicit pedagogy to guard against it. Talib [62], building on Worley's symbiotic medical education model, reported reduced workload and improved staff satisfaction alongside perceived gains in care quality. Tran and colleagues [63] documented a large-scale shift to an

integrated, competency-based curriculum informed by the Lancet Commission, Kotter's change model, RIME, Miller's pyramid, and Bloom's taxonomy together, improving the faculty-to-student ratio from roughly 1:12–15 down to 1:5. Austin and colleagues [64], grounded in AAMC Core EPAs and deliberate practice, found significant improvement on twenty-one of twenty-two training objectives, with hands-on, low-stakes practice valued especially highly. Taylor and colleagues [65] found that peer teachers were rated highly and associated with better subsequent OSCE performance, while Bracken and colleagues [66], applying spiral integration and adult learning theory, found no difference on exit exams but a significant within-rotation improvement on multiple-choice assessment, evidence read as a sign the change was at minimum not doing harm.

Average publication year varied meaningfully across the six clusters, indicating differing stages of maturity across the transformation models identified. The technology-enhanced and active learning cluster (Cluster 2) and the specialty and traditional-rotation redesign cluster (Cluster 5) had the oldest average publication years (2015.2 and 2015.8, respectively), consistent with roots in earlier foundational work. The longitudinal integrated clerkship cluster (Cluster 1), despite containing some of the field's earliest continuity-based models, had a comparatively recent mean (2018.8) and the widest year range of any cluster (2000–2026), suggesting LICs remain a sustained, continuously active stream rather than a historical topic that has plateaued. Across the full 414-document network, publications from 2023 onward and from 2015–2019 each formed the largest bands (n = 106 apiece, 25.6%), followed by 2020–2022 (n = 70, 16.9%), 2010–2014 (n = 71, 17.1%), and pre-2010 work (n = 61, 14.7%), a pattern consistent with the acceleration already observed across the full 774-record sample.

Convergence between keyword co-occurrence and bibliographic coupling clusters

Each of the 414 bibliographic-coupling documents was matched to its Scopus record and its raw author keywords normalized through the same thesaurus used for the keyword-cluster analysis, then checked against the finalized 80-term list. A document's matched keywords each cast one vote for the keyword cluster they belonged to, and the cluster with the most votes became that document's dominant keyword cluster; where two clusters tied for the most votes, the document was recorded as tied between them. Coverage was partial: 277 of 414 documents (66.9%) carried any recorded author keywords, and 247 (59.7%) had at least one keyword matching the finalized list; the shortfall traces mainly to Web of Science and PubMed records lacking the relevant keyword field.

Concordance between the two mapping approaches was weak to moderate across all six bibliographic-coupling clusters, with no case reaching majority agreement. Cluster 3 (nursing clinical education models, dedicated education units, and academic-practice partnerships) came closest to a single dominant theme, with 45.7% of its matched documents aligning to the nursing clinical education keyword cluster. Cluster 4 showed a similar, if weaker, pull toward the same keyword cluster (31.6%). The remaining four bibliographic-coupling clusters split their matched documents across keyword clusters with no group exceeding a quarter of the total (Table 5).

Table 5
Convergence between keyword co-occurrence and bibliographic coupling clusters

BC Cluster	Label	n (matched)	Dominant KW Cluster	Concordance
1	Longitudinal integrated clerkships and rural or community-based longitudinal placements	63	Reform-oriented medical education and longitudinal or community-based models	20.6%
2	Technology-enhanced and active learning formats in clinical teaching	50	Reform-oriented medical education and longitudinal or community-based models	16.0%
3	Nursing clinical education models, dedicated education units, and academic-practice partnerships	46	Nursing clinical education and placement-based learning partnerships	45.7%
4	Interprofessional, system-based, and community-oriented clinical learning models	38	Nursing clinical education and placement-based learning partnerships	31.6%
5	Curricular and clerkship redesign in specialty and traditional clinical rotations	25	Undergraduate medical curriculum and clerkship experiences	24.0%
6	Transition-to-clerkship, integrated curricula, and system-level transformation	25	Undergraduate medical curriculum and clerkship experiences	24.0%

Note. Concordance reflects the share of a bibliographic-coupling cluster's matched documents (author keywords available and mapped to the 80-term list) that voted for the cluster's dominant keyword-cluster. n = number of documents in the bibliographic-coupling cluster with at least one keyword matching the finalized keyword-cluster list, out of 247 total matched documents across the network.

Integrative synthesis: a taxonomy of transformation models

Following Torraco's framework for literature-review synthesis, the 48 closely read documents were classified across the six bibliographic-coupling clusters into five cross-cutting model types. These types are not mutually exclusive with the bibliometric clusters; several documents combine elements of more than one, but each represents a distinct mechanism by which clinical education is restructured.

Continuity-based longitudinal models restructure clinical training around sustained relationships, with patients, preceptors, and peers, across an extended period rather than short block rotations. The longitudinal integrated clerkship literature anchors this type (Cluster 1: Norris et al. 2009; Ogur and Hirsh 2009; Poncelet et al. 2011; Hudson et al. 2012; Latessa et al. 2015), extending into full-year pilots (Ogur

et al. 2007, Cluster 5) and preclinical longitudinal community exposure (Padiko et al. 2025, Cluster 1). The common driver across these studies is loss of continuity and fragmentation in traditional block rotations; common outcomes include comparable-to-superior exam performance, stronger patient-centeredness, and identity or empathy formation.

Academic-practice and workforce-embedded partnership models restructure who teaches and where, shifting primary clinical instruction toward frontline practitioners, staff nurses, or community clinicians, or toward decentralized and rural sites. Dedicated education units exemplify this type (Cluster 3: Masters et al. 2016; Rhodes et al. 2012; Mulready-Shick et al. 2013), alongside decentralized and distributed training models (Talib 2017 and Müller et al. 2025, Cluster 6; Botha et al. 2023, Cluster 5). Faculty and preceptor shortage, paired with capacity strain at tertiary sites, drives this restructuring; outcomes run from comparable-to-superior clinical learning and service-delivery benefits for host sites to workforce and rural-practice pipeline effects, tempered by the ethical risk Müller and colleagues [61] raise under the label of transformative trauma when support structures fall short.

Instructional and technology redesign models restructure how content gets delivered within an otherwise conventional clerkship structure, primarily through flipped classrooms, blended or online delivery, and simulation. Cluster 2 (O'Connor et al. 2016; Belfi et al. 2015; Lin et al. 2017; Morgan et al. 2015; Shahkarami et al. 2025) anchors this type, joined by technology-based transition courses (Austin et al. 2020; Bracken et al. 2021, Cluster 6). Passive, time-inefficient traditional didactics, accelerated by pandemic-era necessity, drive this type; outcomes generally improve engagement, task value, and knowledge gains without loss of exam performance, though the gains are pedagogy-specific rather than uniform, Patel and colleagues [38] found integration could backfire when disciplines were too dissimilar to combine cleanly.

Immersive and specialized experiential models place students into non-traditional, often single-population or single-context clinical experiences designed to build a specific competency or attitude. Cluster 4 supplies most examples (Edwards and Nash 2023 on disability care; Head et al. 2012 on palliative care; Miller-Rosser et al. 2019's chronic-illness camp model; Taoube 2023's community-based micro-curriculum), joined by Patterson and colleagues' [56] recovery-oriented placement in Cluster 5. These models respond to an identified gap in exposure to a specific patient population or care context, and their outcomes, attitude and perspective shift, reduced stigma or discomfort, and greater advocacy motivation, are evaluated almost exclusively through qualitative or reflective methods rather than performance metrics.

Structured transition and whole-curriculum reform models operate at a broader structural level, either bridging a specific transition point (Taylor et al. 2014, Cluster 6; Jacobs et al. 2005, Cluster 5) or reforming an entire program's curriculum architecture around competency-based principles (Dienstag 2011, Cluster 1; Tran et al. 2022, Cluster 6). Systemic misalignment between preparation and clinical demands, or pressure from national and accreditation standards, drives this type; outcomes include

improved readiness and confidence at the individual level and, for whole-curriculum reforms specifically, institutional-level gains in governance, accreditation standing, and faculty development.

Two patterns cut across the taxonomy. Continuity-based and academic-practice partnership models, taken together, account for the largest share of the 48-document subsample (11 plus 8 of 48, close to 40%), suggesting these two mechanisms are the field's most substantively developed transformation strategies rather than simply its most frequently coupled. Evaluation method also tracks model type rather than following a single evidence hierarchy: continuity-based and instructional-redesign models lean on comparative quantitative designs (exam scores, matched cohorts), while workforce-embedded and immersive models rely overwhelmingly on qualitative or reflective designs. That split likely reflects what is feasible to measure, exam performance, against what these models are actually built to produce, relational and identity outcomes, rather than a hierarchy of evidence quality, a distinction worth carrying into the Discussion.

Theoretical framing was genuinely absent in 18 of the 48 closely read documents (37.5%), including several representing mature, multiply replicated models such as longitudinal integrated clerkships and dedicated education units with a decade or more of accumulated implementation experience. Where a framework did appear, the range extended well beyond a narrow default: Kolb's experiential learning cycle and Lave and Wenger's communities of practice recurred most often, but so did Bandura's social cognitive theory, Benner's novice-to-expert progression, Bloom's taxonomy, Mezirow's transformative learning, QSEN, and several discipline-specific frameworks such as Worley's symbiotic model and the PRISMS curriculum model.

Agenda-setting synthesis: directions for future inquiry

Following Torraco's framework, this agenda-setting synthesis identifies gaps and unresolved tensions surfaced across the bibliometric mapping, the 48 closely read documents, and the taxonomy above, and translates them into concrete directions for future research.

Economic evidence is almost entirely absent. Cost and cost-effectiveness went unmeasured or unreported across multiple model types, continuity-based and workforce-embedded alike, and no document in the 48-document subsample reported a formal cost-effectiveness or cost-per-outcome analysis. Future research should prioritize economic evaluation alongside educational outcomes, particularly for continuity-based and academic-practice partnership models, whose resource intensity, faculty time, site coordination, extended placements, is repeatedly noted as a barrier to scale-up but rarely quantified.

Downstream and patient-level outcomes are rarely tracked. Most evaluations stop at student knowledge, skill, or self-reported attitude. A small minority extended to patient-reported outcomes (Vosoughi et al. [46] reported gains in patient satisfaction alongside clinical learning environment and nurse job satisfaction) or to downstream practice patterns (Latessa and colleagues' [31] specialty-choice comparison; Norris and colleagues' [16] tentative signal toward primary care). Future research should

prioritize longitudinal tracking of graduates from these models into residency selection, practice location, and practice patterns, paired with systematic inclusion of patient-level outcomes rather than learner-level outcomes alone.

Evaluation method tracks model type rather than following a single evidence hierarchy. Continuity-based and instructional-redesign models lean on comparative quantitative designs, while workforce-embedded and immersive models rely almost entirely on qualitative or reflective designs, often at a single institution with volunteer or convenience. This is not necessarily a quality gap, but it does mean the two evidence bases are not directly comparable, and neither model family has been triangulated against the other's method type. Future studies should apply mixed-methods or comparative quantitative designs to workforce-embedded and immersive models specifically, testing whether their qualitative benefits, identity formation, empowerment, reduced stigma, correspond to measurable downstream effects.

Theoretical grounding lags implementation. Eighteen of the 48 closely read documents (37.5%) stated no explicit theoretical framework, including several representing mature, multiply replicated models, longitudinal integrated clerkships and dedicated education units among them, with a decade or more of accumulated implementation experience. Where a framework did appear, it drew on a genuinely varied set, Kolb's experiential learning and Lave and Wenger's communities of practice recurred most, but so did Bandura's social cognitive theory, Benner's novice-to-expert model, Bloom's taxonomy, and several discipline-specific frameworks. Future work should explicitly test competing theoretical accounts of why continuity- and partnership-based models work, rather than continuing to describe outcomes atheoretically or lean on a small set of default frameworks without comparison.

Weak convergence between citation-based and vocabulary-based groupings signals conceptual fragmentation. The convergence check between keyword clusters and bibliographic-coupling clusters found one comparatively stronger alignment, Cluster 3 (nursing clinical education models, dedicated education units, and academic-practice partnerships), at 45.7% concordance; every other cluster showed weaker, more dispersed concordance, from 16.0% to 31.6%. This indicates that documents built on the same citation lineage do not consistently converge on shared terminology, different research traditions within clinical education transformation are not citing across a shared vocabulary. A future synthesis or Delphi-type consensus effort could usefully work toward shared terminology for describing transformation models, which would also make future bibliometric mapping of this field more interpretable.

Ethical and psychological-safety dimensions of workforce-embedded and distributed models are under-examined. Müller and colleagues [61] are alone in the subsample in explicitly interrogating the risk that transformative-learning-oriented distributed placements can curdle into what they term transformative trauma absent adequate facilitation and support, a risk with direct relevance to Cluster 6 (transition-to-clerkship and system-level transformation) and, by extension, to any model that removes learners from familiar institutional support structures, rural longitudinal placements and community immersion experiences included. Future research should systematically evaluate the support infrastructure,

supervision capacity, emotional and psychological support, equitable resource access, accompanying distributed and immersive placements, not only their educational outcomes.

Geographic and resource-setting representation remains concentrated. Consistent with the sample-level geographic findings (Table 1), the 48 closely read documents draw overwhelmingly from high-income, English-language settings, the United States, Australia, the United Kingdom, and Canada, with only a small number of studies from low- and middle-income contexts: Tran and colleagues [63] in Vietnam, Botha and colleagues [55] in South Africa, Talib [62] in sub-Saharan Africa, and Shahkarami and colleagues [67] in Iran. These studies disproportionately represent the workforce-embedded and distributed model type, suggesting this family is comparatively better studied outside high-income contexts than within it, while the reverse holds for continuity-based and instructional-redesign models. Future research should deliberately test high-income-context models, longitudinal integrated clerkships and flipped or blended instructional redesign, in low- and middle-income settings, and evaluate workforce-embedded and distributed models more thoroughly in high-income contexts, to determine which mechanisms are context-dependent versus generalizable.

Discussion

Principal Findings

This review of 774 records confirms that clinical education transformation has moved from a peripheral concern to a central, structural priority across health professions. The dominant story emerging from both the bibliometric synthesis and the close reading is a shift away from static, time-based curricula toward structured, multi-component models that intentionally reorganize the relationship between learner, supervisor, and clinical environment. Across clusters and model types, the shared logic is one of integration: integrating experiences over time, integrating academic and practice settings, and integrating technology into the pedagogical core. The agenda-setting gaps derived from this evidence base are therefore not about a lack of activity, but about a lack of cumulative, theory-driven learning from that activity.

The primary theoretical contribution of this review is the five-model taxonomy presented in the Results, developed through close reading of 48 documents selected from the six bibliographic coupling clusters using the CBIS-based ranking procedure described in the Methods. These model types, spanning continuity-based longitudinal approaches to structured transition and whole-curriculum reform, offer a functional classification system for a heterogeneous field. Prior work often splits clinical education into binaries such as hospital versus community or traditional versus innovative. This taxonomy instead organizes the literature by the mechanism of transformation a program claims to enact, whether that mechanism is temporal continuity, partnership structure, technological redesign, immersive experience, or systemic reform. This mechanism-focused classification is important because it aligns descriptions of interventions with the logic of their intended effects, creating a platform for more meaningful cross-study comparison than topic-based labels allow. A program described as a longitudinal integrated

clerkship and one described as an academic-practice partnership may share substantial operational overlap, but the taxonomy differentiates them by their primary lever of change, which is a necessary step for understanding what works, for whom, and why.

A secondary but equally important finding concerns the state of theory in this literature. As reported in the Results, a majority (62.5%) of closely read documents did name an explicit framework, but theory usage is real, shallow, and under-tested. Most papers used frameworks descriptively, as a vocabulary for organizing findings, rather than to explain causal mechanisms or predict transferability of a model to a new setting. The most common grounding was in experiential learning or adult learning theory, yet few studies returned to these theories in their analysis or discussion to interpret results in light of theoretical propositions. This pattern, combined with the geographic concentration (the United States alone accounted for 47.7% of all author affiliations, meaning 47.7% of records had at least one US-affiliated author), suggests a field that is producing abundant local innovation but has not yet developed a shared, portable knowledge base. The temporal trend toward multi-site and partnership models indicates a growing recognition of this limitation, but the theoretical infrastructure has not caught up with the practical ambition.

Comparison With Prior Work

Positioned against the seven most relevant prior reviews and frameworks, this review confirms shared structural observations while extending the field in two distinct ways: methodologically, through a convergence check that quantifies fragmentation, and conceptually, through a mechanism-based taxonomy that, to our knowledge, has not been explicitly applied across clinical education transformation models.

The first point of comparison is with three recent bibliometric analyses of adjacent fields. Wu and colleagues mapped innovation and entrepreneurship education for medical students using a single database, identifying 1,265 records and noting the dominance of developed countries and a three-stage temporal pattern [68]. Wang and colleagues, in their 2024 analysis of artificial intelligence in medical education, retrieved 195 papers from Web of Science alone and concluded that the field was still nascent, with limited interregional collaboration [69]. The more recent 2025 analysis by Wang and colleagues in this journal combined Web of Science and Scopus with a three-pronged search strategy to identify 669 documents on AI in medical education, finding that 2019 marked a critical turning point and that generative AI research had become disconnected from earlier conventional AI work [70]. Each of these reviews relied primarily on keyword co-occurrence or cocitation clustering to map thematic structure. This review, by contrast, paired keyword co-occurrence with bibliographic coupling across three databases and then subjected the coupled clusters to a convergence check. That check revealed something single-method mapping cannot: the dominant-cluster alignment between vocabulary-based and citation-based groupings ranged from 16.0% to 45.7%, suggesting that conceptually related studies are not consistently grouped together when clustering is based on author-keyword similarity versus shared references. Prior bibliometric reviews have described fragmentation; this review measures its depth.

The second point of comparison concerns what the taxonomy contributes relative to prior content-focused syntheses. Nielsen and colleagues described the Oregon Consortium for Nursing Education clinical education model with its five elements, grounding the design explicitly in Benner's novice-to-expert theory and Tanner's clinical judgment model [29]. Pleshkan's systematic review of nurse practitioner clinical education and preceptorship identified persistent problems with clinical site availability, competency assessment, and post-graduation support, arguing for a shift from hour-based to competency-based evaluation [13]. Moya-Mendez and colleagues conducted a scoping review of longitudinal integrated clerkships and surgical education, finding that only 13% of LIC studies addressed surgical education in any substantive way and calling for a consensus statement on essential curriculum components [10]. Each of these prior works operates within a single model type or discipline. The present review's taxonomy encompasses all of them within a single classification scheme: the Oregon model sits within instructional and technology redesign and structured transition, Pleshkan's preceptorship concerns sit squarely within academic-practice and workforce-embedded partnership models, and the LIC literature anchors the continuity-based longitudinal model type. By providing a common vocabulary across these silos, the taxonomy enables comparison that has not previously been possible. It also makes visible a pattern that single-model reviews cannot: the evaluation method follows model type rather than evidence hierarchy. Continuity-based models lean on comparative quantitative designs, while workforce-embedded and immersive models rely on qualitative or reflective designs. This split, noted in the Results, may reflect measurement feasibility rather than an evidence-quality hierarchy, a distinction that matters for interpreting the literature and for planning future research.

The third point of comparison concerns theory and research agendas. Ning and colleagues, in their Viewpoint on AI in medical education, argued for structured, stage-specific educational approaches and for collaborative funding models to support sustainable AI integration [71]. Both Wang et al. bibliometric reviews concluded that future research should prioritize theoretical frameworks and interdisciplinary collaboration, particularly for generative AI [10, 11]. The 2025 Wang et al. review explicitly identified a gap between recent generative AI studies and earlier conventional AI research, arguing for integrative models [70]. This review's findings align with these recommendations but sharpen them. The 37.5% of closely read documents that named no explicit theoretical framework is not, by itself, the core problem; it is that even the 62.5% that did name a framework mostly used it descriptively rather than to explain mechanism or predict transfer. The agenda-setting gaps identified here, including the need for theory-driven evaluation that tests competing accounts of why continuity- and partnership-based models work, the need for longitudinal outcome tracking, and the need for economic evidence, are grounded in this observed pattern. Prior reviews have urged more theory; this review specifies that theory must be operationalized, tested, and linked to mechanism, not merely cited.

Finally, the geographic and temporal patterns observed here confirm prior findings while complicating their interpretation. Wu and colleagues reported US leadership in innovation and entrepreneurship education, with China and other nations growing more recently [68]. Both Wang et al. bibliometric analyses of AI in medical education found similar US dominance, with the 2025 review noting rising activity from China, Qatar, and Malaysia [69, 70]. The present review's geographic data show the same

concentration, with 47.7% of records having at least one US-affiliated author. But the presence of 69 countries across the full sample, and studies from Vietnam, South Africa, sub-Saharan Africa, and Iran among the 48 closely read documents, suggests that this is not a field confined to high-income settings. The concentration is in publication, not necessarily in practice. This distinction matters for the transferability of the taxonomy. A continuity-based model premised on stable student cohorts and long-term preceptor relationships may not translate directly to settings with different workforce structures or training pipelines, a tension that the current literature has only begun to explore.

Strengths and Limitations

This review has several methodological strengths. The dual bibliometric approach, combining keyword co-occurrence with bibliographic coupling, offered a way to map both the conceptual and the citation-based structure of the field. Coupling this with AI-assisted preliminary screening followed by human cross-review and consensus provided a scalable screening workflow while retaining human control over final eligibility decisions. The use of a CBIS-based sampling strategy for the close reading ensured that the qualitative synthesis was grounded in the most influential and densely connected papers within each cluster, rather than an arbitrary subset. Finally, the convergence check, which found dominant-cluster alignment as high as 45.7%, provided an independent bibliometric perspective that partially supported the thematic structure identified through qualitative synthesis.

Several limitations must be acknowledged. First, the bibliographic coupling network is not representative of the entire literature. Of the 774 screened records, 360 (46.5%) lacked mutual reference links dense enough to enter the coupling map. These records were retained in the qualitative synthesis, but the network-based cluster analysis and the CBIS-based close-reading selection were limited to the 414 connected documents. Patterns observed in the coupling clusters may not generalize to the excluded subset. Second, the convergence check itself had partial coverage: 277 of 414 coupled documents carried any recorded author keywords, and only 247 (59.7%) had at least one keyword matching the finalized 80-term list. The concordance figures therefore reflect alignment only among this matched subset, not the full coupling network. Third, the close reading sample of 48 documents, while theoretically justified through CBIS ranking, represents only 6.2% of the full corpus. The close-reading strategy was designed to maximize thematic coverage of highly connected and influential documents within each cluster; nevertheless, the limited sampling fraction may have missed less-cited or peripheral perspectives. Fourth, the geographic concentration in the sample means that the findings, particularly the model descriptions, are heavily skewed toward the United States and other predominantly high-resource settings, which limits global transferability. Fifth, the classification of models was based on published descriptions, which may not always align perfectly with on-the-ground implementation reality. Sixth, although no language restriction was imposed at eligibility, the search remained dependent on indexed databases and therefore may have underrepresented locally published or poorly indexed evidence.

Implications for Practice and Future Research

The agenda-setting synthesis identified seven specific gaps, each of which carries both practical and research implications. For practice, the five-model taxonomy itself is a tool. Educators and curriculum leaders can use it to audit their own programs, asking which mechanisms of transformation they are actually employing and whether those mechanisms are aligned with their stated goals. The finding that theory is underused suggests a practical need for program developers to articulate their theory of change from the outset, not retrospectively. What mechanism should lead to what outcome, and through what causal pathway? This is a planning discipline that moves beyond merely describing activities. The dominance of evaluation studies from the United States signals a need for greater investment in publishing local and regional innovations from other parts of the world, if the field is to learn from a truly global evidence base.

For future research, the gaps point toward a coordinated agenda. The first, and perhaps most critical, is the need for theory-driven evaluation research. Building on the taxonomy, future studies should test whether certain model types are more effective for specific outcomes or learner populations, and why. This requires a shift from descriptive reports to hypothesis-testing designs that engage with the mechanisms of the models themselves. Second, the geographic bias in the literature demands a program of multi-country, comparative research to understand which components of these models are context-dependent and which are portable. Third, the apparent imbalance in evaluation methods across model types calls for investment in mixed-methods evaluation of complex, longitudinal models, which are currently under-studied relative to their immersive counterparts. Fourth, research is needed on the longitudinal outcomes of these models, tracking graduates into practice to assess long-term effects on preparedness, identity formation, and career choices. Fifth, the role of technology as a core component of instructional redesign, rather than an add-on, warrants focused study. Sixth, the patient and community perspective is largely absent from the sampled literature and represents a critical frontier. Finally, the field would benefit from consensus-building work to develop shared language and reporting standards for educational transformation models, directly addressing the fragmentation that the bibliometric clusters and the convergence check reveal.

Conclusions

This review mapped the landscape of clinical education transformation across health professions using a dual bibliometric approach and synthesized a five-model taxonomy from the most influential documents in the field. The taxonomy offers a mechanism-based classification system that moves beyond binary or topic-based labels. Theory usage in the field is real but shallow, and the weak convergence between vocabulary-based and citation-based groupings signals conceptual fragmentation that hampers cumulative progress. Addressing the seven agenda-setting gaps identified here will require coordinated, theory-driven, and internationally comparative research. The taxonomy provides a starting point for that work.

Abbreviations

AI: artificial intelligence

AAMC: Association of American Medical Colleges

ACGME: Accreditation Council for Graduate Medical Education

B-SLR: bibliometric-systematic literature review

CBIS: composite bibliometric influence score

DEU: dedicated education unit

EPA: entrustable professional activity

GMC: General Medical Council

LIC: longitudinal integrated clerkship

MeSH: Medical Subject Headings

NBME: National Board of Medical Examiners

OSCE: objective structured clinical examination

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

QSEN: Quality and Safety Education for Nurses

RIME: Reporter, Interpreter, Manager, Educator

Declarations

Acknowledgments

The authors used DeepSeek-R1 for the initial screening triage of records against the predefined eligibility criteria. All final screening decisions were made by the authors through cross-review and consensus.

Funding

No external funding was received for this study.

Conflicts of Interest

None declared.

Availability of data and materials

The datasets generated and analyzed during this study are included in this published article and its supplementary information files. The complete search strategies for PubMed, Scopus, and Web of Science are provided as Multimedia Appendix A1. The VOSviewer thesaurus file and the full journal and country lists are provided as Multimedia Appendices S1 and D1.

Authors' Contributions

Yaser Davodi was responsible for conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, writing – review & editing, visualization, supervision, and project administration. Mohammad Javad Hajivand contributed to methodology, investigation, data curation, writing – original draft, and writing – review & editing. Omid Gheisavandi contributed to investigation, writing – original draft, and writing – review & editing. Amir Hossein Dehghan contributed to writing – original draft and writing – review & editing. All authors read and approved the final manuscript. Yaser Davodi is the guarantor of the work.

References

1. Dornan T, Boshuizen H, King N, Scherpbier A. Experience-based learning: a model linking the processes and outcomes of medical students' workplace learning. Med Educ. 2007;41(1):84–91.
2. Yardley S, Teunissen PW, Dornan T. Experiential learning: transforming theory into practice. Med Teach. 2012;34(2):161–4.
3. Cruess SR, Cruess RL, Steinert Y. Supporting the development of a professional identity: General principles. Med Teach. 2019;41(6):641–9.
4. Gruppen L, Irby DM, Durning SJ, Maggio LA. Interventions Designed to Improve the Learning Environment in the Health Professions: A Scoping Review. *MedEdPublish (2016)* 2018, 7:211.
5. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, Fineberg H, Garcia P, Ke Y, Kelley P, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. Lancet. 2010;376(9756):1923–58.
6. McIsaac M, Buchan J, Abu-Agla A, Kawar R, Campbell J. Global Strategy on Human Resources for Health: Workforce 2030-A Five-Year Check-In. Hum Resour Health. 2024;22(1):68.
7. Frenk J, Chen LC, Chandran L, Groff EOH, King R, Meleis A, Fineberg HV. Challenges and opportunities for educating health professionals after the COVID-19 pandemic. Lancet. 2022;400(10362):1539–56.
8. Irby DM. Teaching and learning in ambulatory care settings: a thematic review of the literature. Acad Med. 1995;70(10):898–931.
9. Teunissen PW, Westerman M. Opportunity or threat: the ambiguity of the consequences of transitions in medical education. Med Educ. 2011;45(1):51–9.
10. Moya-Mendez ME, Thornton S, Rhodin KE, Gao Q, Leraas HJ, Vatsaas CJ. Longitudinal Integrated Clerkships and Undergraduate Surgical Education: A Scoping Review and Gap Analysis. J Surg Educ.

2024;81(3):367–72.

11. Carraccio C, Englander R, Van Melle E, Ten Cate O, Lockyer J, Chan MK, Frank JR, Snell LS. Advancing Competency-Based Medical Education: A Charter for Clinician-Educators. Acad Med. 2016;91(5):645–9.
12. Goh PS, Sandars J. A vision of the use of technology in medical education after the COVID-19 pandemic. MedEdPublish (2016). 2020;9:49.
13. Pleshkan V. A systematic review: Clinical education and preceptorship during nurse practitioner role transition. J Prof Nurs. 2024;50:16–34.
14. Taoube L, Khanna P, Schneider C, Burgess A, Bleasel J, Haq I, Roberts C. Situated learning in community environments (SLICE): Systems design of an immersive and integrated curriculum for community-based learning. Med Teach. 2023;45(1):80–8.
15. Durning SJ, Artino AR. Situativity theory: a perspective on how participants and the environment can interact: AMEE Guide 52. Med Teach. 2011;33(3):188–99.
16. Norris TE, Schaad DC, DeWitt D, Ogur B, Hunt DD. Longitudinal integrated clerkships for medical students: an innovation adopted by medical schools in Australia, Canada, South Africa, and the United States. Acad Med. 2009;84(7):902–7.
17. Wang S, Yang L, Li M, Zhang X, Tai X. Medical Education and Artificial Intelligence: Web of Science-Based Bibliometric Analysis (2013–2022). JMIR Med Educ. 2024;10:e51411.
18. Marzi G, Balzano M, Caputo A, Pellegrini MM. Guidelines for Bibliometric-Systematic Literature Reviews: 10 steps to combine analysis, synthesis and theory development. Int J Manage Reviews. 2025;27(1):81–103.
19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
20. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210.
21. Pittaway L, Robertson M, Munir K, Denyer D, Neely A. Networking and innovation: a systematic review of the evidence. Int J Manage Reviews. 2004;5–6(3–4):137–68.
22. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 2010;84(2):523–38.
23. Zupic I, Čater T. Bibliometric Methods in Management and Organization. Organizational Res Methods. 2015;18(3):429–72.
24. Petticrew M, Roberts H. Systematic reviews in the social sciences: A practical guide. 2006.
25. Torraco RJ. Writing integrative literature reviews: Using the past and present to explore the future. Hum Resour Dev Rev. 2016;15(4):404–28.
26. Poncelet A, Bokser S, Calton B, Hauer KE, Kirsch H, Jones T, Lai CJ, Mazotti L, Shore W, Teherani A et al. Development of a longitudinal integrated clerkship at an academic medical center. Med Educ

Online 2011, 16.

27. Ogur B, Hirsh D. Learning Through Longitudinal Patient Care—Narratives From the Harvard Medical School–Cambridge Integrated Clerkship. Acad medicine: J Association Am Med Colleges. 2009;84:844–50.
28. Hudson JN, Knight PJ, Weston KM. Patient perceptions of innovative longitudinal integrated clerkships based in regional, rural and remote primary care: a qualitative study. BMC Fam Pract. 2012;13:72.
29. Nielsen AE, Noone J, Voss H, Mathews LR. Preparing nursing students for the future: an innovative approach to clinical education. Nurse Educ Pract. 2013;13(4):301–9.
30. Padiko YY, Suhoyo Y, Prihatiningsih TS. Impact of early, longitudinal community-based education on medical student self-efficacy: a qualitative study. BMC Med Educ. 2025;25(1):1668.
31. Latessa R, Beaty N, Royal K, Colvin G, Pathman DE, Heck J. Academic outcomes of a community-based longitudinal integrated clerkships program. Med Teach. 2015;37(9):862–7.
32. Mazotti L, Adams J, Peyser B, Chretien K, Duffy B, Hirsh DA. Diffusion of innovation and longitudinal integrated clerkships: Results of the clerkship directors in internal medicine annual survey. Med Teach. 2019;41(3):347–53.
33. O'Connor EE, Fried J, McNulty N, Shah P, Hogg JP, Lewis P, Zeffiro T, Agarwal V, Reddy S. Flipping Radiology Education Right Side Up. Acad Radiol. 2016;23(7):810–22.
34. Belfi LM, Bartolotta RJ, Giambrone AE, Davi C, Min RJ. Flipping the introductory clerkship in radiology: impact on medical student performance and perceptions. Acad Radiol. 2015;22(6):794–801.
35. Lin Y, Zhu Y, Chen C, Wang W, Chen T, Li T, Li Y, Liu B, Lian Y, Lu L, et al. Facing the challenges in ophthalmology clerkship teaching: Is flipped classroom the answer? PLoS ONE. 2017;12(4):e0174829.
36. Morgan H, McLean K, Chapman C, Fitzgerald J, Yousuf A, Hammoud M. The flipped classroom for medical students. Clin Teach. 2015;12(3):155–60.
37. Keeley MG, Gusic ME, Morgan HK, Aagaard EM, Santen SA. Moving Toward Summative Competency Assessment to Individualize the Postclerkship Phase. Acad Med. 2019;94(12):1858–64.
38. Patel L, Buck P, Dornan T, Sutton A. Child health and obstetrics-gynaecology in a problem-based learning curriculum: accepting the limits of integration and the need for differentiation. Med Educ. 2002;36(3):261–71.
39. Coates WC, Crooks K, Slavin SJ, Guiton G, Wilkerson L. Medical school curricular reform: fourth-year colleges improve access to career mentoring and overall satisfaction. Acad Med. 2008;83(8):754–60.
40. Mulready-Shick J, Flanagan KM, Banister GE, Mylott L, Curtin LJ. Evaluating dedicated education units for clinical education quality. J Nurs Educ. 2013;52(11):606–14.

41. Masters K. Integrating quality and safety education into clinical nursing education through a dedicated education unit. Nurse Educ Pract. 2016;17:153–60.
42. Rhodes ML, Meyers CC, Underhill ML. Evaluation outcomes of a dedicated education unit in a baccalaureate nursing program. J Prof Nurs. 2012;28(4):223–30.
43. Christiansen A, Bell A. Peer learning partnerships: exploring the experience of pre-registration nursing students. J Clin Nurs. 2010;19(5–6):803–10.
44. Johnsson A, Månsson JS, Pennbrant S. Scaffolding for learning and establishing a professional identity: A qualitative descriptive study of nursing students' experiences of learning with a student-centered supervision model based on patient-oriented care during clinical placement. Nordic J Nurs Res. 2023;43(3–4):20571585231221257.
45. Plemmons C, Clark M, Feng D. Comparing student clinical self-efficacy and team process outcomes for a DEU, blended, and traditional clinical setting: A quasi-experimental research study. Nurse Educ Today. 2018;62:107–11.
46. Vosoughi MN, Zamanzadeh V, Valizadeh L, Ghahramanian A, Lotfi M, Bagheriyeh F, Pourmollamirza A. An introduction to the TPSN model: a comprehensive approach to reducing the theory-practice gap in nursing. BMC Nurs. 2022;21(1):261.
47. Samuriwo R, Laws E, Webb K, Bullock A. I didn't realise they had such a key role. Impact of medical education curriculum change on medical student interactions with nurses: a qualitative exploratory study of student perceptions. Adv Health Sci Educ Theory Pract. 2020;25(1):75–93.
48. Edwards AP, Nash AJ. Transformative care for people with disabilities: Empowering senior nursing students with competency based clinical education-A qualitative study of the impact. Nurse Educ Today. 2023;126:105822.
49. Ottenheijm RP, Zwietering PJ, Scherpbie AJ, Metsemakers JF. Early student-patient contacts in general practice: an approach based on educational principles. Med Teach. 2008;30(8):802–8.
50. Head BA, Earnshaw LA, Greenberg RB, Morehead RC, Pfeifer MP, Shaw MA. I will never forget: what we learned from medical student reflections on a palliative care experience. J Palliat Med. 2012;15(5):535–41.
51. Miller-Rosser K, Colgrave J, Stockhausen L. The student experience of a 'kidney CAMP' as a primary care clinical placement for undergraduate nursing students. Nurse Educ Today. 2019;83:104198.
52. Gray B, Grealish L, Ranse K, Terry V, Armit L, van de Mortel T, Del Fabbro L. The assessment of undergraduate bachelor of nursing students in the collaborative clusters education model: A qualitative descriptive design. Nurse Educ Pract. 2023;70:103675.
53. Ogur B, Hirsh D, Krupat E, Bor D. The Harvard Medical School-Cambridge integrated clerkship: an innovative model of clinical education. Acad Med. 2007;82(4):397–404.
54. Prideaux D, Worley P, Bligh J. Symbiosis: a new model for clinical education. Clin Teach. 2007;4(4):209–12.
55. Botha R, Breedt DS, Barnard D, Couper I. Lessons from innovation in medical education during the COVID-19 pandemic: student perspectives on distributed training. Rural Remote Health.

2023;23(4):8257.

56. Patterson C, Moxham L, Brighton R, Taylor E, Sumskis S, Perlman D, Heffernan T, Hadfield L. Nursing students' reflections on the learning experience of a unique mental health clinical placement. Nurse Educ Today. 2016;46:94–8.
57. Gay S, Bartlett M, McKinley R. Teaching clinical reasoning to medical students. Clin Teach. 2013;10(5):308–12.
58. Negrete Manriquez JA, Bazargan-Hejazi S, Nahm SJ, de Virgilio C. Exploring a novel approach to surgery clerkship didactics during the COVID-19 pandemic: A qualitative study. Am J Surg. 2022;223(4):662–9.
59. Jacobs JC, Bolhuis S, Bulte JA, Laan R, Holdrinet RS. Starting learning in medical practice: an evaluation of a new Introductory Clerkship. Med Teach. 2005;27(5):408–14.
60. Atherley A, Dolmans D, Hu W, Hegazi I, Alexander S, Teunissen PW. Beyond the struggles: a scoping review on the transition to undergraduate clinical training. Med Educ. 2019;53(6):559–70.
61. Müller J, Meyer R, Bantjes J, Archer E, Couper I. Handle with Care: Transformative Learning as Pedagogy in an Under-Resourced Health Care Context. Teach Learn Med. 2025;37(3):412–21.
62. Talib Z, van Schalkwyk S, Couper I, Pattanaik S, Turay K, Sagay AS, Baingana R, Baird S, Gaede B, Iputo J, et al. Medical Education in Decentralized Settings: How Medical Students Contribute to Health Care in 10 Sub-Saharan African Countries. Acad Med. 2017;92(12):1723–32.
63. Tran TD, Vu PM, Pham HTM, Au LN, Do HP, Doan HTT, Huynh N, Huynh QTV, Le BK, Ngo DQ, et al. Transforming medical education to strengthen the health professional training in Viet Nam: A case study. Lancet Reg Health West Pac. 2022;27:100543.
64. Austin JP, Baskerville M, Bumsted T, Haedinger L, Nonas S, Pohoata E, Rogers M, Spickerman M, Thuillier P, Mitchell SH. Development and evaluation of a simulation-based transition to clerkship course. Perspect Med Educ. 2020;9(6):379–84.
65. Taylor JS, George PF, MacNamara MM, Zink D, Patel NK, Gainor J, Dollase RH. A new clinical skills clerkship for medical students. Fam Med. 2014;46(6):433–9.
66. Bracken K, Levinson AJ, Mahmud M, Allice I, Vanstone M, Grierson L. Spiralling Pre-clerkship Concepts into the Clinical Phase: Augmenting Knowledge Transfer Using Innovative Technology-Enhanced Curriculum Activities. Med Sci Educ. 2021;31(5):1607–20.
67. Shahkarami F, Taghavi AA, Dolama RH, Khajavirad N, Mirzazadeh A, Gandomkar R. Students’ and faculty members’ perceptions of the online component of a blended internal medicine clerkship course: a mixed-method evaluation. BMC Med Educ. 2025;25(1):1221.
68. Wu X, Zhan F, Zhang X, Wang T. Innovation and entrepreneurship education for medical students: a global bibliometric analysis (2000–2024). Med Educ Online. 2025;30(1):2515385.
69. Wang S, Yang L, Li M, Zhang X, Tai X. Medical Education and Artificial Intelligence: Web of Science–Based Bibliometric Analysis (2013–2022). JMIR Med Educ. 2024;10:e51411.

70. Wang Y, Chang C, Shi W, Liu H, Huang X, Jiao Y. How AI Is Transforming Medical Education: Bibliometric Analysis. JMIR Med Educ. 2025;11:e75911.
71. Ning Y, Ong JCL, Cheng H, Wang H, Ting DSW, Tham YC, Wong TY, Liu N. How can artificial intelligence transform the training of medical students and physicians? Lancet Digit Health. 2025;7(10):100900.

Figures

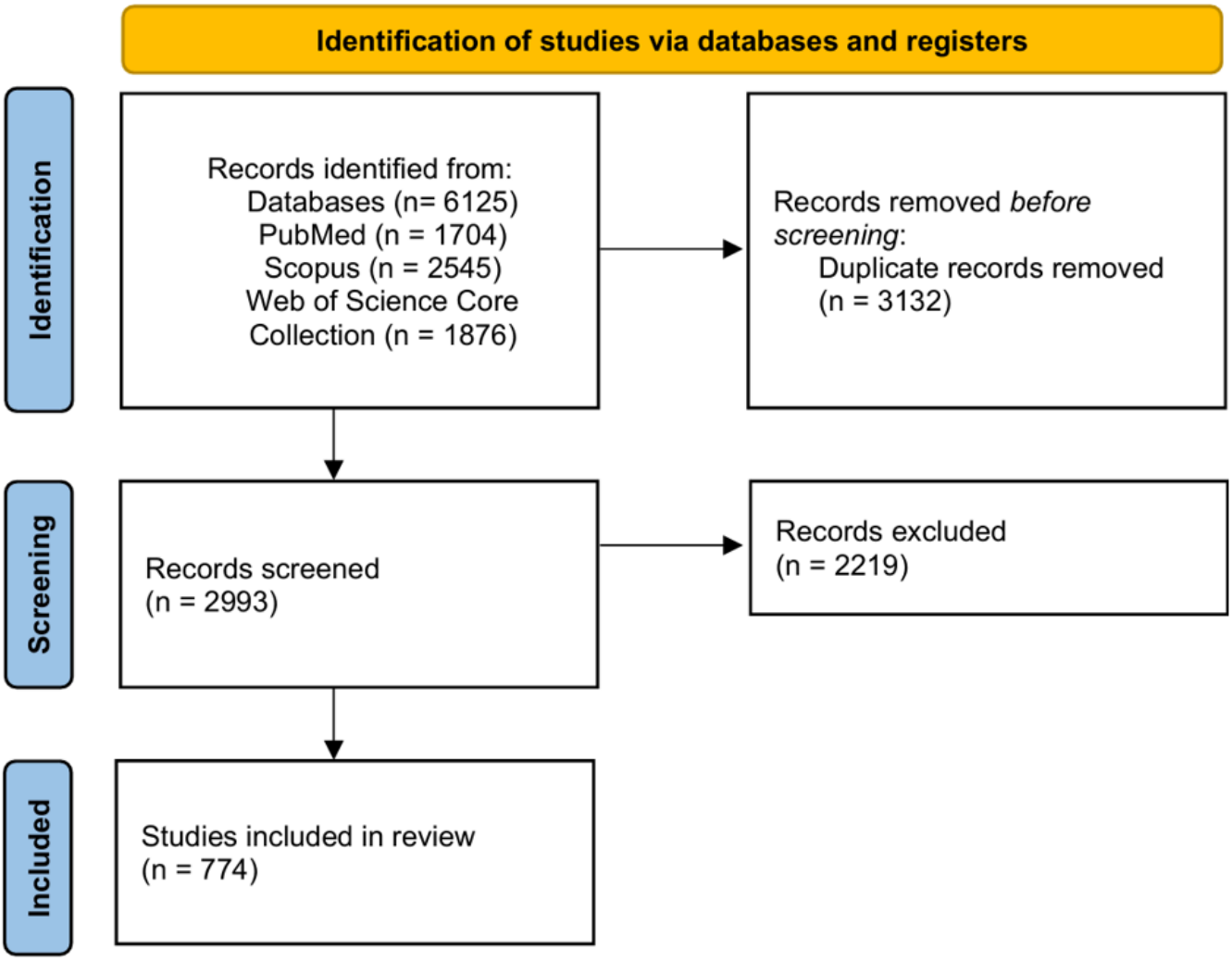


Figure 1

PRISMA 2020 flow diagram

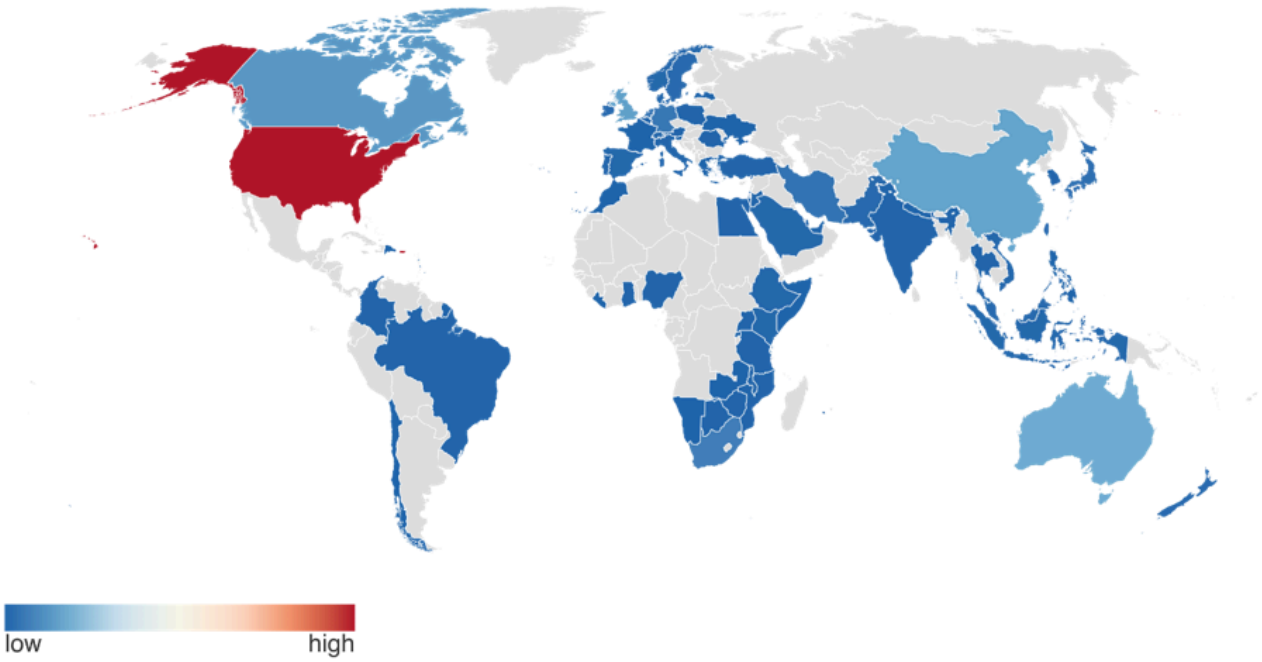


Figure 2

World map of contributing countries

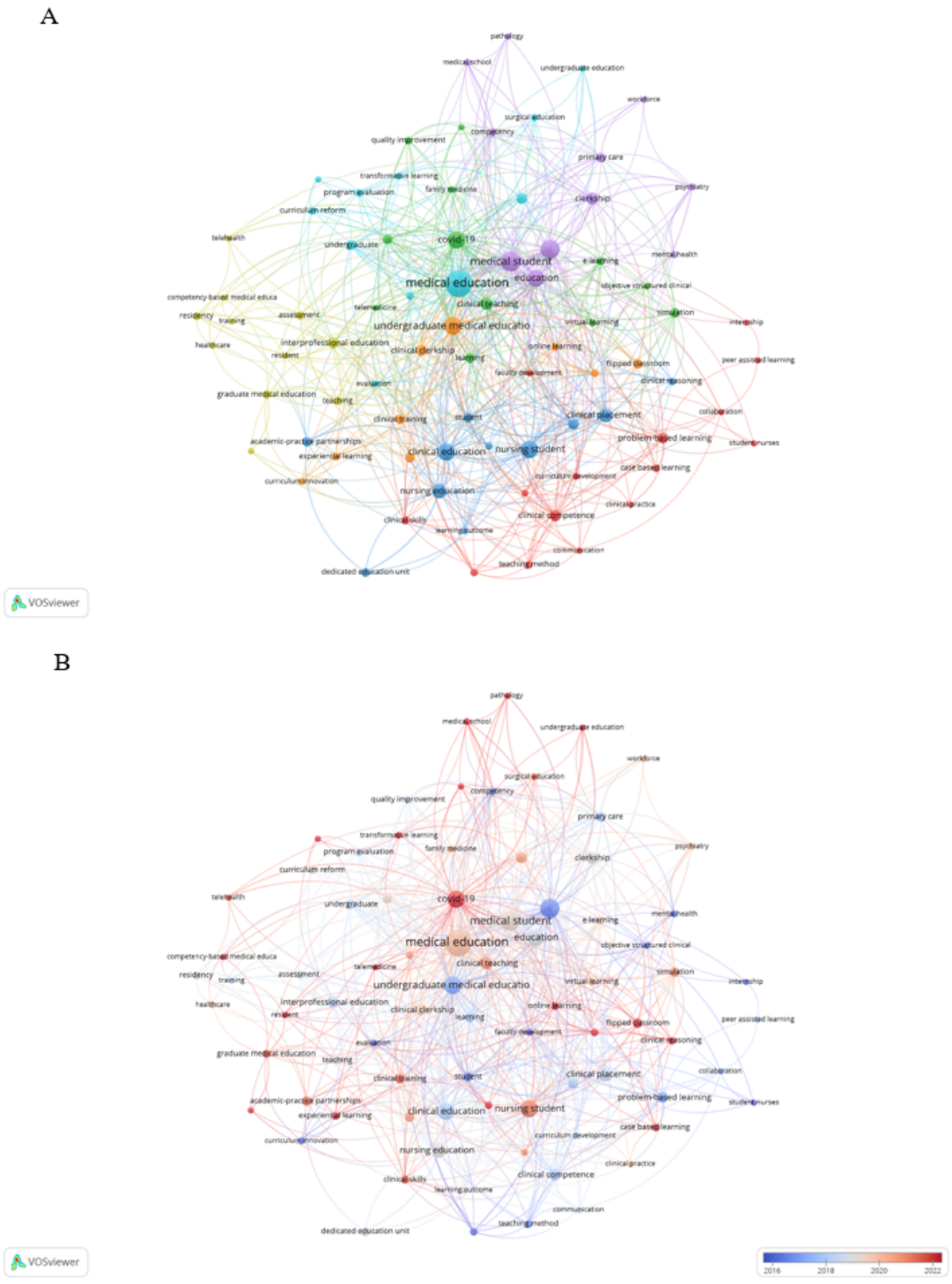


Figure 3

Keyword co-occurrence network (VOSviewer); (A) Network visualization, colored by cluster membership. (B) Overlay visualization, colored by the average publication year of each term's occurrences.

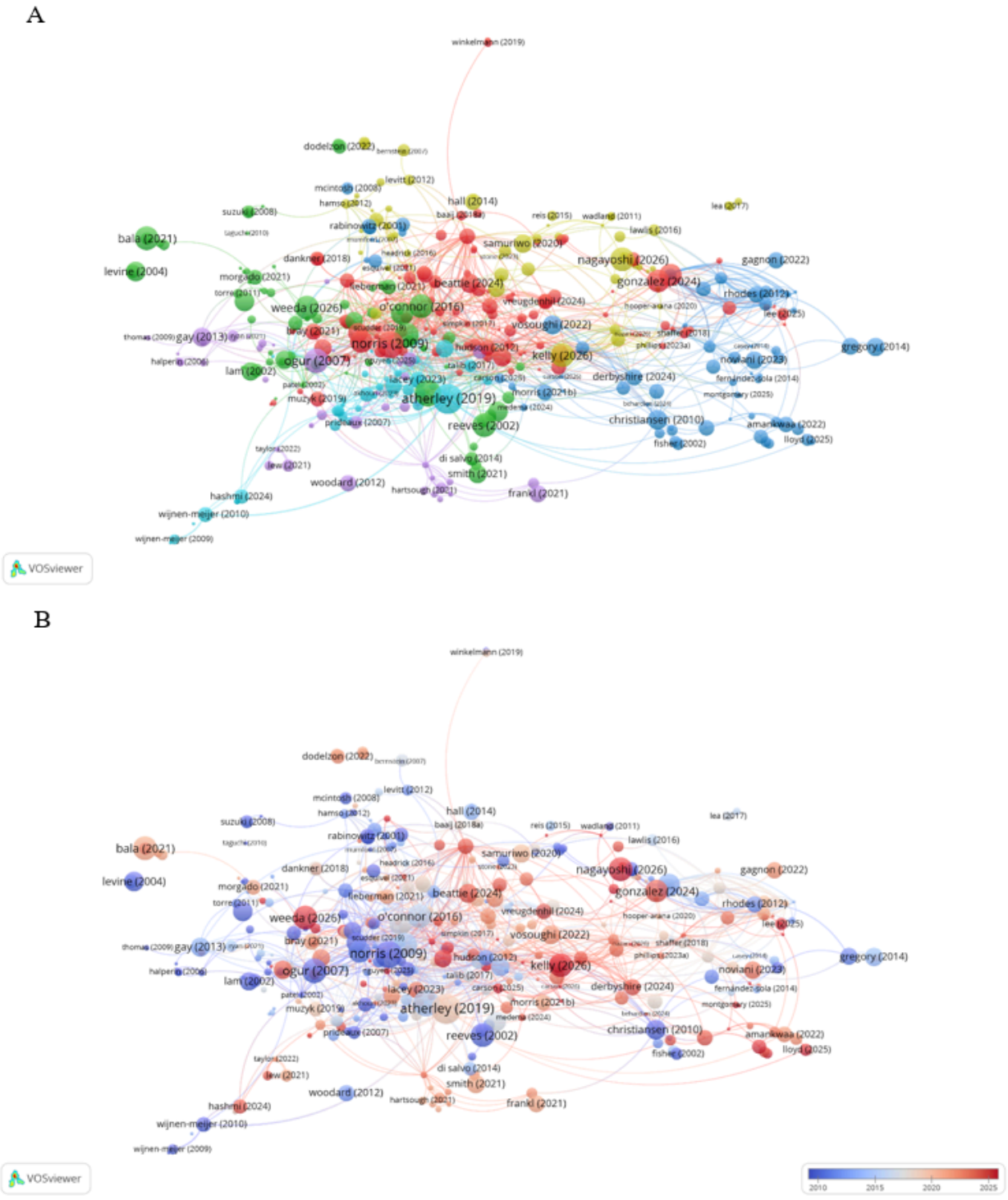


Figure 4

Bibliographic coupling network (VOSviewer), based on 414 documents (of 774 screened) (A) Network visualization, colored by cluster membership. (B) Overlay visualization, colored by publication year

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- A1.docx
- S1.zip
- D1.xlsx
- S2.zip