Creative Education, 2026, 17(11), 2041-2066
https://www.scirp.org/journal/ce
ISSN Online: 2151-4771
ISSN Print: 2151-4755

Beyond Instructional Labels: A Realist Review of Active-Learning Variation in Medical Education

Anatoli Em. Argyrou*, **Andreas C. Lazaris**

First Department of Pathology, Medical School, National and Kapodistrian University of Athens, Athens, Greece
Email: *a.argyrou@outlook.com

How to cite this paper: Argyrou, A. E., & Lazaris, A. C. (2026). Beyond Instructional Labels: A Realist Review of Active-Learning Variation in Medical Education. *Creative Education, 17,* 2041-2066.
https://doi.org/10.4236/ce.2026.1711115

Received: September 28, 2026
Accepted: October 31, 2026
Published: November 3, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

Introduction: A realist review examined how learning mechanisms explain variation across teaching approaches in undergraduate medical education (UME). **Methods:** PubMed/MEDLINE, Scopus, Web of Science, and ERIC were searched through August 2026. Of 2383 records, 27 met inclusion criteria. Evidence was synthesized using context-mechanism-outcome configurations, with Interactive, Constructive, Active, Passive (ICAP) characterizing engagement. **Results:** Three mechanism-level demi-regularities were identified. Mechanism activation, cognitive engagement, constructive alignment, and implementation fidelity explained variation, while instructional labels showed heterogeneity. **Conclusion:** Instructional labels are family indicators rather than interventions. Mechanism-ICAP profiles and implementation fidelity provide analytical units for understanding outcomes across UME contexts.

Keywords

Medical Education, Active Learning, Realist Review, ICAP Framework, Cognitive Engagement

1. Introduction

1.1. The Educational Problem

Medical education prepares students to acquire knowledge, interpret evidence, solve clinical problems, communicate effectively, collaborate, and apply learning in patient care. These demands have prompted the adoption of active-participation approaches that engage students directly rather than relying solely on lectures. Across STEM education, active learning has been associated with improved examination performance and reduced course failure compared with traditional

teaching (Freeman et al., 2014). Active-participation approaches—including flipped classrooms, team-based learning (TBL), problem-based learning (PBL), peer-assisted learning (PAL), BOPPPS [bridge-in, outcomes (or objectives), pre-assessment, participatory learning, post-assessment, and summary], and Jigsaw teaching—combine retrieval practice, discussion, problem solving, application, explanation, and feedback, yet identical instructional labels often represent markedly different activities across institutions. Accordingly, instructional labels may function as family indicators, with outcomes determined by the learning activities and cognitive mechanisms that are actually enacted. Reported effects vary widely: flipped classrooms (SMD 0.33 - 0.93) (Hew & Lo, 2018; Spaić et al., 2025; Hughes & Lyons, 2017), team-based learning (d ≈ 0.55 - SMD 2.63) (Chen et al., 2018; Swanson et al., 2019), problem-based learning shows domain-specific effects (Trullàs et al., 2022), and team-based learning outperformed problem-based learning (SMD 0.37, 95% CI 0.02 - 0.73) (Zhang et al., 2024). Together with evidence from cognitive-architecture (Bordage, 2009; Sweller et al., 2019; Leppink & van den Heuvel, 2015; Young et al., 2014), testing-effect (Pyc & Rawson, 2010; Pan & Rickard, 2018; Adesope et al., 2017; Rowland, 2014; Yang et al., 2021), and peer-learning literatures (Loda et al., 2019; Loda et al., 2020; ten Cate et al., 2012; Brierley et al., 2022,) this variability motivates a mechanism-based rather than label-based synthesis.

1.2. Conceptual Framework

A realist conceptual framework integrates theory, intervention, evidence, and educational practice so that a reader can recognize what advance in knowledge a synthesis offers (Bordage, 2009; Bucklin et al., 2021). This review is organized around such a framework. The framework has one anchor, one engagement classifier, six mechanism resources, and one cross-cutting moderator; their relationships are derived rather than merely enumerated.

Anchor: constructive alignment. Constructive alignment of intended learning outcomes, teaching activities, and assessment is the condition under which mechanisms can be observed consistently. When teaching, learner activity, and assessment target the same construct, an intervention resource can plausibly activate a corresponding learner response; when alignment is absent, observed effects cannot be cleanly attributed to any mechanism. The anchor therefore establishes the analytic boundary inside which the six resources can be observed.

Engagement classifier: ICAP. ICAP distinguishes Passive, Active, Constructive, and Interactive engagement by what the learner produces. ICAP is not a seventh mechanism; it specifies the level of engagement at which the anchor can be observed to operate. ICAP classifies the engagement a resource is expected to elicit, while the mechanism resources explain what the resource does (**Figure 1**).

Six mechanism resources. Working-memory redistribution, retrieval, accountability, context anchoring, constructive alignment (operating again at the resource tier when applied to specific outcomes), and cognitive and social con-

gruence are specified underneath the anchor and classifier as the design features through which mechanism activation plausibly occurs. Resources do not stand alone. They require anchor-level alignment to be observed consistently, and they require Constructive or Interactive engagement to take effect at the level realist intervention theory predicts. Their theoretical grounding, mechanism-level evidence base, and operationalization are detailed in §1.3.

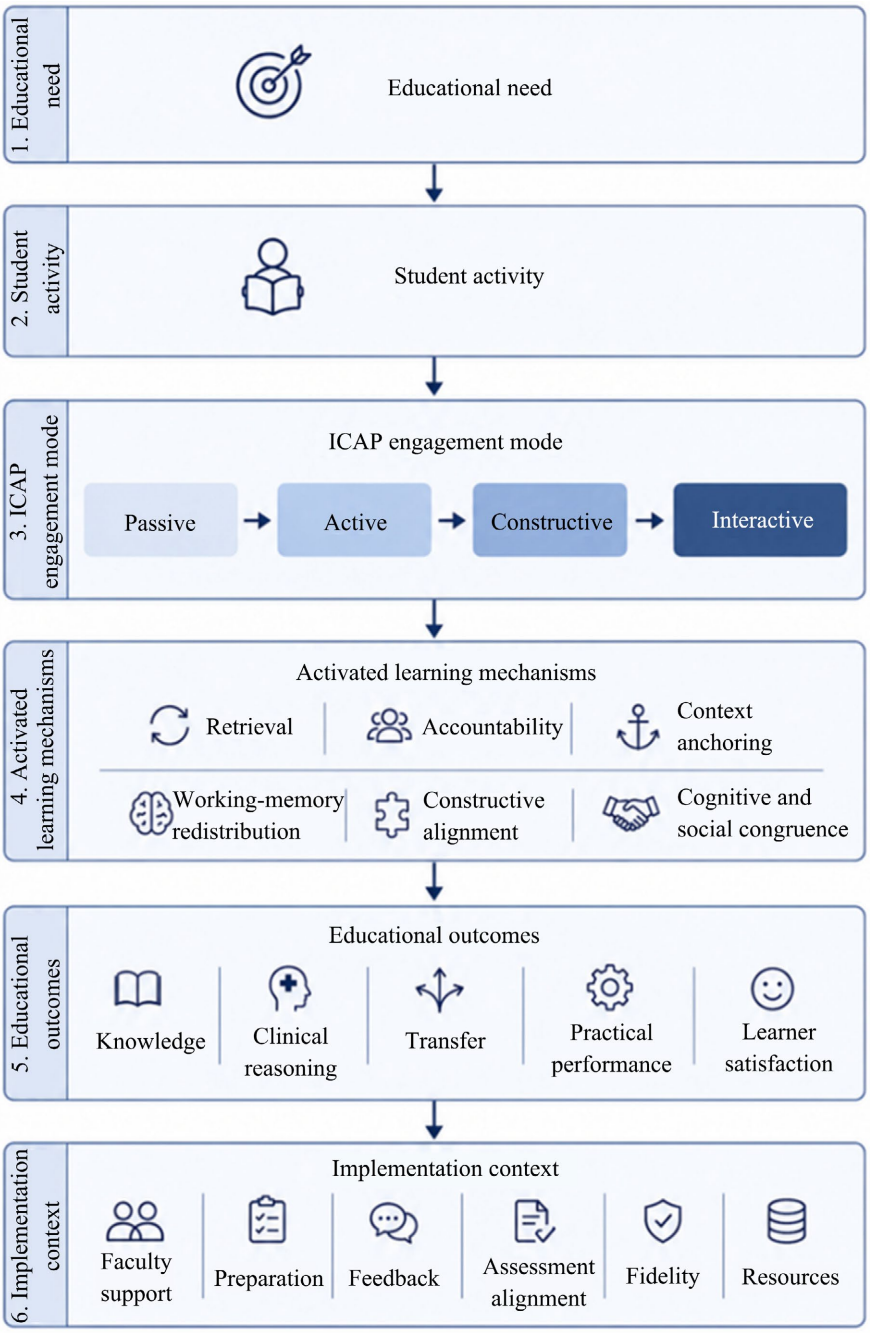


Figure 1. ICAP-mechanism conceptual framework. The diagram links educational need to student activity, to ICAP engagement mode, to activated mechanism, and to educational outcome.

Cross-cutting moderator: implementation fidelity. Fidelity is treated as a moderator on activation rather than as a seventh mechanism. Fidelity determines whether the design feature presented to learners manifests as mechanism activation in any given implementation, and is recorded during extraction and again during CMO adjudication.

The framework maintains a single hierarchy throughout the review: context supplies the condition for activation; the intervention resource is the design feature presented to learners; the mechanism is the inferred learner response; ICAP specifies the engagement level at which the mechanism operates; the outcome is the educational result; and fidelity moderates whether the resource actually manifests as activation.

1.3. Mechanism Profiles

Six mechanisms structure the realist analysis. Each is presented as an intervention resource—a design feature engaging a learner response within an appropriate context.

Working-memory redistribution. Pre-class preparation shifts information transmission outside class, freeing time for application and feedback; activation depends on repurposing this time for high-demand activities (Sweller et al., 2019; Leppink & van den Heuvel, 2015; Young et al., 2014).

Retrieval. Questions, exercises, tests, and case prompts require the retrieval of information, plausibly strengthening retention and transfer (Pyc & Rawson, 2010; Pan & Rickard, 2018; Adesope et al., 2017; Rowland, 2014; Yang et al., 2021). Within ICAP, retrieval may operate at Active, Constructive, or Interactive tiers.

Accountability. Readiness tests, quizzes, peer check-ins, and participation expectations plausibly activate preparation, channeling effort toward assessment (Ngoc et al., 2020). Accountability may interact with retrieval through the incentive context.

Context anchoring. Authentic clinical problems plausibly activate encoding and retrieval alongside clinical cues, potentially explaining effects on clinical reasoning (Trullàs et al., 2022).

Constructive alignment. Alignment of outcomes, teaching, and assessment plausibly channels effort toward targeted processes, with stronger activation when all three target the same construct (Pereira et al., 2024).

Cognitive and social congruence. Structured peer interaction may activate appropriately pitched explanatory processes (Loda et al., 2019; Loda et al., 2020; ten Cate et al., 2012; Brierley et al., 2022; Grijpma et al., 2024). Responding to and building on peers may reach ICAP-Interactive engagement, supporting conceptual integration; opinion exchange without elaboration remains closer to Active engagement.

1.4. Initial Program Theory

Following realist review methodology, we articulated an a priori program theory

specifying three context-mechanism-outcome configurations. Context denotes conditional features relevant to mechanism activation; mechanism denotes the inferred causal process—intervention resource plus learner response—through which change occurs; outcome denotes the learner-level result. The configurations served as an analytic scaffold rather than a fixed taxonomy, with evidence used to retain, qualify, or refine each CMO. ICAP describes how mechanisms operate at different engagement levels, not as an additional mechanism.

1) CMO-1. Working-memory redistribution + retrieval + accountability → knowledge acquisition. Where preparation is required and verified, retrieval is embedded in classroom activity, and accountability is present, resources plausibly activate preparatory and retrieval effort, supporting foundational knowledge above lecture-only instruction.

2) CMO-2. Constructive alignment + context anchoring → clinical reasoning and transfer. Where teaching and assessment target applied judgement and learning is anchored in authentic clinical problems, resources plausibly activate coherent context-specific processing, supporting clinical reasoning and transfer beyond factual recall.

3) CMO-3. Cognitive/social congruence at the ICAP-Interactive tier → durable understanding through peer explanation. Where peers provide appropriately pitched explanations and learners respond constructively, reciprocal peer reasoning may support deeper conceptual integration than individual study. These configurations informed extraction and iterative theory refinement.

2. Methods

2.1. Reporting Standards

The review followed the STORIES statement (Gordon & Gibbs, 2014) and PRISMA 2020 (Page et al., 2021). Titles and abstracts were screened independently in duplicate, with disagreements resolved by discussion. A PRISMA flow diagram records the numbers of records identified, screened, eligible, and included, with reasons for exclusion at full-text review.

Framework-Derived Study Design

All methodological decisions were guided by the conceptual framework. The research question was formulated on the basis of the ICAP prediction that engagement mode may moderate outcomes. Eligibility criteria and the extraction form were constructed to ensure that variables recorded for each study map directly onto framework constructs (ICAP mode per activity; six mechanisms present/absent; intended outcome, activity, and assessment for constructive alignment; implementation fidelity (Stains & Vickrey, 2017); and educational outcomes). The subgroup structure was defined a priori by ICAP mode rather than by instructional label.

Context was captured alongside mechanisms as a condition for activation rather than solely as a source of heterogeneity. Recorded contextual fields included

learner preparation, curricular structure, faculty involvement, preparation requirements, assessment structure, educational setting, learner characteristics, group structure, and supervision/facilitation. Implementation fidelity was treated as a cross-cutting condition that could affect whether intended resources and mechanisms were enacted in a given implementation; it was not assumed to be an independently established cause of effect variation.

Mechanism profiles were coded from the learner-facing activities described in the included reports, where sufficient detail was available, and were used to adjudicate the three initial CMO configurations through realist synthesis. Because many included reports synthesized multiple primary studies, activity-level coding should be understood as limited to the activities explicitly described in the source reports, not as a complete re-extraction or recoding of every underlying primary study.

2.2. Search Strategy and Eligibility

PubMed/MEDLINE, Scopus, Web of Science Core Collection, and ERIC were searched from inception to August 2026. Because the contemporaneous search log did not preserve the verbatim platform strings, we report the reproducible conceptual strategy used to translate the eligibility criteria into each database rather than reconstructing an exact historical string. PubMed/MEDLINE combined controlled vocabulary and text words for medical/health-professions education with active-learning terms (including flipped classroom, team-based learning, problem-based learning, peer-assisted learning, BOPPPS, and Jigsaw); Scopus and Web of Science used title/abstract/keyword topic fields with the corresponding free-text concepts; ERIC used education thesaurus/free-text equivalents. The core concept was: (medical OR health-professions OR undergraduate) AND (education OR teaching OR learning) AND (active learning OR flipped classroom OR team-based learning OR problem-based learning OR peer-assisted learning OR BOPPPS OR Jigsaw), with outcome terms used for screening rather than as mandatory search terms. Citation chasing was conducted from conceptually central reports identified during searching and screening. Backward and forward citations were logged as citation-chasing records, entered into the same screening pool after duplicate removal, and subjected to the same title/abstract and full-text eligibility criteria as database-derived records; a citation-chasing record could therefore enter the final synthesis only after passing the same eligibility process. Citation chasing was used to improve conceptual completeness and was not treated as a separate evidence stream.

2.3. Study Selection and Data Extraction

We identified 2383 records addressing interactive teaching methods in higher education. After removal of 2183 duplicates, 200 records were screened for relevance to medical education; 27 reports met all predefined inclusion criteria and were included in the synthesis. The PRISMA flow diagram details the number of rec-

ords identified through database searching and citation tracking, screened by title and abstract, retrieved for full-text review, and finally included in the synthesis. The 27 reports comprised systematic reviews, meta-analyses, scoping/realist reviews, and primary comparative or quasi-experimental educational reports, covering predominantly undergraduate medical students and, where mechanistically relevant, broader undergraduate health-professions learners. To avoid conflating reviews with the primary studies they summarized, the unit of inclusion was the report, while the unit of evidence was the primary study or study set described by that report. Several reviews addressed overlapping literatures (particularly flipped classrooms, team-based learning, peer-assisted learning, and testing-enhanced learning). Because the source reports did not consistently provide machine-readable identifiers for all underlying primary studies, a complete deduplicated count of primary studies across reviews could not be established; overlap was therefore treated as a source of evidence dependence rather than as independent replication. Citation-chasing records entered the same screening and eligibility pathway described above. A standardized extraction form captured study/report design, learner population, intervention, comparator, mechanism profile, ICAP engagement mode where coding was possible, outcome, effect estimate, heterogeneity, funding, conflicts of interest, and the source of each effect or quality judgement. A report-level evidence map was used to distinguish report type, learner population, and overlapping evidence domains (Figure 2).

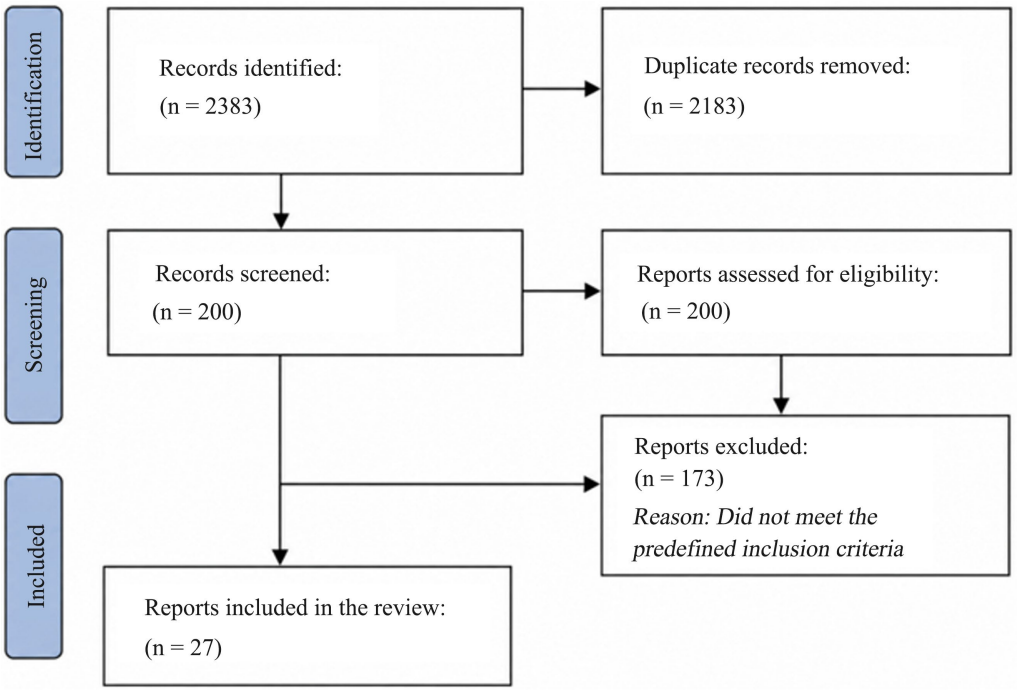


Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) diagram.

A report-level evidence map is provided in Table 1, including study type, learner population, and the evidence domain in which each report contributed.

Table 1. The table identifies the 27 reports treated as substantive evidence sources in this review. Study type and learner population are reported at the broad level supported by the manuscript and the cited reports.

Report	Study type	Learner population	Evidence domain/overlap
Brierley et al. (2022)	Systematic review/meta-analysis	Medical students/health-professions learners	PAL; overlap with near-peer/PAL literature
Bucklin et al. (2021)	Survey study	Continuing medical education professionals/units	Active-learning implementation; survey of educational practice
Chen et al. (2018)	Meta-analysis	Medical students in China	TBL; overlap with broader TBL evidence
Daly et al. (2026)	Realist review	Undergraduate health-professions learners	Case-based learning; mechanistic context
Grijpma et al. (2024)	Qualitative interview study (constructivist grounded theory)	Experienced educators in small-group learning	Small-group engagement; educator perspectives
Hew & Lo (2018)	Meta-analysis	Health-professions students	Flipped classroom; overlap with Hughes and Spaić
Hughes & Lyons (2017)	Systematic review	Medical students	Flipped classroom; overlap with Hew and Spaić
Loda et al. (2019)	Scoping review	Medical students/medical trainees	Cognitive and social congruence; PAL
Loda et al. (2020)	Instrument-development study	Medical trainees	Cognitive/social congruence; related PAL evidence
Ma et al. (2022)	Systematic review/meta-analysis	Medical students in China	BOPPPS
Maggio et al. (2018)	Meta-synthesis	Medical education evidence/guidance literature	Evidence synthesis; methodological/contextual support
Ngoc et al. (2020)	Meta-analysis	Medical students in TBL	Readiness assurance/accountability
Pahwa et al. (2022)	Quasi-experimental primary study	Undergraduate medical students	Jigsaw pathology teaching
Secomb (2008)	Systematic review	Clinical education learners/health-professions students	Peer teaching and learning
Spaić et al. (2025)	Systematic review/meta-analysis	Medical students/medical education learners	Flipped classroom; overlap with Hew and Hughes
Swanson et al. (2019)	Meta-analysis	Higher-education learners, including health-professions evidence	TBL; overlap with Chen and Zhang
ten Cate et al. (2012)	Retrospective comparative study	Medical students taught by near-peer tutors or faculty	Near-peer tutoring; comparison of academic performance
Trullàs et al. (2022)	Scoping review	Undergraduate medical students	PBL

Continued

Vernon & Blake (1993)	Meta-analysis	Medical education learners	PBL; historical evidence base
Yang et al. (2021)	Systematic review/meta-analysis	Classroom learners across educational settings	Testing/quizzing; mechanistic support for retrieval
Yiou & Goodenough (2006)	Educational report	Medical/anatomy students	PBL implementation
Zhang et al. (2024)	Systematic review	Medical education learners	TBL versus PBL; overlap with TBL/PBL reviews
Belezini et al. (2021)	Primary comparative educational study	Undergraduate medical students	Interactive pathology teaching
Lazari et al. (2023a)	Primary educational study	Medical students	Experiential student study groups
Lazari et al. (2023b)	Primary educational study	Medical students	Tele-pathology/image-based learning
Lazaris et al. (2015)	Primary educational study	Medical students	Experiential pathology learning
Lee et al. (2025)	Systematic review	Medical education learners	Student engagement

Two reviewers independently coded the included reports using a prespecified coding framework. The primary coding unit was a distinct learner-facing activity within an intervention, provided that the report described the activity sufficiently to support classification. When an intervention comprised multiple activities, each activity was coded separately; report-level profiles were subsequently derived from the activities described. Instructional labels alone were not considered sufficient evidence for assigning a mechanism or an ICAP mode.

Mechanism indicators were coded as *present* when an explicitly reported intervention feature met the prespecified operational definition, *absent* when the report explicitly indicated that the feature was not used, and *unclear/not reported* when the available description did not permit a defensible determination. ICAP classification was based on the observable learner activity or product: Passive involved receiving information; Active involved manipulating or selecting information without generating a new explanatory product; Constructive involved generating an explanation, prediction, solution, or other new knowledge product; and Interactive required reciprocal contributions through which learners built on one another's ideas.

When activity details were insufficient to support an ICAP classification, the activity was recorded as *not classifiable* rather than assigned a default mode. Such activities were excluded from mode-specific agreement calculations and synthesis, but were not treated as evidence that the corresponding engagement mode was absent. Disagreements were resolved through discussion and consensus. Because the retained coding counts were insufficient to independently verify a numerical inter-rater agreement coefficient, no numerical coefficient is reported. Accordingly, the synthesis distinguishes explicit evidence, explicit absence, and insufficient reporting, and does not interpret unclassifiable activities as evidence against a proposed mechanism or engagement mode.

2.4. Risk of Bias and Methodological Quality

Methodological quality was considered in accordance with the level of appraisal information available in each included report. Because this review synthesized both evidence syntheses and primary educational studies, appraisal information was not uniformly available or directly comparable across all 27 reports. We therefore distinguished between risk-of-bias or methodological-quality judgements reported by the included sources and assessments independently conducted for the present review. No new study-level ratings were assigned where the information required for a defensible appraisal was unavailable. Where reported, methodological-quality judgements were used as interpretive qualifiers during the realist synthesis, informing the strength and certainty with which evidence was considered to support, qualify, or challenge a context-mechanism-outcome (CMO) configuration. They were not treated as independent evidence of mechanism activation, nor were they used to attribute differences in outcomes to methodological quality alone. Given the incomplete and heterogeneous availability of appraisal information, the contribution of risk of bias to individual CMO judgements could not be systematically quantified. This limitation was taken into account when interpreting the strength of the synthesized propositions.

2.5. Realist Synthesis Methods

This review is a realist review in the methodological sense of that term. It specifies, a priori, the configurations of context, intervention resource, mechanism, and outcome—the CMO configurations—through which active-participation teaching is hypothesized to produce learning, and it adjudicates these configurations against evidence drawn from included systematic reviews and meta-analyses.

Generative causation. Realist synthesis treats outcomes as the joint product of mechanism, context, and resource, rather than as the separable effect of an intervention labelled in a particular way. The analytical question is therefore not whether a labelled approach works on average, but which mechanisms are activated, for whom, under which conditions, and through which engagement modes. Instructional labels are treated in this review as family indicators, and the variation observed across labels in headline effects is interpreted as evidence of mechanism variation rather than as statistical noise requiring partition.

Retroduction. A realist review relies on retroduction—the inference, from patterns of outcome in the evidence, of what must be true of the underlying mechanism for those outcomes to occur. Each of the three CMOs in this review is derived from a learning-theory expectation that makes explicit what must be true of the learner for the outcome to occur, rather than being inferred post-hoc from effect estimates. CMO-1 is derived from the cognitive-architecture prediction that freeing cognitive capacity through pre-class preparation, in combination with retrieval and accountability, produces durable knowledge acquisition. CMO-2 is derived from the alignment prediction that learning activities and assessments targeting the same construct produce stronger transfer when also anchored in au-

thentic clinical problems. CMO-3 is derived from the congruence prediction that reciprocal peer explanation at appropriate pitch produces elaborated understanding beyond what is achieved by individual study. The synthesis therefore moves from theoretical expectation to pattern of evidence, rather than from pattern of evidence to post-hoc theory.

Quantitative evidence as realist evidence. Effect estimates, prediction intervals, I^2 statistics, and risk-of-bias judgements drawn from included systematic reviews and meta-analyses are used in this review as realist evidence rather than as data to be re-analyzed; the interpretation and reporting of prediction intervals are informed by methodological literature on meta-analytic heterogeneity and prediction intervals (Borenstein, 2023; IntHout et al., 2016; Mátrai et al., 2026; Siemens et al., 2025). High heterogeneity—most notably the I^2 = 95.2% reported in the flipped-classroom meta-analyses—is informative about variation in mechanism activation across implementations. It is not partitioned here as methodological nuisance but is explained as substantive evidence about under which conditions the mechanism fires. No de novo meta-analysis is undertaken; the synthesis does not re-pool estimates or re-analyze included sources.

Analytical sequence. The synthesis follows a defined sequence. An initial program theory specifying three CMO configurations is articulated a priori (§1.4). Evidence is then extracted against dichotomous mechanism-resource indicators, ICAP engagement mode, intended outcome/activity/assessment for alignment, implementation fidelity, and educational outcomes (§2.3). Supporting and contradictory evidence is identified for each CMO and for the corresponding failure configurations. CMO adjudication proceeds through triangulation of mechanism-profile coding, ICAP mode, intervention components, and effect estimates from the included meta-analyses. Cross-case comparison is undertaken across instructional labels rather than within them, so that mechanism activation is compared rather than labels. Finally, the program theory is refined into demi-regularities (§4.8).

Demi-regularities as middle-range propositions. The demi-regularities derived through this sequence are middle-range propositions about mechanism activation in undergraduate medical education. They are not grand-theory claims; their transferability is conditional on context conditions—preparation required and verified; teaching and assessment targeting applied judgement; peer interaction structured with appropriate pitch. Demi-regularities should therefore be read as falsifiable propositions for prospective primary studies rather than as confirmations of effective practice.

Six mechanisms as analytic scaffold. The six mechanism resources are used as an a priori analytic scaffold for CMO adjudication. They were specified from learning theory rather than arrived at inductively from the evidence. Evidence that did not map onto the six resources was retained and used to qualify, refine, or refute the initial program theory. Iteration therefore retains the configuration where evidence supports it, qualifies it where context-dependent, and revises it where contradicted.

This methodology is reflected consistently across the framing of the research question (§1.2), the extraction variables (§2.3), the appraisal strategy (§2.4), and the criteria used to refine the demi-regularities (§4.8). The methodological identity of the review—realist rather than positivist meta-analytic—is preserved across these elements without need to invoke any additional source.

3. Results

3.1. Overall Findings

Active-participation teaching was associated with better outcomes than passive instruction across the included evidence base, but the reported effects were not directly comparable because the underlying syntheses differed in learner populations, comparators, outcome definitions, study designs, and evidence bases. Reported effects ranged from SMD 0.33 to SMD 2.63 across instructional labels. Mechanism activation, ICAP engagement, constructive alignment, and implementation fidelity are considered possible explanations for some between-study variation, but the cited reviews do not isolate these factors experimentally from differences in populations, comparators, outcomes, faculty involvement, or implementation setting. Accordingly, mechanism activation and fidelity are treated as explanatory hypotheses supported by pattern consistency, not as established independent moderators.

3.2. Flipped Classroom: Variation in CMO-1 Activation

Flipped-classroom effects ranged from SMD 0.33 to 0.93 (Hew & Lo, 2018; Spaić et al., 2025; Hughes & Lyons, 2017). Large syntheses report very high heterogeneity (Spaić et al., 2025). This degree of heterogeneity is consistent with "flipped classroom" representing a family of interventions rather than a uniform treatment, but does not identify which implementation features account for the variation. Differences in learner populations, comparators, outcome measures, curricular setting, and implementation are alternative explanations for the observed variation. Within those constraints, differences in mechanism-activating features—such as pre-class preparation, retrieval opportunities, accountability structures, classroom application, feedback, and assessment alignment—and differences in implementation fidelity provide plausible additional explanations. The evidence does not permit these factors to be separated from the other sources of heterogeneity; therefore, the CMO-1 interpretation is hypothesis-generating rather than a demonstrated causal decomposition.

3.3. TBL: Retrieval and Accountability Configuration

A Chinese medical-education meta-analysis reported an SMD of 2.46 (95% CI 1.53 - 3.40) (Chen et al., 2018), whereas a broader meta-analysis reported approximately d = 0.55 (Swanson et al., 2019). A systematic review comparing team-based learning with problem-based learning found a modest advantage for TBL (SMD 0.37, 95% CI 0.02 - 0.73) (Zhang et al., 2024). These findings are consistent

with the proposition that TBL is more than group work: readiness assurance plausibly activates retrieval and accountability, team discussion supports Constructive or Interactive engagement, and application exercises anchor learning to clinical decisions. Variation across estimates is plausibly explained by differences in CMO-1 mechanism activation rather than by instructional label alone. Differences in inclusion base, comparator structure, source-study risk-of-bias profiles, and curricular context are consistent candidates for the observed gap, and the methodological features that may account for the contrast should be considered across both source syntheses.

3.4. PBL: Context Anchoring and Constructive Alignment

The 2022 scoping review by Trullàs and colleagues concluded that problem-based learning was at least as effective as traditional teaching for knowledge retention and academic performance, while also supporting broader professional competencies (Trullàs et al., 2022). Because effect sizes were not pooled, within-label heterogeneity in PBL implementation and assessment was documented but not quantified. These findings are consistent with constructive alignment: when learners interpret cases, identify learning needs, prioritize problems, and justify decisions, assessments should evaluate these competencies, plausibly supporting CMO-2 when teaching activities and assessment both target applied clinical judgement. Support for CMO-2 appears to have been more indirect than support for CMO-1, because several supporting syntheses assessed knowledge or general academic outcomes rather than clinical reasoning directly.

3.5. PAL: Congruence and Interactive Engagement

In Brierley and colleagues' 21-trial meta-analysis, peer-assisted learning produced an aggregate SMD of 0.52 (0.18 - 0.85; $p = 0.003$), with effects of 0.63 for clinical outcomes and 0.69 for practical outcomes (Brierley et al., 2022). Near-peer tutoring produced a mean effect size of d = 0.23 across 36 course-level contrasts, with 29 positive contrasts (ten Cate et al., 2012). The evidence base on cognitive and social congruence provides a mechanistic basis for interpreting these effects (Loda et al., 2019; Loda et al., 2020; Maggio et al., 2018).

These findings are consistent with the value of peer teaching, but do not establish that all peer-led activities are equivalent. Tutor preparation, supervision, interaction structure, learner needs, and assessment alignment moderated outcomes in ways plausible for CMO-3.

3.6. BOPPPS and Jigsaw: Boundary Conditions and Failure Configurations

Across 40 comparisons with lecture-based learning, BOPPPS produced SMDs of 1.56 (95% CI 1.24 - 1.89) for knowledge, 1.15 (1.00 - 1.30) for skills, and 1.22 (0.85 - 1.59) for comprehensive ability, with higher teaching satisfaction (OR 3.64, 2.97 - 4.46) (Ma et al., 2022). Although these effects are substantial, limited study quality warrants cautious interpretation.

Generalizability flag. The BOPPPS evidence derives largely from a single national context. The largest team-based-learning synthesis is similarly concentrated (Chen et al., 2018), contrasting with more modest international estimates (Swanson et al., 2019). International peer-assisted-learning and problem-based-learning syntheses also report smaller, more heterogeneous effects. Consequently, support for CMO-1 is strongest within the educational structures in which these reports were conducted, and transferability to settings with different learner preparation, faculty development, and curricular structures requires cautious interpretation.

Jigsaw teaching may promote Constructive or Interactive engagement when learners explain content, respond to peers, and integrate shared understanding. However, its evidence base is dominated by quasi-experimental studies and remains hypothesis-generating.

3.7. Realist Synthesis: CMO Adjudication

Across the 27 included reports, all three CMOs received consistent but predominantly indirect support, with findings that were consistent with the proposed mechanism but did not directly measure learner-level processes. The learner responses shown in Table 1 should therefore be interpreted as theory-informed inferences rather than directly observed mechanisms. Support for CMO-2 was more indirect than for CMO-1, because several supporting syntheses assessed knowledge or general academic outcomes rather than clinical reasoning directly. Some of the observed heterogeneity may reflect meaningful variation in intervention design rather than instructional labels alone. Table 2 summarises the CMO evidence matrix.

Table 2. CMO evidence matrix.

CMO	Context	Mechanism/ resource	Learner response (inferred)	Outcome	Supporting evidence	Strength of support
CMO-1	Pre-class preparation required and verified; readiness testing	Working-memory redistribution; retrieval; accountability	Preparatory and retrieval effort	Knowledge acquisition above lecture-only	Flipped syntheses; TBL readiness assurance; testing-effect reviews	Moderate; consistent with the proposed configuration, but mechanism activation is inferred rather than directly measured
CMO-2	Teaching and assessment target applied judgement; authentic clinical problems	Constructive alignment; context anchoring	Coherent effort toward context-specific clinical processing	Clinical reasoning and transfer beyond factual recall	PBL scoping review; BOPPPS	Indirect; consistent with mechanism; clinical reasoning not directly measured in BOPPPS

Continued

CMO-3	Peers explain at an appropriate level within a supportive environment; learners generate, exchange, and respond	Cognitive/social congruence; ICAP-Interactive peer exchange	Reciprocal constructive peer reasoning	Deeper conceptual integration than instructionally identical individual study	PAL meta-analysis; near-peer tutoring (ten Cate et al., 2012); congruence work	Moderate; mechanism plausible from secondary evidence
CMO-1 failure	Preparation weakly required or unverified; no accountability; passive absorption	Intervention retains label but omits activating components	Limited preparatory effort; mechanism not activated	Smaller knowledge gains than label-comparable reports	Within-label variation in flipped classroom	Indirect; hypothesis-level; fidelity is a plausible moderator
CMO-2 failure	Teaching activity and assessment target different constructs	Misaligned activities and assessment	Effort directed toward assessable but unrelated content	Limited transfer/clinical-reasoning gains	Within-label variation in PBL; BOPPPS effects concentrated on knowledge/skill	Indirect; plausible candidate explanation for part of variation
CMO-3 failure	Peer interaction lacks explanation structure	Participatory discussion only	Opinion exchange without elaboration	Shallow conceptual integration	Jigsaw quasi-experimental evidence; PAL heterogeneity	Hypothesis-level
Fidelity failure (cross-cutting)	Intervention retains label; mechanism-activating components omitted or weakly implemented	Any of the six resources	Mechanism not activated	Effect attenuated relative to fidelity-comparable reports	High I^2 = 95.2% in a flipped-classroom synthesis (Spaić et al., 2025); limited fidelity operationalisation across syntheses (Stains & Vickrey, 2017)	Indirect; heterogeneity shows variation, not its cause. Fidelity is a candidate moderator requiring direct testing.

4. Discussion

4.1. Principal Finding

The principal finding of this realist review is that instructional labels do not adequately describe educational interventions. The same label may involve different combinations of retrieval, accountability, clinical application, peer interaction, explanation, and assessment alignment. ICAP is not treated as an additional mechanism here; it specifies the level of learner engagement at which the relevant mechanism may operate. ICAP therefore distinguishes what learners actually produce during an activity: they may receive information passively, manipulate it actively, generate explanations constructively, or engage interactively by responding to and developing one another's contributions. A flipped classroom is not necessarily active merely because students watch content before class; a team activity is not necessarily interactive merely because students sit together.

The realist question examined here is therefore not whether intervention X works, but which mechanisms are activated, for whom, under which conditions, and with what outcomes? Variation in headline effects across labels may become interpretable once mechanisms, ICAP mode, and implementation fidelity are considered.

4.2. What This Study Adds

Rather than proposing another active-learning taxonomy, the review is organized around a different unit of analysis: the mechanism-ICAP configuration rather than the instructional label. Three contributions beyond existing meta-analyses of active-participation teaching are identified (Freeman et al., 2014; Hew & Lo, 2018; Hughes & Lyons, 2017; Trullàs et al., 2022; Brierley et al., 2022). First, it reframes analysis from instructional labels (flipped, TBL, PBL, PAL [Peer Assisted Learning], BOPPPS, Jigsaw) to mechanism-ICAP configurations, providing a theoretically grounded basis for selecting learning activities according to the cognitive engagement they activate. Second, it identifies three CMO-level demi-regularities (retrieval + accountability → knowledge; alignment + context → clinical reasoning; congruence + Interactive peer engagement → conceptual integration) that recur across the evidence base but are obscured in label-level syntheses. Third, it identifies implementation fidelity (Stains & Vickrey, 2017) as a cross-cutting moderator and a plausible contributor to the marked heterogeneity reported in flipped-classroom meta-analyses (Hew & Lo, 2018; Spaić et al., 2025; Hughes & Lyons, 2017), supporting routine fidelity reporting in future reviews. The flipped-classroom literature illustrates this argument: interventions sharing the same label differ substantially in mechanism activation, ranging from configurations combining preparation, retrieval, accountability, application, and aligned assessment to simple lecture replacement with recordings. Consequently, some of the observed heterogeneity (Hew & Lo, 2018; Spaić et al., 2025; Hughes & Lyons, 2017) may reflect meaningful variation in intervention design rather than instructional labels alone, supporting explanation-based evidence synthesis as a plausible complement to label-based synthesis in health-professions education.

4.3. Constructive Alignment and ICAP

Constructive alignment links intended outcomes, teaching activities, and assessment, originating with Biggs and illustrated by Pereira and colleagues (Pereira et al., 2024). ICAP provides the engagement structure: retrieval may operate at Active, Constructive, or Interactive tiers depending on learner output (Pyc & Rawson, 2010; Pan & Rickard, 2018; Adesope et al., 2017; Rowland, 2014; Yang et al., 2021). Cognitive and social congruence support Constructive or Interactive peer engagement (Loda et al., 2019; Loda et al., 2020). ICAP complements, rather than supersedes, sociocultural, workplace-learning, and complexity-informed perspectives, which explain participation within professional communities, organizational cultures, clinical practice, and adaptive educational systems. ICAP is used

in this synthesis to explain cognitive engagement while wider contextual determinants of learning are recognized.

4.4. Practical Implications for Curriculum Design

Curriculum planning should begin with the educational problem students need to solve:

1) Students with factual knowledge but limited capacity for application may need clinical cases and context anchoring.

2) Students demonstrating insufficient preparation may need clearer accountability structures.

3) Students who engage in case discussion but demonstrate poor assessment performance may need improved constructive alignment.

4) Students who are reluctant to ask questions may need stronger cognitive and social congruence.

5) Students who forget previously taught material may need more retrieval opportunities.

The student activity and intended ICAP mode should be defined. The activity design should specify whether students are expected to select information, generate explanations, apply knowledge, critique reasoning, or jointly construct a decision (Daly et al., 2026). Reducing lecture time without replacing it with purposeful learning activity is unlikely to produce the intended change. Shared classroom time should be used for application, explanation, discussion, feedback, and interaction.

Inspired by the mechanism-profile synthesis reported above, the undergraduate Pathology curriculum at the Medical School of the National and Kapodistrian University of Athens was redesigned as a mechanism-aligned, evidence-informed educational framework rather than the adoption of a single pedagogic label. The redesign draws on the active-learning syntheses identified in this review—Ma and colleagues (Ma et al., 2022) and Spaić and colleagues (Spaić et al., 2025) for the integrated-mechanism rationale; Freeman and colleagues (Freeman et al., 2014) for the performance baseline; Burgess, van Diggele, Roberts, and Mellis for institutional-translation guidance; Stains and Vickrey (Burgess et al., 2020; Stains & Vickrey, 2017) for the implementation-fidelity requirement; and Loda and colleagues (Loda et al., 2019) for the cognitive-and-social congruence rationale that motivates the peer-component design.

4.4.1. Mechanism Presence in the Redesigned Curriculum

The redesign operationalises all six mechanisms identified in §1.3:

- **M1 (working-memory redistribution).** Each thematic unit begins with pre-circulated digital whole-slide images and structured reading lists, so that in-class contact hours are reserved for schema application rather than transmissive intake (Sweller et al., 2019; Leppink & van den Heuvel, 2015; Young et al., 2014).

- **M2 (retrieval as a causal learning event).** Course examinations use selected digital whole-slide images accompanied by structured dialogue with the examiner. Students are guided toward a microscopic diagnosis and then asked either to identify lesions in pre-specified fields of view or to locate diagnostically relevant features in response to targeted questions, instrumentally operationalising retrieval rather than passive recognition (Pyc & Rawson, 2010; Pan & Rickard, 2018; Yang et al., 2021).
- **M3 (accountability-driven effort transformation).** Voluntary small-group sessions are organised in a two-session structure, so that preparation for the second session presents a structurally obligatory rather than volitional commitment (Chen et al., 2018; Swanson et al., 2019).
- **M4 (context-anchored encoding).** Plenary teaching is systematically followed by case-based exercises requiring students to integrate histopathological findings with clinical, imaging, and laboratory data in support of diagnostic and therapeutic reasoning. Each thematic unit concludes with discussion of representative clinical cases that illustrate the clinical significance of pathological findings in establishing the final diagnosis, determining disease prognosis, and guiding individualised treatment within the framework of precision medicine.
- **M5 (constructive alignment).** Examination design has been aligned with the cognitive operations classroom teaching elicits rather than retaining transmissive items (Pereira et al., 2024; McCartney, 2014). The whole-slide-image discussion format makes this alignment operational.
- **M6 (social congruence).** Small-group cohorts are constructed with explicit attention to congruence between tutor and learner, mirroring the cognitive-and-social congruence moderators documented by Loda and colleagues (Loda et al., 2019; Loda et al., 2020) and the PAL aggregate reported by Brierley and colleagues (Brierley et al., 2022); SMD 0.52 (0.18 - 0.85; $p = 0.003$).

4.4.2. From Transmissive to Participatory Engagement

Students who have become accustomed to an impersonal model of education—one in which their studies are perceived primarily as a sequence of theoretical courses culminating in equally theoretical examinations that must simply be passed—often experience considerable difficulty participating in meaningful communication with their instructors, principally because of fear of making mistakes. Within an active participatory educational process, however, mistakes are not merely inevitable but essential: learning from them substantially reduces the likelihood of their repetition in clinical practice, where the consequences are more serious. The premise underlying the redesign is that errors made in dialogic engagement with whole-slide images are precisely the kinds of errors whose correction later improves prognosis-related reasoning—analogous to the Vernon and Blake (Vernon & Blake, 1993) clinical-versus-factual divergence, but operating at the level of within-institution curriculum design rather than cross-institution synthesis.

4.4.3. Plenary, Dialogic, and Laboratory Components

Plenary teaching is systematically followed by case-based exercises requiring students to integrate histopathological findings with clinical, imaging, and laboratory data to support diagnostic and therapeutic reasoning. Each thematic unit concludes with discussion of representative clinical cases illustrating the clinical significance of pathological findings in establishing the final diagnosis, determining disease prognosis, and guiding individualized therapeutic decisions within the framework of precision medicine. Students engage in open dialogue with the instructor while examining pathological specimens and histopathological images, correlating morphological findings with the patient's clinical history, symptoms, physical examination findings, imaging studies, and other laboratory investigations. Through this process, students develop an appreciation of the multidisciplinary nature of patient management and the importance of effective collaboration among healthcare professionals, founded on mutual respect for each discipline's expertise and role (Kassab et al., 2023; Lee et al., 2025). Group dynamics in particular significantly influence learning outcomes in health-professions education (Belezini et al., 2021).

In parallel, laboratory sessions provide hands-on opportunities for the macroscopic examination of pathological lesions in surgical specimens. Using educational digital whole-slide images (Lazari et al., 2023a; Lazari et al., 2023b), students examine histopathological lesions simultaneously with the instructor, actively identifying and interpreting microscopic abnormalities in real time. This approach constitutes an authentic simulation-based training experience that closely reflects the routine practice of a contemporary pathology laboratory (Lazaris et al., 2015) (**Figure 3**).

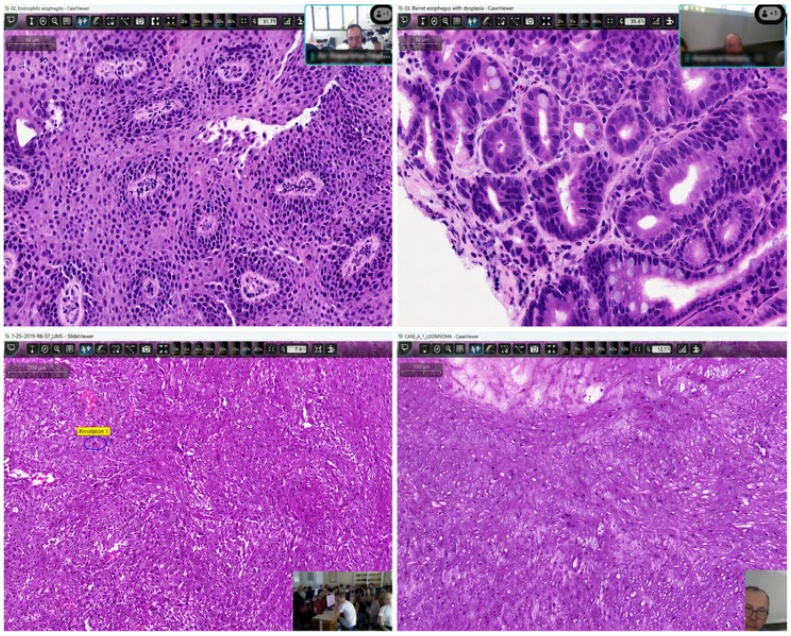


Figure 3. Examples of digital whole-slide images presented during synchronous virtual pathology teaching.

4.4.4. Voluntary Small-Group Sessions: A Jigsaw-Mechanism Translation

Aligned with the PAL evidence base (ten Cate et al., 2012; Brierley et al., 2022) and the principle that PBL remains most effective when used to support integration of foundational science with clinical reasoning rather than short-term knowledge acquisition (Trullàs et al., 2022; Vernon & Blake, 1993) (Secomb, 2008; Yiou & Goodenough, 2006), and consistent with the mechanism profile of Jigsaw cooperative learning documented in two MBBS-Phase-II pathology quasi-experiments (Pahwa et al., 2022)—pre-to-post gains of 11.1 ± 3.50 to 14.57 ± 3.20; 89.3% scoring ≥50%; satisfaction 86.73; n = 141—voluntary small-group pathology sessions were introduced to reinforce collaborative, learner-centred engagement. Each semester, students may voluntarily participate in small-group interactive learning activities focused on selected specimens organised by organ system or disease category (e.g., dermatopathies or breast pathology). During the first session, all student groups undertake an in-depth review of the selected cases under the joint supervision of a pathology-faculty member and a clinician specialising in the corresponding field (e.g., a dermatologist or breast surgeon). During the second session, each student group presents its assigned case, describing the clinical presentation, the patient's diagnostic work-up, the relevant investigations performed, the key pathological features required for differential diagnosis, the demonstration of these findings on the corresponding digital whole-slide image, and the implications for subsequent patient management (**Figure 4**).



Figure 4. Voluntary small-group pathology sessions.

4.4.5. Assessment Alignment (M5)

The interactive-learning principles have been incorporated into student assessment. Part of the course examination is conducted using selected digital whole-slide images accompanied by appropriate clinical case histories. Through a structured discussion with the examiner, students are guided toward establishing the correct microscopic diagnosis. They are subsequently asked either to identify key microscopic lesions and histopathological patterns at specific microscopy fields indicated by the examiner or to independently locate and interpret diagnostically relevant histological features within the digital slide in response to targeted questions. This instrument directly addresses the M5 (constructive-alignment) implementation gap documented in §4.2 and targeted by Stains and Vickrey's (Stains & Vickrey, 2017) implementation-fidelity framework (Pereira et al., 2024; McCartney, 2014). Per the framework a vanced in this review, redesign of classroom teaching without corresponding alignment of assessment items is the principal documented failure mode; this example operationalises the alignment.

The example illustrates, at a single institution, how the six-mechanism framework advanced in this paper translates into a concrete curriculum. The same mechanism profile that accounts for systematic variation in observed effects across the 27-report corpus also supplies the design vocabulary for curriculum committees to audit their own active-learning initiatives. Implementation fidelity (Stains & Vickrey, 2017), institutional support (Graffam, 2007), and rigorous measurement of learner engagement (Kassab et al., 2023; Lee et al., 2025) are the conditions without which mechanism presence does not translate into achieved gains—and the conditions under which this institutional example should be read as a mechanism-aligned rather than merely label-aligned redesign.

4.5. Implementation Fidelity

Implementation fidelity emerged as one plausible, cross-cutting explanation for heterogeneity (Stains & Vickrey, 2017). When the same label is delivered with different degrees of preparation, retrieval practice, accountability, faculty facilitation, classroom application, feedback, and assessment alignment, the enacted intervention may differ despite a shared label. However, fidelity is not isolated in the cited reviews from learner population, comparator, outcome, or setting differences. It is therefore presented as a candidate moderator of mechanism activation, not as a demonstrated independent explanation of the observed effect variation. Future primary studies and reviews should report fidelity explicitly and, where possible, test it alongside population, comparator, and outcome characteristics.

4.6. Choosing Appropriate Outcomes

A method intended to improve clinical reasoning should not be evaluated only through factual recall (Daly et al., 2026). A method intended to improve foundational knowledge should include valid measures of retention. Depending on the intervention, relevant outcomes may include knowledge acquisition, clinical rea-

soning, transfer, practical performance, and learner satisfaction. Satisfaction provides information about acceptability, but it should not be treated as a substitute for evidence of learning.

4.7. Strengths and Limitations

Because the synthesis relied predominantly on systematic reviews and meta-analyses, contextual and mechanistic interpretation was constrained by the level of detail reported in the underlying evidence syntheses; this limitation does not invalidate the review but should be considered when interpreting fine-grained mechanism inference. In particular, the report-level evidence base contains overlapping primary studies, and the available reports did not consistently permit complete reconstruction of those overlaps. In addition, activity-level ICAP coding was necessarily unavailable for reports that did not describe learner activity in sufficient detail, and the original inter-rater κ could not be independently verified from the retained coding counts. These limitations favor qualitative, theory-informed interpretation over claims of independent quantitative moderation by mechanism activation or fidelity.

4.8. Refined Program Theory and Demi-Regularities

Theory refinement proceeded through the following stages: initial program theory, evidence extraction, CMO adjudication, cross-case comparison, and refinement, culminating in three mechanism-level demi-regularities and a cross-cutting fidelity proposition. The CMO configurations were refined into three demi-regularities, while implementation fidelity emerged as a cross-cutting moderator.

1) Where retrieval and accountability are activated, knowledge acquisition tends to improve more than when these mechanisms are weak, regardless of instructional label.

2) Where teaching and assessment align with applied judgement, clinical reasoning tends to improve, whereas factual-recall outcomes may change less.

3) Where peer interaction shifts from Active to Interactive ICAP engagement, conceptual integration is more likely when congruence is supported.

Heterogeneity in flipped-classroom syntheses (Hew & Lo, 2018; Spaić et al., 2025; Hughes & Lyons, 2017) is compatible with fidelity variation as one possible explanation, but does not establish fidelity as the cause of the observed variation (Chen et al., 2018; Stains & Vickrey, 2017). Future studies should report mechanism-ICAP profiles alongside labels, including preparation, retrieval, case-solving, explanation, peer teaching, critique, and aligned assessment. Mechanism-level meta-regression should be reserved for sufficiently large evidence bases.

5. Conclusion

Active-participation teaching generally appears to outperform passive instruction in the included evidence base, but its effectiveness varies substantially. This variability is consistent with—though not reducible to—variation in mechanism acti-

vation, ICAP engagement mode, constructive alignment, and implementation fidelity, alongside differences in learner populations, comparators, outcomes, and settings. The mechanism-ICAP configuration, rather than the instructional label, is offered as a plausible explanatory unit for variation in educational outcomes; labels function as imperfect proxies for the configurations they instantiate. Implementation fidelity is treated as a cross-cutting condition that may influence mechanism activation, not as an independently established moderator. The demi-regularities should therefore be understood as theory-refining propositions requiring prospective empirical testing across diverse undergraduate medical education settings.

Acknowledgements

The authors gratefully acknowledge the support provided by *BioAnalytica S. A.* towards the publication of this article.

Funding

The authors received no specific funding for this study.

Author Contributions

[Anatoli Em. Argyrou] contributed to the conceptualization of the clarification synthesis, design of the search strategy, screening of records, quality appraisal, thematic coding, drafting of the framework, and writing of the manuscript. [Andreas C. Lazaris] contributed to screening, quality appraisal, and critical revision of the manuscript for important intellectual content. All authors approved the final version.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

Ethics Approval

This study is a synthesis of previously published literature. No human participants were recruited and no primary data were collected. Ethical approval was therefore not applicable.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the Use of Tests: A Meta-Analysis of Practice Testing. *Review of Educational Research, 87,* 659-701. https://doi.org/10.3102/0034654316689306

Belezini, E., Katsoulas, N., Thomopoulou, G., & Lazaris, A. C. (2021). The Superiority of Interactive Courses Combined with the Teacher's Physical Presence in the Undergraduate Pathology Curriculum. *Journal of Contemporary Medical Education, 11,* 1-5.

Bordage, G. (2009). Conceptual Frameworks to Illuminate and Magnify. *Medical Education, 43,* 312-319. https://doi.org/10.1111/j.1365-2923.2009.03295.x

Borenstein, M. (2023). How to Understand and Report Heterogeneity in a Meta-Analysis: The Difference between I-Squared and Prediction Intervals. *Integrative Medicine Research, 12,* Article ID: 101014. https://doi.org/10.1016/j.imr.2023.101014

Brierley, C., Ellis, L., & Reid, E. R. (2022). Peer-Assisted Learning in Medical Education: A Systematic Review and Meta-Analysis. *Medical Education, 56,* 365-373. https://doi.org/10.1111/medu.14672

Bucklin, B. A., Asdigian, N. L., Hawkins, J. L., & Klein, U. (2021). Making It Stick: Use of Active Learning Strategies in Continuing Medical Education. *BMC Medical Education, 21,* Article No. 44. https://doi.org/10.1186/s12909-020-02447-0

Burgess, A., van Diggele, C., Roberts, C., & Mellis, C. (2020). Feedback in the Clinical Setting. *BMC Medical Education, 20,* Article No. 460. https://doi.org/10.1186/s12909-020-02280-5

Chen, M., Ni, C., Hu, Y., Wang, M., Liu, L., Ji, X. et al. (2018). Meta-Analysis on the Effectiveness of Team-Based Learning on Medical Education in China. *BMC Medical Education, 18,* Article No. 117. https://doi.org/10.1186/s12909-018-1179-1

Daly, R., Tunney, E., Spooner, M., Offiah, G., Flood, K., & Kent, F. (2026). Case-Based Learning (CBL) in Undergraduate Health Professions Education: A Realist Review. *Medical Education, 60,* 846-864. https://doi.org/10.1111/medu.70179

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H. et al. (2014). Active Learning Increases Student Performance in Science, Engineering, and Mathematics. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 8410-8415. https://doi.org/10.1073/pnas.1319030111

Gordon, M., & Gibbs, T. (2014). STORIES Statement: Publication Standards for Healthcare Education Evidence Synthesis. *BMC Medicine, 12,* Article No. 143. https://doi.org/10.1186/s12916-014-0143-0

Graffam, B. (2007). Active Learning in Medical Education: Strategies for Beginning Implementation. *Medical Teacher, 29,* 38-42. https://doi.org/10.1080/01421590601176398

Grijpma, J. W., Ramdas, S., Broeksma, L., Meeter, M., Kusurkar, R. A., & De la Croix, A. (2024). Learning from the Experts: Stimulating Student Engagement in Small-Group Active Learning. *Perspectives on Medical Education, 13,* 229-238. https://doi.org/10.5334/pme.1245

Hew, K. F., & Lo, C. K. (2018). Flipped Classroom Improves Student Learning in Health Professions Education: A Meta-Analysis. *BMC Medical Education, 18,* Article No. 38. https://doi.org/10.1186/s12909-018-1144-z

Hughes, Y., & Lyons, N. (2017). Does the Flipped Classroom Improve Exam Performance in Medical Education? A Systematic Review. *MedEdPublish, 6,* Article 100. https://doi.org/10.15694/mep.2017.000100

IntHout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for Routinely Presenting Prediction Intervals in Meta-Analysis. *BMJ Open, 6,* e010247. https://doi.org/10.1136/bmjopen-2015-010247

Kassab, S. E., Al-Eraky, M., El-Sayed, W., Hamdy, H., & Schmidt, H. (2023). Measurement of Student Engagement in Health Professions Education: A Review of Literature. *BMC Medical Education, 23,* Article No. 354. https://doi.org/10.1186/s12909-023-04344-8

Lazari, E. C., Mylonas, C. C., Thomopoulou, G. E., Manou, E., Nastos, C., Kavantzas, N. et al. (2023a). Experiential Student Study Groups: Perspectives on Medical Education in the Post-Covid-19 Period. *BMC Medical Education, 23,* Article No. 42. https://doi.org/10.1186/s12909-023-04006-9

Lazari, E., Lazaris, A. C., Manou, E., Agrogiannis, G., Nastos, C., Pikoulis, E. et al. (2023b).

"Starting from the Image": A Tele-Pathology Pre-Graduate Course Aimed at Motivating Medical Students. *Medical Science Educator, 33,* 589-593. https://doi.org/10.1007/s40670-023-01770-7

Lazaris, A. C., Riccioni, O., Solomou, M., Nikolakopoulos, I., Vemmou, E., Karamaroudis, S. et al. (2015). Implementation of Experimental Learning in Pathology: Impact of HIPON Project Concept and Attainment. *International Archives of Medicine, 8,* 1-7. https://doi.org/10.3823/1810

Lee, J. W. Y., Balaji, A., & Bello, F. (2025). *A Systematic Review of Engagement in Medical Education.* Research Square. https://doi.org/10.21203/rs.3.rs-7560693/v1

Leppink, J., & Van den Heuvel, A. (2015). The Evolution of Cognitive Load Theory and Its Application to Medical Education. *Perspectives on Medical Education, 4,* 119-127. https://doi.org/10.1007/s40037-015-0192-x

Loda, T., Erschens, R., Loenneker, H., Keifenheim, K. E., Nikendei, C., Junne, F. et al. (2019). Cognitive and Social Congruence in Peer-Assisted Learning—A Scoping Review. *PLOS ONE, 14,* e0222224. https://doi.org/10.1371/journal.pone.0222224

Loda, T., Erschens, R., Nikendei, C., Giel, K., Junne, F., Zipfel, S. et al. (2020). A Novel Instrument of Cognitive and Social Congruence within Peer-Assisted Learning in Medical Training: Construction of a Questionnaire by Factor Analyses. *BMC Medical Education, 20,* Article No. 214. https://doi.org/10.1186/s12909-020-02129-x

Ma, X., Zeng, D., Wang, J., Xu, K., & Li, L. (2022). Effectiveness of Bridge-In, Objective, Pre-Assessment, Participatory Learning, Post-Assessment, and Summary Teaching Strategy in Chinese Medical Education: A Systematic Review and Meta-Analysis. *Frontiers in Medicine, 9,* Article 975229. https://doi.org/10.3389/fmed.2022.975229

Maggio, L. A., Thomas, A., Chen, H. C., Ioannidis, J. P. A., Kanter, S. L., Norton, C. et al. (2018). Examining the Readiness of Best Evidence in Medical Education Guides for Integration into Educational Practice: A Meta-Synthesis. *Perspectives on Medical Education, 7,* 292-301. https://doi.org/10.1007/s40037-018-0450-9

Mátrai, P., Kói, T., Sipos, Z., & Farkas, N. (2026). Assessing the Properties of the Prediction Interval in Random-Effects Meta-Analysis. *Research Synthesis Methods, 17,* 517-537. https://doi.org/10.1017/rsm.2025.10055

McCartney, M. (2014). Active Learning: The Twilight of Chem 101? *Science, 344,* 1239-1239. https://doi.org/10.1126/science.344.6189.1239-c

Ngoc, P. N., Cheng, C., Lin, Y., Wu, M., Chu, J., & Tang, K. (2020). A Meta-Analysis of Students' Readiness Assurance Test Performance with Team-Based Learning. *BMC Medical Education, 20,* Article No. 223. https://doi.org/10.1186/s12909-020-02139-9

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D. et al. (2021). The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *PLOS Medicine, 18,* e1003583. https://doi.org/10.1371/journal.pmed.1003583

Pahwa, A. R., Dudani, S., Gangadharan, V., & Gulati, R. (2022). Introduction of the Jigsaw Technique of Cooperative Learning in Teaching Pathology to Medical Undergraduates. *CHRISMED Journal of Health and Research, 9,* 252-257. https://doi.org/10.4103/cjhr.cjhr_19_22

Pan, S. C., & Rickard, T. C. (2018). Transfer of Test-Enhanced Learning: Meta-Analytic Review and Synthesis. *Psychological Bulletin, 144,* 710-756. https://doi.org/10.1037/bul0000151

Pereira, E., Nsair, S., Pereira, L. R., & Grant, K. (2024). Constructive Alignment in a Graduate-Level Project Management Course: An Innovative Framework Using Large Lan-

guage Models. *International Journal of Educational Technology in Higher Education, 21,* Article No. 25. https://doi.org/10.1186/s41239-024-00457-2

Pyc, M. A., & Rawson, K. A. (2010). Why Testing Improves Memory: Mediator Effectiveness Hypothesis. *Science, 330,* 335-335. https://doi.org/10.1126/science.1191465

Rowland, C. A. (2014). The Effect of Testing versus Restudy on Retention: A Meta-Analytic Review of the Testing Effect. *Psychological Bulletin, 140,* 1432-1463. https://doi.org/10.1037/a0037559

Secomb, J. (2008). A Systematic Review of Peer Teaching and Learning in Clinical Education. *Journal of Clinical Nursing, 17,* 703-716. https://doi.org/10.1111/j.1365-2702.2007.01954.x

Siemens, W., Borenstein, M., Evrenoglou, T., Meerpohl, J. J., & Schwarzer, G. (2025). Beyond Prediction Intervals in Meta-Analysis: Reporting the Expected Proportion of Comparable Studies with Clinically Relevant Benefit or Harm. *BMC Medical Research Methodology, 25,* Article No. 215. https://doi.org/10.1186/s12874-025-02733-9

Spaić, D., Bukumiric, Z., Rajovic, N., Markovic, K., Savic, M., Milin-Lazovic, J. et al. (2025). The Flipped Classroom in Medical Education: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research, 27,* e60757. https://doi.org/10.2196/60757

Stains, M., & Vickrey, T. (2017). Fidelity of Implementation: An Overlooked Yet Critical Construct to Establish Effectiveness of Evidence-Based Instructional Practices. *CBE—Life Sciences Education, 16,* rm1. https://doi.org/10.1187/cbe.16-03-0113

Swanson, E., McCulley, L. V., Osman, D. J., Scammacca Lewis, N., & Solis, M. (2019). The Effect of Team-Based Learning on Content Knowledge: A Meta-Analysis. *Active Learning in Higher Education, 20,* 39-50. https://doi.org/10.1177/1469787417731201

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review, 31,* 261-292. https://doi.org/10.1007/s10648-019-09465-5

ten Cate, O., van de Vorst, I., & van den Broek, S. (2012). Academic Achievement of Students Tutored by Near-Peers. *International Journal of Medical Education, 3,* 6-13. https://doi.org/10.5116/ijme.4f0c.9ed2

Trullàs, J. C., Blay, C., Sarri, E., & Pujol, R. (2022). Effectiveness of Problem-Based Learning Methodology in Undergraduate Medical Education: A Scoping Review. *BMC Medical Education, 22,* Article No. 104. https://doi.org/10.1186/s12909-022-03154-8

Vernon, D. T., & Blake, R. L. (1993). Does Problem-Based Learning Work? A Meta-Analysis of Evaluative Research. *Academic Medicine, 68,* 550-63. https://doi.org/10.1097/00001888-199307000-00015

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (Quizzing) Boosts Classroom Learning: A Systematic and Meta-Analytic Review. *Psychological Bulletin, 147,* 399-435. https://doi.org/10.1037/bul0000309

Yiou, R., & Goodenough, D. (2006). Applying Problem-Based Learning to the Teaching of Anatomy: The Example of Harvard Medical School. *Surgical and Radiologic Anatomy, 28,* 189-194. https://doi.org/10.1007/s00276-005-0062-z

Young, J. Q., Van Merrienboer, J., Durning, S., & Ten Cate, O. (2014). Cognitive Load Theory: Implications for Medical Education: AMEE Guide No. 86. *Medical Teacher, 36,* 371-384. https://doi.org/10.3109/0142159x.2014.889290

Zhang, W., Wei, J., Guo, W., Wang, Z., & Chen, S. (2024). Comparing the Effects of Team-Based and Problem-Based Learning Strategies in Medical Education: A Systematic Review. *BMC Medical Education, 24,* Article No. 172. https://doi.org/10.1186/s12909-024-05107-9