education sciences

MDPI

Article

Artificial Intelligence-Mediated Reasoning in Higher Education: A Pedagogical Framework from a Preliminary Observational Study

Alberto Melián Ortiz , Elena Chamorro-Rebollo and Raúl Quintana-Alonso *

Salus Infirmorum Nursing and Physiotherapy Faculty, Pontifical University of Salamanca, 28015 Madrid, Spain; amelianor@upsa.es (A.M.O.); echamorrore@upsa.es (E.C.-R.)

* Correspondence: rquintanaal@upsa.es

Abstract

Background: Artificial intelligence (AI) is increasingly incorporated into higher education. However, most empirical studies focus on technological adoption or learner satisfaction rather than on how pedagogical design, perceived learning impact, and student experience interact within AI-mediated learning environments. Understanding these relationships is essential to determine whether AI supports higher-order reasoning processes rather than merely increasing technological engagement. Objective: This preliminary study aimed to develop and evaluate a theory-driven AI-mediated pedagogical framework and examine relationships between pedagogical design, perceived learning impact, and student satisfaction in a university learning context. Methods: An observational educational evaluation was conducted during implementation of an AI-mediated instructional framework in an undergraduate physiotherapy course. The full academic cohort (n = 22) completed a 24-item questionnaire assessing seven pedagogical domains on a 5-point Likert scale. Descriptive statistics, Pearson correlations, exploratory regression modeling, and factor analysis were used to examine relationships among domains. Academic performance indicators were summarized descriptively. Results: Students reported high evaluations across all domains (means > 4.5/5). The strongest association with satisfaction was perceived learning impact (r = 0.79, $p < 0.001$). Moderate correlations were found for usability (r = 0.66), AI content quality (r = 0.61), pedagogical coherence (r = 0.58), critical thinking (r = 0.52), and ethical integration (r = 0.47). Academic pass rates exceeded 90%. Conclusions: Perceived learning impact emerged as the central mechanism linking AI-mediated instructional design to student satisfaction, suggesting that the educational value of AI depends on alignment with cognitively demanding learning processes.

Keywords: education; professional; students; health occupations; clinical competence; artificial intelligence; problem-based learning

Check for updates

Academic Editors: Marta Montenegro-Rueda and José Fernández-Cerero

Received: 7 March 2026

Revised: 19 March 2026

Accepted: 20 March 2026

Published: 26 March 2026

1. Introduction

Artificial intelligence is increasingly embedded in the epistemic and operational fabric of contemporary healthcare. Predictive analytics, machine learning-based diagnostics, and generative models are reshaping how information is processed, interpreted, and translated into clinical action (Al Kuwaiti et al., 2023; Goralski & Tan, 2020). Beyond technological efficiency, AI represents a shift in the distribution of cognitive labor between human professionals and algorithmic systems. This transformation carries profound implications for health profession education, where the development of professional judgment, contextual interpretation, and ethical accountability remains foundational.

Physiotherapy education offers a particularly relevant context in which to examine this question. Clinical reasoning in physiotherapy requires iterative hypothesis testing, integration of biomechanical and psychosocial factors, and adaptive decision-making in response to patient-specific variability. Educational strategies such as problem-based learning have demonstrated effectiveness in promoting analytic depth and learner motivation (Gunn et al., 2012; D'Sa, 2015). Structured case triggers and collaborative scenario design further support reflective engagement and conceptual transfer (Roche et al., 2016). These approaches align with contemporary models of cognitive development that emphasize analysis, evaluation, and creative problem-solving as central educational outcomes (Anderson & Krathwohl, 2001). Recent studies have also begun to explore the integration of artificial intelligence and large language models within physiotherapy education, including AI-assisted clinical reasoning training and simulated case analysis (Ferrer-Peña et al., 2025; Ergezen Sahin et al., 2025). Conceptual analyses further suggest that generative AI may support curriculum development, learning resource generation, and clinical reasoning practice in physical therapy training (Naqvi et al., 2024).

Recent analyses have highlighted the growing role of artificial intelligence in advancing system-level goals, including sustainability, accessibility, and scalability of health services (Palliyil et al., 2025; Ametepey et al., 2024). However, the educational implications of this shift are more complex than technological adoption alone. Preparing future clinicians for AI-integrated practice requires more than technical familiarity. It demands the cultivation of critical engagement, algorithmic literacy, and reflective reasoning within environments where machine-generated outputs are readily available.

Within educational settings, AI technologies have been introduced through adaptive tutoring systems, conversational agents, and automated feedback platforms. These systems are frequently evaluated in terms of usability, efficiency, and learner satisfaction (Çayır, 2023; Choi & Drumwright, 2021). Acceptance research suggests that perceived usefulness and trust significantly influence attitudes toward AI adoption (Albarrán Lozano & Molina, 2021). Yet satisfaction metrics do not necessarily capture deeper epistemic outcomes. In clinical disciplines, the central question is not whether students appreciate technological support but whether such support enhances or attenuates higher-order reasoning processes (Almansour & Alfhaid, 2024; Shorey et al., 2024).

The integration of generative AI into such pedagogical traditions introduces both opportunity and tension. AI systems are capable of generating simulated patient narratives, proposing diagnostic alternatives, and facilitating structured Socratic dialogue, potentially supporting reasoning scaffolds in complex cases (Hao et al., 2025). At the same time, concerns persist regarding cognitive offloading, reduced effortful processing, and dependency on algorithmic suggestions. Research on technology-mediated engagement suggests that sustained reliance on digital systems can alter patterns of cognitive investment and attention (Ngai, 2012). In clinical education, where professional responsibility and patient safety are paramount, uncritical integration of AI risks undermining the very competencies it aims to enhance.

Emerging scholarship in health professions education argues that artificial intelligence should be conceptualized within broader curricular transformation rather than as a peripheral technological add-on (Almansour & Alfhaid, 2024). Despite this recognition, empirical investigations often examine isolated tools or discrete learning outcomes without modeling the structural relationships among pedagogical components. The extent to which usability, pedagogical coherence, and perceived cognitive impact interact to shape overall learner experience remains insufficiently explored. Understanding these relationships is essential for determining whether AI-mediated instruction strengthens clinical reasoning or merely augments instructional convenience.

Against this background, there is a need for theory-driven evaluations that move beyond descriptive perception studies and examine the internal architecture of AI-mediated learning environments. Such evaluations should consider not only student endorsement but also the interplay between instructional design, technological interaction, and perceived cognitive development.

The aim of this study was to develop and empirically evaluate a theory-driven artificial intelligence-mediated clinical reasoning framework in undergraduate physiotherapy education and examine the structural relationships among pedagogical domains, perceived learning impact, and student satisfaction.

2. Materials and Methods

2.1. Study Design and Reporting Framework

This study was conducted as an educational evaluation of an AI-mediated clinical reasoning framework implemented in an undergraduate physiotherapy course. The evaluation followed an observational design with a mixed-methods orientation, combining descriptive quantitative analysis of student survey responses and academic outcomes with complementary qualitative educational evidence derived from systematic teaching observation and performance during simulated clinical activities. Given the observational nature of the evaluation and the inclusion of cross-sectional post-intervention survey data, the reporting structure was aligned with STROBE recommendations where applicable. To enhance replicability of the educational intervention, the instructional design was mapped to the TIDieR framework for intervention description.

2.2. Setting and Participants

The intervention was integrated into routine teaching of the course Physiotherapy of the Locomotor System (FAL) during the 2024/2025 academic year in a university setting supported by a virtual learning environment. The evaluation sample comprised the full cohort of students enrolled in the course who completed the post-intervention satisfaction questionnaire ($n = 22$). No personally identifiable demographic data were collected, consistent with the minimal-risk nature of the educational evaluation and the anonymized survey procedure.

2.3. Educational Intervention: IA-LOCOM AI-Mediated Framework

The IA-LOCOM framework was designed as a multi-layered pedagogical architecture integrating generative AI within established active learning strategies. The framework combines three sequential instructional components: (1) flipped classroom preparation supported by AI-generated learning materials, (2) in-class active learning through AI-mediated Socratic dialogue and collaborative problem-based learning activities, and (3) a final integrative project in which students applied clinical reasoning to complex simulated cases using multiple AI systems. This structure was intended to ensure that AI tools functioned as scaffolding for reasoning processes rather than as isolated technological aids.

Layer 1: Pre-class preparation (flipped classroom).

Students accessed course materials in advance through the virtual campus, including: (a) instructor-prepared PDF content and guided readings, (b) an approximately 17-min podcast generated from course materials using a dedicated tool (NotebookLM), (c) an approximately 5 min video summary generated via the same tool, and (d) an AI-developed formative self-assessment resource (a "Trivial"-style quiz with 50 questions, built with Vercel v0).

Layer 2: In-class active learning mediated by LLMs.

Each in-person session followed a structured and consistent sequence.

Consolidation of pre-class learning (≈30 min): An initial plenary discussion was conducted to clarify questions arising from pre-class materials and reinforce core conceptual understanding.

AI-assisted Socratic dialogue: Students worked in small groups using a master prompt developed by the teaching team. LLM tools (e.g., ChatGPT (GPT-4), Gemini (gemini-2.5-pro), or Perplexity (Perplexity-Sonar)) were instructed to adopt a Socratic tutoring role, guiding students through iterative questioning rather than delivering direct explanatory responses. The dialogue process continued until students demonstrated adequate conceptual and technical reasoning, as verified by the instructor.

Collaborative clinical simulation using PBL and LLMs (≈45 min): Students were divided into two collaborative groups, each assigned a different LLM tool to analyze the same clinical case through a structured prompt. A key objective of this activity was to examine similarities and divergences in reasoning processes and conclusions generated by different AI systems. The session concluded with a whole-class synthesis aimed at reaching consensus on functional diagnosis and therapeutic planning.

Before the LLM-mediated activities, the first two sessions were devoted to (a) introducing the pedagogical logic and assessment system and (b) delivering a workshop on generative AI literacy, focusing on capabilities, limitations, bias/risk recognition, ethical use, and academic integrity to ensure that AI support did not replace students' own reasoning.

Layer 3: Final integrative project.

At the end of the course, students completed a final integrative project designed to synthesize learning across topics. The cohort developed a simulated "conference" on course pathologies, supported by explicit production guidelines and assessment rubrics defined by the teaching team. Students developed dissemination materials (e.g., posters) using creative AI tools (e.g., Canva AI and Freepik AI) after prompt development in an LLM and participated in structured oral presentations with explicit requirements for ethical and transparent AI use (including documentation of how AI was used).

2.4. Outcomes and Instruments

2.4.1. Academic Performance Indicators

Course assessment was continuous and comprised three graded components: problem-based learning (PBL) and collaborative work (25%), final integrative project (45%), and a written examination (30%). Overall performance and component-specific outcomes were summarized descriptively.

2.4.2. Student Satisfaction and Perceived Pedagogical Impact

A purpose-built satisfaction questionnaire was administered post-intervention via QR code. The instrument contained 24 items distributed across seven predefined pedagogical domains: (1) AI content quality, (2) pedagogical coherence, (3) usability/experience, (4) perceived learning impact, (5) critical thinking/clinical reasoning, (6) ethical integration, and (7) overall satisfaction. Items were rated on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree).

To minimize potential response bias associated with instructor-led interventions, the questionnaire was completed anonymously using a QR-based digital survey system, and no personally identifiable information was collected. Students were informed that participation was voluntary and that their responses would not influence course grading or academic evaluation. In addition, the instrument assessed multiple pedagogical domains related to learning experience rather than direct evaluation of the instructor, which helps reduce the risk of social desirability bias.

2.5. Complementary Qualitative Educational Evidence

Qualitative evidence derived from systematic teaching observation and students' performance during simulated clinical activities was used to contextualize and interpret quantitative findings (e.g., patterns of engagement, autonomy, and reasoning quality during classroom tasks).

2.6. Data Coding and Statistical Analysis

All questionnaire responses were exported from the digital survey platform into IBM SPSS Statistics version 27.0 for analysis. Items were measured on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Prior to analysis, data were screened for completeness and accuracy. Given the anonymized data collection procedure and the absence of missing responses within submitted surveys, complete case analysis was conducted.

Items were conceptually grouped into predefined pedagogical domains based on the theoretical structure of the AI-mediated framework. Domain scores were calculated as the arithmetic mean of the items comprising each construct, preserving the original response scale. This approach was selected to maintain interpretability and comparability across domains while reducing dimensionality for modeling purposes. Although individual Likert items are ordinal, composite scores derived from multiple Likert items can approximate interval-level measurement when aggregated across domains. This approach is widely accepted in educational and psychological research when items measure related constructs and distributions do not substantially deviate from normality (Norman, 2010). Accordingly, parametric techniques were considered appropriate for exploratory analysis of relationships among pedagogical domains in this pilot study.

Descriptive statistics were computed for each domain, including mean, standard deviation, median, interquartile range, minimum, and maximum values. Distributional properties were examined using graphical inspection and Shapiro–Wilk testing. Given the approximately symmetric distributions and the use of composite domain scores, parametric analyses were deemed appropriate.

Bivariate associations between pedagogical domains and overall satisfaction were assessed using two-tailed Pearson correlation coefficients. Effect sizes were interpreted according to conventional thresholds. Ninety-five percent confidence intervals for correlation coefficients were calculated using Fisher's z transformation to enhance precision in estimation.

Given the pilot and preliminary nature of the study, multivariable modeling was conducted for exploratory purposes and interpreted cautiously.

To explore potential independent patterns of association with satisfaction, multiple linear regression analysis was performed with satisfaction as the dependent variable and all remaining domains entered simultaneously as predictors. Assumptions of linearity, homoscedasticity, independence of errors, and absence of multicollinearity were assessed through residual diagnostics and variance inflation factors. Model fit was evaluated using R^2, adjusted R^2, and the overall F statistic. Unstandardized regression coefficients, standard errors, *p* values, and 95% confidence intervals are reported.

Given the limited sample size, a post-hoc power analysis was conducted for the regression model. Assuming a moderate effect size ($f^2 = 0.15$), an alpha level of 0.05, and six predictors, the estimated statistical power was low ($\approx$0.16), indicating insufficient sensitivity to detect small to moderate effects. Furthermore, the ratio of observations to predictors was below commonly recommended thresholds for stable multivariable estimation. Accordingly, the regression analysis was treated as exploratory and hypothesis-generating rather than confirmatory.

An exploratory factor analysis was conducted to examine the latent structure of the questionnaire and evaluate alignment between empirical response patterns and the theoretical framework. Principal axis factoring was applied, given the exploratory nature of the analysis. Factor retention was guided by eigenvalues greater than one and inspection of the scree plot. Given the pilot sample size, factor solutions were interpreted cautiously and considered hypothesis-generating rather than confirmatory.

All statistical tests were two-tailed. Statistical significance was defined as $p < 0.05$. Emphasis was placed on effect sizes and confidence intervals to support interpretation beyond dichotomous significance testing.

2.7. Participant Characteristics

A total of 22 third-year undergraduate physiotherapy students completed the post-intervention evaluation in full. Data collection was conducted anonymously via a QR-based digital survey to maximize confidentiality and minimize response bias. No individually identifiable demographic variables were collected, in line with the study's minimal-risk educational design.

The sample therefore represents a complete academic cohort exposed to the AI-mediated pedagogical framework during a full course implementation.

2.8. Descriptive Outcomes Across Pedagogical Domains

Students reported consistently high ratings across all evaluated pedagogical domains, with mean scores exceeding 4.5 on a 5-point Likert scale. The highest evaluations were observed for:

Overall satisfaction
Ethical integration of artificial intelligence
Pedagogical coherence of the instructional design

Domains related to higher-order cognitive processes—such as clinical reasoning and critical thinking—also demonstrated strong ratings, albeit with slightly greater dispersion, suggesting individual variability in perceived cognitive gains (Table 1).

Table 1. Descriptive statistics across pedagogical domains.

Domain	Mean	SD	Median	IQR	Min	Max
AI Content Quality	4.64	0.33	4.75	0.5	3.75	5.0
Pedagogical Coherence	4.7	0.29	4.75	0.5	4.0	5.0
Usability	4.61	0.39	4.67	0.67	3.33	5.0
Perceived Learning Impact	4.65	0.3	4.75	0.5	4.25	5.0
Critical Thinking	4.53	0.29	4.5	0.5	4.0	5.0
Ethical Integration	4.71	0.31	4.75	0.5	4.0	5.0
Satisfaction	4.77	0.26	4.83	0.33	4.33	5.0

Note. Scores range from 1 to 5. Higher scores indicate stronger agreement. IQR = interquartile range.

2.9. Bivariate Associations with Satisfaction

Pearson correlation analyses were conducted to explore associations between pedagogical domains and overall satisfaction. Effect sizes were interpreted according to conventional thresholds (small ≥ 0.10, moderate ≥ 0.30, large ≥ 0.50).

The strongest association emerged between Perceived Impact on Learning and Overall Satisfaction (r = 0.79, $p < 0.001$; 95% CI 0.55–0.91), indicating a large effect size.

Moderate-to-large associations were also observed for:

Usability (r = 0.66, $p = 0.001$; 95% CI 0.33–0.85)
AI Content Quality (r = 0.61, $p = 0.002$; 95% CI 0.25–0.82)
Pedagogical Coherence (r = 0.58, $p = 0.004$; 95% CI 0.21–0.80)
Critical Thinking (r = 0.52, $p = 0.012$; 95% CI 0.13–0.77)
Ethical Integration (r = 0.47, $p = 0.028$; 95% CI 0.06–0.74)

These findings indicate that satisfaction was primarily associated with perceived cognitive benefit, rather than technological novelty alone (Table 2).

Table 2. Pearson correlations between pedagogical domains and satisfaction.

Domain	r	*p* Value	95% CI Lower	95% CI Upper
AI Content Quality	0.61	0.002	0.25	0.82
Pedagogical Coherence	0.58	0.004	0.21	0.8
Usability	0.66	0.001	0.33	0.85
Perceived Learning Impact	0.79	0.0	0.55	0.91
Critical Thinking	0.52	0.012	0.13	0.77
Ethical Integration	0.47	0.028	0.06	0.74

Note. Two-tailed Pearson correlations. Confidence intervals computed using Fisher's z transformation.

2.10. Multivariable Modeling of Satisfaction

To explore potential independent associations between pedagogical domains and satisfaction within this preliminary sample, a multiple linear regression model was estimated with satisfaction as the dependent variable and all domains entered simultaneously as predictors.

The model accounted for 39.4% of the variance in satisfaction ($R^2 = 0.394$; adjusted $R^2 = 0.151$). Although the overall F-test did not reach statistical significance ($F(6, 15) = 1.62$, $p = 0.208$), this result must be interpreted in the context of limited statistical power inherent to pilot samples and the restricted variance observed across high-scoring domains.

When examining individual predictors:

Pedagogical coherence demonstrated a near-significant association (B = −0.429, SE = 0.208, $p = 0.057$), suggesting potential suppression effects due to intercorrelations among domains.
Critical thinking showed a positive trend (B = 0.329, SE = 0.189, $p = 0.103$).

Remaining predictors did not demonstrate independent significance after adjustment.

Importantly, the attenuation of effects under simultaneous adjustment likely reflects conceptual collinearity among highly rated pedagogical constructs rather than absence of substantive relationships (Table 3).

Table 3. Multiple linear regression predicting satisfaction.

Predictor	B	SE	*p* Value	95% CI Lower	95% CI Upper
Constant	4.912	1.842	0.02	0.902	8.922
AI Content Quality	0.082	0.195	0.688	−0.333	0.498
Pedagogical Coherence	−0.429	0.208	0.057	−0.872	0.014
Usability	−0.089	0.161	0.587	−0.432	0.254
Perceived Learning Impact	0.154	0.275	0.585	−0.431	0.739
Critical Thinking	0.329	0.189	0.103	−0.074	0.732
Ethical Integration	−0.132	0.215	0.553	−0.591	0.327

Note. $R^2 = 0.394$; Adjusted $R^2 = 0.151$; $F(6, 15) = 1.62$; $p = 0.208$.

2.11. Exploratory Structural Analysis Supporting a Multidimensional Framework

An exploratory factor analysis was conducted to examine whether the observed response pattern aligned with the theoretical structure of the AI-mediated framework. Given the pilot nature and sample size, results are interpreted as preliminary.

The solution suggested a coherent multidimensional configuration broadly consistent with three latent domains:

Cognitive Domain (learning impact and reasoning enhancement)
Technological Experience Domain (usability and interaction with AI tools)
Pedagogical-Ethical Domain (coherence, alignment, and responsible AI integration)

This structural convergence between empirical item clustering and instructional design strengthens the conceptual validity of the IA-LOCOM framework as an integrated pedagogical model rather than a collection of isolated technological interventions.

Academic Performance as External Educational Indicator

Academic performance outcomes following the intervention demonstrated high achievement levels, with pass rates exceeding 90% across all assessment components and a strong overall mean course grade (Table 4).

Table 4. Academic performance following implementation of the AI-Mediated Framework.

Assessment Component	Pass Rate	Mean $\pm$ SD
Problem-Based Learning & Collaborative Work	100%	8.7 ± 0.6
Final Integrative Project	95.5%	8.3 ± 0.8
Written Examination	90.9%	7.6 ± 1.0
Overall Course Result	95.5%	8.2 ± 0.7

Note. Course grades are reported on a 0–10 scale.

Although no causal inference is implied, the convergence between high perceived learning impact and strong academic outcomes provides preliminary external validity for the framework.

3. Discussion

Artificial intelligence is increasingly embedded in health profession education, yet empirical work has largely focused on tool adoption rather than on the pedagogical architecture through which AI reshapes learning processes. This pilot study advances the field by evaluating a theory-driven, multidimensional AI-mediated framework implemented within undergraduate physiotherapy education. Rather than examining isolated technological features, we modeled the structural relationships between pedagogical coherence, usability, perceived learning impact, and student satisfaction.

The principal finding is that perceived learning impact demonstrated the strongest association with satisfaction, with a large effect size. This suggests that students' acceptance of AI-mediated learning environments is fundamentally anchored in cognitive value rather than technological appeal. The magnitude of this association exceeds what is commonly reported in perception-based studies of AI in medical education, where technological novelty and ease of use often dominate evaluative metrics (Safranek et al., 2023; Jackson et al., 2024). In contrast, our data indicate that satisfaction is primarily driven by perceived enhancement of reasoning processes.

A potential concern in technology-enhanced learning environments is the novelty effect, whereby favorable student responses may partly reflect initial enthusiasm toward newly introduced technologies rather than genuine pedagogical benefit. Several features of the IA-LOCOM framework were designed to mitigate this risk. First, before the imple-

https://doi.org/10.3390/educsci16040514

mentation of AI-supported activities, students participated in dedicated sessions on AI literacy addressing capabilities, limitations, ethical considerations, and responsible use. Second, AI tools were embedded within established pedagogical strategies, including flipped classroom preparation, Socratic dialogue, and problem-based learning, so that the primary focus remained on reasoning processes rather than on the technology itself. Third, multiple AI systems were used in parallel during case-based activities to encourage comparison of reasoning outputs across tools rather than attachment to a single platform. Nevertheless, because this was a pilot educational evaluation without a control group, the potential influence of novelty effects cannot be completely excluded.

The strength of the relationship between perceived learning impact and satisfaction aligns with emerging randomized evidence demonstrating that AI-supported learning can improve clinical reasoning and higher-order thinking. For example, ChatGPT-assisted problem-based learning has shown measurable gains in reasoning performance in controlled settings (Hui et al., 2025), and meta-analytic evidence suggests that generative AI tools can enhance learning outcomes under structured implementation (J. Wang & Fan, 2025). However, most prior studies evaluate performance outcomes or perception metrics independently. Our contribution lies in demonstrating how these dimensions interact within a unified explanatory model.

When all pedagogical domains were entered simultaneously into the multivariable regression model, individual effects were attenuated, and the global model did not reach statistical significance. Given the modest sample size of this pilot cohort and the number of predictors included, limited statistical power and potential overfitting represent plausible explanations for this pattern. At the same time, moderate-to-high correlations observed between several pedagogical domains suggest conceptual overlap among constructs such as pedagogical coherence, usability, and perceived learning impact. Consequently, the regression results should be interpreted as exploratory and hypothesis-generating rather than as evidence of independent causal relationships.

From a pedagogical perspective, the observed pattern may also reflect interdependence among domains within an integrated learning environment. When pedagogical coherence, usability, and cognitive engagement are all rated highly, their explanatory variance may overlap. In this context, attenuation of individual predictors under simultaneous modeling does not necessarily imply the absence of meaningful relationships but rather the presence of interacting pedagogical dimensions. Similar patterns have been described in AI-supported tutoring environments, where learning outcomes emerge from the interaction between instructional design and adaptive feedback mechanisms (Kestin et al., 2025; Möller et al., 2024).

The exploratory structural analysis further supports this interpretation. The emergent multidimensional configuration suggests that students differentiate between technological interaction, pedagogical alignment, and cognitive development. This convergence between theoretical design and empirical structure strengthens the conceptual validity of the framework and positions it beyond descriptive innovation reporting.

During implementation of the IA-LOCOM project, the course instructor conducted continuous and systematic observation of student performance across active learning activities, particularly within problem-based learning sessions, clinical simulations, and AI-mediated Socratic dialogue. From this qualitative perspective, informed by the instructor's prior teaching experience in previous iterations of the course, an appreciable improvement in the quality of students' clinical reasoning was observed.

Although overall evaluations were consistently high, some variability was observed in domains such as usability and critical thinking. Informal classroom observations suggested that this variability was partly associated with differences in students' familiarity with generative AI tools and their confidence in interacting with large language models. A small subset of students initially required additional guidance to formulate effective prompts or critically evaluate AI-generated responses during case discussions. These observations highlight the importance of explicit AI literacy training and instructor facilitation when integrating generative AI into learning environments.

From a subjective instructional standpoint, students demonstrated greater ability to formulate well-grounded diagnostic hypotheses, justify therapeutic decisions based on biomechanical and clinical reasoning, and articulate their thinking processes in a structured manner. Participation in clinical discussions appeared more reflective, with reduced reliance on memorized responses compared with prior cohorts in which clinical reasoning competencies were also assessed, but through more traditional methodological approaches.

Although these observations do not constitute quantitative evidence, they provide a relevant qualitative indicator suggesting that the combined use of active methodologies and guided artificial intelligence within IA-LOCOM may have facilitated deeper, more explicit, and more transferable reasoning processes in clinical contexts. These preliminary insights warrant confirmation through future comparative studies incorporating objective clinical performance metrics.

3.1. Theoretical Implications for Clinical Reasoning and Higher-Order Cognition

Clinical reasoning remains a central epistemic objective in physiotherapy education. It requires integration of domain knowledge, contextual interpretation, hypothesis generation, and reflective evaluation under conditions of uncertainty. The present findings suggest that students perceive AI-mediated instruction as cognitively meaningful when it supports these higher-order processes rather than merely accelerating information retrieval.

The strong association between perceived learning impact and satisfaction indicates that students respond positively when AI use aligns with reasoning development. This aligns with evidence from randomized trials demonstrating improvements in reasoning performance following AI-supported clinical learning interventions (Ergezen Sahin et al., 2025; Hui et al., 2025). However, the current study extends this understanding by examining the internal structure of learner perception. Rather than measuring outcome gains alone, it evaluates how students cognitively situate AI within their reasoning process.

The exploratory structural findings suggest differentiation between technological interaction and cognitive development. This distinction is theoretically significant. In health profession education, superficial engagement with digital tools may enhance efficiency but does not necessarily promote epistemic growth. The present data support the interpretation that when AI-mediated dialogue and simulation are integrated within structured pedagogical scaffolding, learners interpret the technology as contributing to analytic depth rather than replacing cognitive effort.

This interpretation resonates with contemporary work on AI-supported tutoring systems, demonstrating that adaptive feedback can enhance higher-order thinking when embedded within meaningful problem contexts (Kestin et al., 2025; Möller et al., 2024). It also aligns with meta-analytic evidence indicating that generative AI has greater educational impact when structured within cognitively demanding learning tasks (J. Wang & Fan, 2025). The framework evaluated in this study appears to operate within this paradigm, emphasizing dialogic reasoning and clinical scenario analysis rather than answer generation alone.

From a cognitive perspective, the findings suggest that AI-mediated instruction may function as a scaffold for metacognitive regulation. Students reported strong perceived gains in reasoning while maintaining high ratings of ethical integration and pedagogical coherence. This pattern indicates that the technology was experienced as augmenting reflective thinking rather than automating decision-making. Such integration is essential in clinical domains where professional judgment cannot be outsourced.

3.2. Methodological Contribution and Structural Modeling Implications

Beyond content findings, this study contributes methodologically by modeling relationships among pedagogical domains within an AI-mediated educational framework. Much of the existing literature in AI and medical education relies on either descriptive surveys or experimental comparisons of performance outcomes (Jackson et al., 2024; Sallam, 2023). While valuable, these approaches often treat technological acceptance, cognitive benefit, and satisfaction as independent endpoints.

The present approach adopts a structural perspective. By examining correlations and multivariable associations among domains, the study conceptualizes AI integration as a system of interdependent pedagogical components. The attenuation of predictors under simultaneous regression adjustment should not be interpreted as absence of effect. Rather, it reflects conceptual overlap among highly rated domains within a coherent instructional design. When usability, pedagogical coherence, and cognitive engagement are all strong, their explanatory variance converges. This convergence is characteristic of integrated learning environments and supports the interpretation that educational effectiveness arises from interaction among components rather than from isolated variables.

The exploratory factor structure further strengthens this systemic interpretation. The alignment between emergent latent domains and the theoretical design of the intervention provides preliminary construct validation of the framework. Even within a pilot cohort, the structural coherence of responses suggests that students cognitively differentiate between technological interface, pedagogical alignment, and reasoning development. This internal organization supports the plausibility of future confirmatory modeling in larger samples.

Importantly, the study demonstrates that meaningful structural analysis can be conducted even within small pilot cohorts when effect sizes are substantial and variance is interpretable. Although statistical power limits definitive causal claims, the modeling strategy establishes a foundation for prospective, multicenter investigations designed to test mediated pathways formally. In this respect, the contribution of the present work lies in its empirical findings and its analytic architecture.

3.3. Ethical Integration and Trust in AI-Mediated Learning

The high ratings observed for ethical integration suggest that students did not perceive the AI-mediated framework as epistemically threatening. This finding is particularly relevant given the growing concerns surrounding overreliance, bias propagation, and erosion of professional judgment associated with generative AI tools in healthcare (Choung et al., 2023; Das et al., 2023). In many contexts, student enthusiasm toward AI coexists with uncertainty regarding trust, authorship, and accountability (Sallam, 2023; Shorey et al., 2024). The present results indicate that explicit pedagogical framing of AI use may mitigate such tensions.

Trust in AI systems is shaped by perceived alignment with professional norms as well as by technical performance (Gesk & Leyer, 2022). When AI is introduced as a dialogic and reflective aid rather than as an authority, learners appear more willing to integrate it constructively into their reasoning process. The strong association between cognitive impact and satisfaction, alongside high ethical ratings, suggests that responsible implementation can support engagement while preserving critical distance. This balance is particularly important in physiotherapy, where clinical reasoning relies heavily on contextual interpretation and patient-specific decision making.

Qualitative research in physiotherapy education has identified ambivalence toward generative AI, with students expressing both excitement and apprehension regarding its influence on professional identity (Lindbäck et al., 2025). The present findings provide preliminary evidence that structured integration accompanied by ethical discussion may stabilize this ambivalence. When students understand the boundaries of AI capability and the necessity of human oversight, the technology may be experienced as an augmentation of expertise rather than a substitute for it.

3.4. Implications for Health Sciences Education

The findings of this pilot study have broader implications for health sciences education beyond physiotherapy. First, they reinforce the importance of aligning AI implementation with clearly articulated cognitive objectives. Educational benefit appears to emerge when AI tools are embedded within pedagogical frameworks that emphasize reasoning, reflection, and dialogue. This is consistent with evidence that problem-based and case-based approaches enhance motivation and deeper learning in health professions training (Abugri et al., 2026; Korpi et al., 2019). AI-mediated dialogue may function as an extension of these approaches when carefully scaffolded.

Second, the modeling results suggest that technological usability alone does not guarantee educational value. In settings where AI is introduced primarily as a productivity tool, gains may remain superficial. In contrast, when integrated into structured clinical scenarios that demand analytic engagement, AI can support higher-order thinking processes. This aligns with emerging work on adaptive tutoring systems, where educational impact depends on the interplay between feedback mechanisms and curricular design (Kestin et al., 2025).

Third, the findings underscore the importance of AI literacy within health profession curricula. As students increasingly encounter AI systems in clinical practice, their capacity to critically evaluate outputs becomes central to patient safety and professional responsibility. Although literacy measures were not directly assessed in this study, the strong ratings for ethical integration suggest that explicit framing may enhance learners' reflective engagement with AI tools. Future research should incorporate validated measures of AI self-efficacy and literacy to examine how these constructs interact with clinical reasoning development (Y. Y. Wang & Chuang, 2023; El-Sayed et al., 2025).

At a systemic level, the results support the view that AI integration in health education should be conceptualized as curricular transformation rather than technological supplementation. Educational design, ethical governance, and cognitive scaffolding must evolve concurrently. When these elements are aligned, AI-mediated learning may enhance performance outcomes and epistemic maturity.

3.5. Limitations and Future Directions

Several limitations warrant consideration. The pilot design and modest sample size restrict statistical power and preclude causal inference. Although effect sizes were substantial, confidence intervals were wide, and multivariable estimates should therefore be interpreted cautiously. The absence of demographic variables also limits examination of potential moderating factors such as prior digital experience or familiarity with AI tools.

In addition, the multivariable regression model should be interpreted with caution, given the limited sample size relative to the number of predictors included. Post-hoc power analysis indicated low statistical power, and the observations-to-predictor ratio was below commonly recommended thresholds for stable estimation. Therefore, regression findings are best understood as exploratory and hypothesis-generating, aimed at identifying potential patterns of association rather than establishing independent effects.

The exploratory factor analysis was conducted within a small cohort and should therefore be regarded as preliminary. Confirmatory modeling in larger, multi-institutional samples will be necessary to validate the structural pathways proposed here. Randomized or quasi-experimental designs would also allow more definitive conclusions regarding mediated effects.

An additional limitation concerns the reliance on perceived learning indicators. Although perceived learning impact showed a strong association with student satisfaction, subjective perceptions do not necessarily correspond to objective knowledge acquisition or improvements in reasoning performance. Future research should therefore incorporate objective measures of clinical reasoning, such as blinded evaluation of student case analyses, standardized reasoning assessments, or comparative study designs including control groups. Such approaches would help determine whether AI-mediated instructional frameworks translate into measurable improvements in higher-order cognitive outcomes.

Despite these limitations, the present findings establish a foundation for prospective research. By articulating a coherent pedagogical framework and modeling its internal relationships, this study provides a structural blueprint for future confirmatory investigations.

4. Conclusions

This pilot explanatory study developed and empirically evaluated a theory-driven AI-mediated clinical reasoning framework within undergraduate physiotherapy education and examined the structural relationships among pedagogical coherence, usability, perceived learning impact, and student satisfaction.

The findings indicate that perceived learning impact emerged as the central explanatory mechanism linking instructional design to learner endorsement. While usability and pedagogical coherence were positively associated with satisfaction at the bivariate level, their independent contributions attenuated under simultaneous modeling, reflecting structural interdependence within an integrated educational system. These results suggest that the educational value of artificial intelligence in health professions training is contingent upon its alignment with cognitively demanding instructional goals rather than on technological functionality alone.

Although limited by its pilot design and sample size, this study provides preliminary structural evidence supporting AI-mediated clinical reasoning as a coherent pedagogical framework. Future multicenter and confirmatory investigations are warranted to test mediated pathways prospectively and examine whether similar structural patterns generalize across disciplines within health sciences education.

Author Contributions: Conceptualization, A.M.O.; Methodology, A.M.O., E.C.-R. and R.Q.-A.; Software, A.M.O. and R.Q.-A.; Validation, A.M.O., E.C.-R. and R.Q.-A.; Formal analysis, A.M.O.; Investigation, A.M.O., E.C.-R. and R.Q.-A.; Data curation, A.M.O. and R.Q.-A.; Writing—original draft, A.M.O., E.C.-R. and R.Q.-A.; Writing—review & editing, A.M.O., E.C.-R. and R.Q.-A.; Visualization, A.M.O. and R.Q.-A.; Supervision, A.M.O.; Project administration, A.M.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study. The study described in the manuscript forms part of a teaching innovation project conducted within a regular university course at the Universidad Pontificia de Salamanca. According to the institutional guidelines governing research and innovation activities involving students, projects of educational innovation aimed at improving teaching methodologies within a course do not require prior review by the institutional Ethics Committee, provided that specific safeguards are met. In particular, the institutional guidelines state that in educational innovation or teaching-improvement projects involving one's own students, the principal investigator must submit a responsible declaration committing to ethical conduct, voluntary participation, and data confidentiality, and must obtain authorization from the Faculty Dean before the project begins. These requirements were fully complied with in this case. Specifically: "A Responsible Declaration for an Educational Innovation Project was submitted by the principal investigator, committing to ensure voluntary participation, confidentiality, and compliance with applicable regulations." The project received formal authorization from the Faculty Dean, confirming that participation procedures respected participants' rights and did not interfere with academic activities. Both documents were submitted together with the project proposal prior to implementation. The study therefore followed the ethical and legal procedures established by the university for educational innovation projects involving students, ensuring voluntary participation, protection of participants' rights, and compliance with applicable data protection regulations. Additionally, the study adhered to the ethical principles outlined in the Declaration of Helsinki (1975, revised 2013).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The dataset generated and analyzed during the current study is publicly available in the Zenodo repository at https://doi.org/10.5281/zenodo.19097499.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
CI	Confidence interval
FA	Physiotherapy of the Locomotor System (course name)
IQR	Interquartile range
LLM	Large language model
PBL	Problem-based learning
QR	Quick response code
SD	Standard deviation
SPSS	Statistical Package for the Social Sciences

References

Abugri, B. A., Assibi, M. A., Amalba, A., Gaa, P. K., Kpebu, S. E. A., Karikari, P. A. A., & Mogre, V. (2026). The effects of problem-based learning (PBL) on undergraduate medical students' critical thinking and communication skills development: A scoping review across resource-rich and resource-limited settings (2015–2024). *PLoS ONE*, *21*(2), e0342599. [CrossRef]

Albarrán Lozano, I., & Molina, J. M. (2021). Perception of artificial intelligence in Spain. *Telematics and Informatics*, *63*, 101672. [CrossRef]

Al Kuwaiti, A., Nazer, K., Al-Reedy, A., Al-Shehri, S., Al-Muhanna, A., & Subbarayalu, A. V. (2023). A review of the role of artificial intelligence in healthcare. *Journal of Personalized Medicine*, *13*(6), 951. [CrossRef] [PubMed]

Almansour, M., & Alfhaid, F. M. (2024). Generative artificial intelligence and the personalization of health professional education: A narrative review. *Medicine*, *103*(31), e38955. [CrossRef] [PubMed]

Ametepey, S. O., Aigbavboa, C., Thwala, W. D., & Addy, H. (2024). The impact of AI in sustainable development goal implementation: A Delphi study. *Sustainability*, *16*(9), 3858. [CrossRef]

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Choi, T. R., & Drumwright, M. E. (2021). "OK, Google, why do I use you?" Motivations, post-consumption evaluations, and perceptions of voice AI assistants. *Telematics and Informatics*, *62*, 101628. [CrossRef]

Choung, H., David, P., & Ross, A. (2023). Trust and ethics in AI. *AI & Society*, *38*(2), 733–745. [CrossRef]

Çayır, A. (2023). A literature review on the effect of artificial intelligence on education. *İnsan ve Sosyal Bilimler Dergisi*, *6*(2), 276–288. [CrossRef]

Das, K., Pattanaik, M., Basantia, S., Mishra, R., Das, D., Sahoo, K., & Paital, B. (2023). Informatics on a social view and need of ethical interventions for wellbeing via interference of artificial intelligence. *Telematics and Informatics Reports*, *11*, 100065. [CrossRef]

D'Sa, J. L. (2015). Effect of problem based learning on motivation of nursing students. *International Journal of Current Research and Review*, *7*, 34–38.

El-Sayed, B. K. M., El-Sayed, A. A. I., Alsenany, S. A., & Asal, M. G. R. (2025). The role of artificial intelligence literacy and innovation mindset in shaping nursing students' career and talent self-efficacy. *Nurse Education in Practice*, *82*, 104208. [CrossRef]

Ergezen Sahin, G., Aras Bayram, G., Sanchez Sierra, A., Akdemir, S., Kurc, D., Tarakci, D., & Tunali, A. N. (2025). Effects of artificial intelligence–based physiotherapy educational approach in developing clinical reasoning skills: A randomized controlled trial. *BMC Medical Education*, *25*, 1378. [CrossRef]

Ferrer-Peña, R., Di-Bonaventura, S., Pérez-González, A., & Lerín-Calvo, A. (2025). Feasibility of a randomized controlled trial of large AI-based linguistic models for clinical reasoning training of physical therapy students: Pilot randomized parallel-group study. *JMIR From Res*, *11*, e66126. [CrossRef]

Gesk, T. S., & Leyer, M. (2022). Artificial intelligence in public services: When and why citizens accept its usage. *Government Information Quarterly*, *39*(3), 101704. [CrossRef]

Goralski, M. A., & Tan, T. K. (2020). Artificial intelligence and sustainable development. *The International Journal of Management Education*, *18*(1), 100330. [CrossRef]

Gunn, H., Hunter, H., & Haas, B. (2012). Problem-based learning in physiotherapy education: A practice perspective. *Physiotherapy*, *98*(4), 330–335. [CrossRef] [PubMed]

Hao, J., Yao, Z., Tang, Y., Remis, A., Wu, K., & Yu, X. (2025). Artificial intelligence in physical therapy: Evaluating ChatGPT's role in clinical decision support for musculoskeletal care. *Annals of Biomedical Engineering*, *53*(1), 9–13. [CrossRef]

Hui, Z., Zewu, Z., Jiao, H., & Yu, C. (2025). Application of ChatGPT-assisted problem-based learning teaching method in clinical medical education. *BMC Medical Education*, *25*(1), 50. [CrossRef]

Jackson, P., Ponath Sukumaran, G., Babu, C., Tony, M. C., Jack, D. S., Reshma, V. R., Davis, D., Kurian, N., & John, A. (2024). Artificial intelligence in medical education: Perception among medical students. *BMC Medical Education*, *24*, 804. [CrossRef]

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, *15*(1), 17458. [CrossRef]

Korpi, H., Peltokallio, L., & Piirainen, A. (2019). Problem-based learning in professional studies from the physiotherapy students' perspective. *Interdisciplinary Journal of Problem-Based Learning*, *13*(1), 4. [CrossRef]

Lindbäck, Y., Schröder, K., Engström, T., Valeskog, K., & Sonesson, S. (2025). Generative artificial intelligence in physiotherapy education: Great potential amidst challenges—A qualitative interview study. *BMC Medical Education*, *24*, 603. [CrossRef] [PubMed]

Möller, M., Nirmal, G., Fabietti, D., Stierstorfer, Q., Zakhvatkin, M., Sommerfeld, H., & Schütt, S. (2024). Revolutionising distance learning: A comparative study of learning progress with AI-driven tutoring. *arXiv*. [CrossRef]

Naqvi, W. M., Shaikh, S. Z., & Mishra, G. V. (2024). The role of large language models in physiotherapy education and practice: Opportunities and challenges. *Frontiers in Public Health*, *12*, 1364660. [CrossRef]

Ngai, E. W. T. (2012). The impact of information technology on cognitive processes and learning behaviors: Implications for technology-mediated education. *Computers & Education*, *58*(4), 1135–1145. [CrossRef]

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. [CrossRef]

Palliyil, V., Cai, M., Karam, H., Phatthanachaisuksiri, L., Suhre, N., & Kaßens-Noor, E. (2025). Artificial intelligence use cases adopted by people and their impact on achieving sustainable development goals: A systematic review. *Open Research Europe*, *5*, 117. [CrossRef]

Roche, M., DeWitt, D. E., & Marks, S. (2016). PBL trigger design by medical students: An effective active learning strategy outside the classroom. *Journal of Clinical and Diagnostic Research*, *10*(12), JC06–JC08. [CrossRef]

Safranek, C. W., Sidamonidze, N., Zou, J., DiPietro, M., & Chiel, H. J. (2023). The role of large language models in medical education: Applications and implications. *JMIR Medical Education*, *9*, e50945. [CrossRef]

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, *11*(6), 887. [CrossRef]

Shorey, S., Mattar, C., Pereira, T. L. B., & Choolani, M. (2024). A scoping review of ChatGPT's role in healthcare education and research. *Nurse Education Today*, *135*, 106121. [CrossRef] [PubMed]

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications*, *12*, 621. [CrossRef]

Wang, Y. Y., & Chuang, Y. W. (2023). Artificial intelligence self-efficacy: Scale development and validation. *Education and Information Technologies*, *29*(4), 4785–4808. [CrossRef]