Research Square

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

Artificial Intelligence in Pediatric Education by SNAPPS and Clinical Reasoning Among Postgraduate Residents in a Tertiary Care Hospital in Central India

Rizwan Ahmed[1]

rizwanahmed1811@gmail.com

NKP Salve Institute of Medical Sciences and Lata Mangeshkar Hospital

Jayant Vagha[2]

Datta Meghe Institute of Higher Education and Research

Anjali Edbor

NKP Salve Institute of Medical Sciences and Lata Mangeshkar Hospital

Research Article

Keywords: Artificial intelligence, SNAPPS, clinical reasoning, Script Concordance Test, competency-based medical education, pediatrics, postgraduate training

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10783779/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Competency-based medical education (CBME) emphasizes clinical reasoning as a core postgraduate competency, yet structured teaching strategies such as SNAPPS remain under-optimized in pediatric postgraduate training. Artificial intelligence (AI) has been proposed as a supervised cognitive scaffold within SNAPPS, but empirical outcome-level evidence, particularly from pediatric postgraduate education in India, is limited.

Objective

To evaluate the effect of AI-assisted SNAPPS on clinical reasoning, structured case-presentation quality, and knowledge acquisition among pediatric postgraduate residents, and to assess learner and faculty perceptions of the intervention.

Methods

A prospective, single-arm, within-subject educational interventional study was conducted over six months among 30 pediatric postgraduate residents (10 each from JR1, JR2, and JR3) in the Department of Pediatrics, NKP Salve Institute of Medical Sciences & Research Centre and Lata Mangeshkar Hospital, Nagpur. Following sensitization, all participants underwent baseline assessment, a traditional-SNAPPS teaching phase, and an AI-assisted SNAPPS phase using a standardized SNAPPS-mapped AI prompt protocol with mandatory faculty validation at every step. Outcomes were measured using the Script Concordance Test (SCT), a validated 30-point SNAPPS analytical rubric, a multiple-choice question (MCQ) knowledge test, and five-point Likert perception surveys for residents and faculty. Paired comparisons were analyzed using appropriate parametric and non-parametric tests; $p < 0.05$ was considered significant.

Results

Mean SCT scores improved significantly with AI-assisted SNAPPS (12.0 ± 2.0 to 19.6 ± 2.8, $p < 0.001$), a gain that exceeded the improvement achieved with traditional SNAPPS alone (11.8 ± 2.1 to 15.2 ± 2.5, $p < 0.01$). The total SNAPPS rubric score rose from 19.1 ± 3.2 to 27.6 ± 2.9 ($p < 0.001$), with significant gains across all six domains. MCQ knowledge scores improved similarly (12.7 ± 2.5 to 18.8 ± 1.9, $p < 0.001$). SCT and SNAPPS rubric scores were strongly and positively correlated ($r = 0.72$, $p < 0.001$). Both residents and faculty reported significantly higher satisfaction with AI-assisted SNAPPS across every measured domain (all $p < 0.001$).

Conclusion

AI-assisted SNAPPS, implemented as a faculty-supervised cognitive scaffold, is feasible, well accepted, and educationally effective for strengthening clinical reasoning, structured presentation and knowledge retention among pediatric postgraduate residents. The model merits validation in larger, multicentric, competency-based curricula.

INTRODUCTION

Competency-based medical education (CBME) in India, introduced by the National Medical Commission, has reoriented postgraduate training away from rote factual recall toward demonstrable competencies in clinical reasoning, communication, professionalism and lifelong learning **[1]**. International competency frameworks similarly hold that a graduating physician must integrate scientific knowledge with sound clinical judgement under conditions of diagnostic uncertainty, and that curricula should be organized around observable milestones rather than time-based exposure alone **[2,3]**. Clinical reasoning — the cognitive process of collecting data, generating and weighing diagnostic hypotheses, and arriving at a safe management plan — is consistently identified as the competency most closely linked to patient safety, since deficiencies in reasoning underlie a disproportionate share of diagnostic error **[4,5]**.

Pediatric postgraduate training poses distinctive challenges to the development of this competency. Children present across a continuum of developmental stages, communicate symptoms unreliably or not at all, and frequently manifest disease non-specifically, so that trainees must reason under greater diagnostic uncertainty than their counterparts in adult medicine. Traditional bedside case presentations, although central to postgraduate teaching, often emphasize transfer of factual findings rather than exposition of the reasoning that connects findings to diagnosis and plan; faculty frequently supply answers rather than eliciting the trainee's own reasoning, which can blunt engagement, discourage acknowledgement of uncertainty, and limit opportunities for reflection.

The SNAPPS model — Summarize, Narrow, Analyze, Probe, Plan, Select — was developed by Wolpaw and colleagues as a learner-centred alternative to conventional case presentation **[6]**. By requiring residents to summarize the case succinctly, narrow and analyze a differential diagnosis, probe the preceptor about areas of uncertainty, commit to a management plan, and select a self-directed learning issue, SNAPPS operationalizes constructivist and metacognitive principles of adult learning **[7,8]**. Multiple studies have shown that SNAPPS increases learner autonomy, structures preceptor feedback, and enriches resident-faculty dialogue, although the magnitude of benefit is contingent on learner preparedness and faculty skill in facilitating each step **[8]**.

Over the past several years, generative artificial intelligence (AI) and large language models — including ChatGPT, Gemini and Claude — have entered mainstream discussion in health-professions education **[9,10]**. These systems can generate differential diagnoses, explain pathophysiological reasoning, draft evidence-based management options, and pose Socratic follow-up questions, positioning them as

potential cognitive scaffolds analogous to the temporary instructional support described in Vygotsky's sociocultural theory of learning, in which assistance is calibrated to the learner's zone of proximal development and gradually withdrawn as competence grows **[11]**. Early reports indicate that large language models can perform creditably on licensing-style examinations and can support self-directed learning when appropriately supervised **[12,13]**. At the same time, these systems remain prone to factual inaccuracy, outdated content, and confident but incorrect ("hallucinated") output; uncritical use risks eroding rather than strengthening independent reasoning, and international guidance therefore emphasizes human oversight, verification of AI-generated content, and protection of patient confidentiality as prerequisites for responsible educational use **[14,20]**.

Despite this rapidly expanding interest, direct empirical evidence for the role of AI in structured postgraduate clinical teaching remains sparse. Recent scoping reviews of AI in medical education consistently report that the available literature is dominated by narrative commentary and perception surveys, with comparatively few controlled or quasi-experimental studies that measure reasoning outcomes using validated instruments **[15,16,17]**. Reviews focused specifically on pediatric postgraduate education highlight the near-absence of studies that combine AI exposure with an established teaching framework such as SNAPPS; to our knowledge, none has evaluated this combination using the Script Concordance Test, the instrument regarded as most sensitive to reasoning under diagnostic uncertainty **[15]**. This represents an important and largely unaddressed research gap, particularly from low- and middle-income settings such as India, where adoption of digital-education tools is accelerating faster than the generation of outcome-level evidence to guide their use **[17,18,19]**.

A further consideration motivating this study is the growing recognition that AI literacy is itself becoming a professional competency. Just as earlier generations of postgraduate curricula incorporated evidence-based medicine and simulation-based learning as core skills, current and future clinicians will be expected to use AI tools critically, ethically, and in a manner that augments rather than supplants independent judgement **[9,13]**. Structured frameworks such as SNAPPS offer a natural scaffold onto which such AI literacy can be layered, because each step already demands an explicit, externally verifiable reasoning product — a differential list, a justification, a question, a plan — against which AI-generated suggestions can be compared, critiqued, and either adopted or discarded under faculty supervision. This positions AI-assisted SNAPPS not merely as a technological add-on but as a pedagogically coherent extension of an already-validated reasoning framework.

The present study was therefore designed to evaluate, within a functioning postgraduate pediatric training programme, whether embedding AI as a faculty-supervised cognitive scaffold within each step of the SNAPPS framework improves clinical reasoning (measured objectively using the Script Concordance Test), structured case-presentation quality (measured using a validated SNAPPS rubric), and knowledge acquisition (measured using an MCQ test), and how residents and faculty perceive this innovation. By using each resident as their own control across sequential traditional and AI-assisted phases, the study sought to generate methodologically rigorous, context-specific evidence to inform the responsible integration of AI into competency-based pediatric postgraduate curricula **[1,20]**.

METHODS

Study design

This was a prospective, single-arm, within-subject (repeated-measures) educational interventional study. Rather than comparing two independent groups, every participant sequentially completed a traditional-SNAPPS teaching phase followed by an AI-assisted SNAPPS phase, with structured assessment before and after each phase. This design allowed each resident to act as their own control, a strategy that minimizes inter-individual variability that would otherwise confound comparisons drawn from small postgraduate cohorts, and is consistent with recommended approaches for pilot educational interventions where randomization to withhold a promising teaching method is not considered justifiable.

Setting and duration

The study was conducted over six months in the Department of Pediatrics, NKP Salve Institute of Medical Sciences & Research Centre and Lata Mangeshkar Hospital, Nagpur, a tertiary-care teaching hospital in Central India, as part of the Advanced Course in Medical Education (ACME), with Jawaharlal Nehru Medical College, Wardha, as the nodal centre.

Participants

All thirty eligible postgraduate residents pursuing MD Pediatrics (ten each from Junior Resident Year 1, Year 2, and Year 3) were enrolled by complete enumeration of the departmental cohort, after written informed consent. Residents with prior formal SCT experience within the preceding three months and those unwilling to provide consent were excluded.

Ethical considerations

Institutional Ethics Committee approval was obtained from NKP Salve institute of Medical Sciences and Research Centre prior to commencement. All case material used for AI interaction was anonymized before entry into any AI platform, and no patient identifiers were disclosed at any stage, consistent with contemporary guidance on the ethical governance of AI in health professions education **[14]**.

Assessment instruments

Four complementary instruments were used. (i) The Script Concordance Test, comprising pediatric clinical scenarios developed and validated by a panel of five senior pediatric faculty and scored by the aggregate expert-scoring method, served as the primary outcome measure of reasoning under

uncertainty. (ii) A 30-point SNAPPS analytical rubric assessed the quality of case presentation across all six SNAPPS domains. (iii) A 20-item MCQ test assessed factual and applied knowledge. (iv) Five-point Likert-scale perception questionnaires captured resident confidence, satisfaction and perceived learning benefit, and a parallel faculty perception survey captured educator-rated presentation quality, depth of analysis, and satisfaction. A Faculty AI-Use Monitoring Checklist was additionally completed at every AI-assisted session to confirm that independent reasoning preceded AI consultation, that AI was used only as a learning aid, that no patient identifiers were entered, that AI output was critically appraised, and that the final reasoning was faculty-validated.

Procedure

Data were collected at three sequential time points. At baseline (T0), separate sensitization sessions oriented residents and faculty to the SNAPPS framework, principles of clinical reasoning, and ethical norms for AI use, after which the Script Concordance Test and baseline SNAPPS rubric scoring were administered during standard case presentations. In the traditional-SNAPPS phase (T1), residents presented pediatric cases using the unmodified SNAPPS framework under faculty supervision, after which the SCT, SNAPPS rubric, and MCQ were re-administered. In the AI-assisted SNAPPS phase (T2), residents used a standardized SNAPPS-mapped AI prompt protocol at each of the six steps — generating candidate differential diagnoses, contrasting supporting and opposing features, drafting probing questions, and reviewing the management plan against current evidence — with every AI-generated suggestion subjected to independent resident appraisal and mandatory faculty validation before use; the SCT, SNAPPS rubric, MCQ and perception surveys were then re-administered. Study data were recorded against unique, de-identified participant codes and entered into Microsoft Excel for analysis.

Standardized AI prompt protocol

To ensure uniformity of AI use across residents and reduce inter-rater variability in how the technology was applied, a fixed set of SNAPPS-mapped prompts was used at every AI-assisted session. For the Summarize step, residents were directed to have the AI review an anonymized case summary and flag missing clinically relevant information without offering a diagnosis. For Narrow, the AI was prompted to suggest differentials categorized as common, serious ("not-to-miss"), and uncommon, each with brief justification. For Analyze, the AI was asked to contrast supporting and opposing features of the leading differentials and identify additional information needed. For Probe, the AI generated two to three higher-order discussion questions for the faculty preceptor. For Plan, the AI reviewed the proposed management for gaps, monitoring needs, and evidence-based considerations without substituting for faculty decision-making. For Select, the AI proposed candidate self-directed learning topics linked to the case. At every step, residents were required to complete their own independent reasoning before

consulting the AI, and faculty reviewed and validated the final reasoning before it was accepted for educational credit, consistent with the Faculty AI-Use Monitoring Checklist described above.

Statistical analysis

Continuous variables were expressed as mean ± standard deviation and compared using the paired t-test for normally distributed data; ordinal Likert-scale data were summarized as mean ± SD and compared using the Wilcoxon signed-rank test. Categorical variables were expressed as frequencies and percentages. The relationship between SCT and SNAPPS rubric scores was assessed using Pearson's correlation coefficient. A two-tailed p-value <0.05 was considered statistically significant. Analyses were performed using standard statistical software.

RESULTS AND OBSERVATION

Thirty pediatric postgraduate residents completed the study, comprising ten residents each from JR1, JR2 and JR3, with no attrition. The mean age was 27.1 ± 1.2 years, and female residents (n=16, 53.3%) slightly outnumbered male residents (n=14, 46.7%) (Table 1). The balanced distribution of training year and comparable demographic profile minimized selection bias, supporting attribution of the observed changes to the educational intervention rather than to baseline cohort differences, and the demographic distribution is illustrated in Figure 1.

Table 1. Demographic Profile of Participants

Variable	Value
Mean age (years)	27.1 ± 1.2
JR1	10
JR2	10
JR3	10
Female	16 (53.3%)
Male	14 (46.7%)

Clinical reasoning, assessed by the Script Concordance Test, improved significantly across both teaching phases, with a substantially larger gain following AI-assisted SNAPPS. Mean SCT scores rose from 11.8 ± 2.1 to 15.2 ± 2.5 ($p<0.01$) after the traditional-SNAPPS phase, and from 12.0 ± 2.0 to 19.6 ± 2.8 ($p<0.001$) after the AI-assisted phase (Table 2, Figure 2). Post-intervention scores were significantly higher after the AI-assisted phase than after the traditional phase (19.6 ± 2.8 vs 15.2 ± 2.5, $p<0.001$; Table 2, Figure 3), indicating that AI-assisted scaffolding produced a measurably greater improvement in reasoning under diagnostic uncertainty than structured teaching alone.

Table 2a. Comparison of Script Concordance Test (SCT) Scores

Phase	Pre-test Mean ± SD	Post-test Mean ± SD	p-value
Traditional SNAPPS	11.8 ± 2.1	15.2 ± 2.5	<0.01
AI-Assisted SNAPPS	12.0 ± 2.0	19.6 ± 2.8	<0.001

Table 2b. Post-Intervention SCT Score: Traditional vs AI-Assisted Phase

Variable	Traditional	AI-Assisted	p-value
Post-test SCT score	15.2 ± 2.5	19.6 ± 2.8	<0.001

Structured case-presentation quality, assessed using the six-domain SNAPPS rubric, showed parallel improvement. The total rubric score increased from 19.1 ± 3.2 to 27.6 ± 2.9 ($p<0.001$), with statistically significant gains in every domain (Table 3, Figures 4–5). The largest relative gains were observed in probing-question quality and self-directed learning, the two domains most dependent on active generation of higher-order questions.

Table 3. SNAPPS Rubric Domain Scores

Domain	Traditional (mean)	AI-Assisted (mean)	p-value
Case summary	3.4 ± 0.7	4.5 ± 0.5	<0.001
Differential diagnosis	3.2 ± 0.6	4.6 ± 0.4	<0.001
Analytical reasoning	3.1 ± 0.8	4.7 ± 0.5	<0.001
Probing questions	2.9 ± 0.7	4.4 ± 0.6	<0.001
Management planning	3.5 ± 0.6	4.6 ± 0.5	<0.001
Self-directed learning	3.0 ± 0.8	4.8 ± 0.4	<0.001
Total score (/30)	19.1 ± 3.2	27.6 ± 2.9	<0.001

Knowledge acquisition, assessed by the MCQ test, improved in both phases but to a greater degree with AI assistance (12.7 ± 2.5 to 18.8 ± 1.9, $p<0.001$) than with traditional teaching alone (12.4 ± 2.3 to 15.1 ± 2.1, $p=0.01$) (Table 4, Figure 6), suggesting that AI-supported discussion strengthened factual retention in addition to reasoning skill.

Table 4. MCQ Knowledge Test Scores

Phase	Pre-test	Post-test	p-value
Traditional	12.4 ± 2.3	15.1 ± 2.1	0.01
AI-Assisted	12.7 ± 2.5	18.8 ± 1.9	<0.001

Resident perception of the AI-assisted intervention was highly favourable across all measured domains (Table 5), and faculty perception mirrored this pattern (Table 6, Figure 7).

Table 5. Postgraduate Satisfaction Scores (5-point Likert scale)

Parameter	Traditional	AI-Assisted	p-value
Improved confidence	3.2 ± 0.8	4.6 ± 0.5	<0.001
Better reasoning	3.1 ± 0.7	4.7 ± 0.4	<0.001
Structured learning	3.4 ± 0.6	4.8 ± 0.3	<0.001
Self-directed learning	3.0 ± 0.8	4.7 ± 0.5	<0.001
Overall satisfaction	3.3 ± 0.7	4.8 ± 0.4	<0.001

Table 6. Faculty Feedback Scores

Parameter	Traditional	AI-Assisted	p-value
Presentation quality	3.5 ± 0.5	4.7 ± 0.4	<0.001
Depth of analysis	3.1 ± 0.6	4.6 ± 0.5	<0.001
Clinical reasoning	3.2 ± 0.6	4.8 ± 0.3	<0.001
Faculty satisfaction	3.4 ± 0.5	4.7 ± 0.4	<0.001

A strong, statistically significant positive correlation was observed between SCT scores and total SNAPPS rubric scores ($r=0.72$, $p<0.001$; Table 7, Figure 8), indicating that residents who achieved higher structured-presentation scores also demonstrated superior clinical reasoning, and supporting the construct validity of the SNAPPS framework as used in this intervention.

Table 7. Correlation Between SCT and SNAPPS Rubric Scores

Variable	Correlation coefficient (r)	p-value
SCT vs SNAPPS total score	0.72	<0.001

DISCUSSION

This prospective, within-subject educational interventional study demonstrates that embedding AI as a faculty-supervised cognitive scaffold within the SNAPPS framework produces significantly greater improvement in clinical reasoning, structured case presentation, and knowledge acquisition than traditional SNAPPS teaching alone, and is met with high satisfaction among both postgraduate residents and faculty. These findings extend a small but growing body of evidence on responsible AI integration in postgraduate medical education and address a specific gap identified in recent scoping reviews — the near-absence of outcome-level evidence combining AI exposure with an established clinical-reasoning framework in pediatric postgraduate training **[15,17]**.

Clinical reasoning is widely regarded as the cornerstone of safe medical practice. Charlin and colleagues conceptualize reasoning as the dynamic organization and application of illness scripts under conditions of uncertainty, and proposed the Script Concordance Test specifically to capture this process rather than simple factual recall **[21,22]**. The significant improvement in SCT scores observed after AI-assisted SNAPPS in the present study suggests that AI-generated prompts encouraged deeper analytical engagement — requiring residents to justify differential diagnoses, weigh contradictory evidence, and reconsider management options — rather than passive acceptance of AI output. This interpretation is consistent with reports that structured, question-driven interaction with AI tools promotes higher-order reasoning more effectively than simple information retrieval **[9,13]**. It is also consistent with the well-established observation that clinical reasoning is highly context- and case-specific rather than a single generic skill transferable in the abstract, so that repeated, case-anchored practice — of the kind provided across sequential SNAPPS cases in this study — is more likely to yield durable gains than reasoning taught in isolation from real clinical content **[28,29]**.

The improvement in SNAPPS rubric scores reinforces the educational value of learner-centred structured teaching. Wolpaw and colleagues, who originally described SNAPPS, reported that the framework enhances learner autonomy and stimulates active discussion between residents and preceptors **[6]**. The present findings extend this observation by showing that AI can enrich each SNAPPS step through structured cognitive prompting, yielding better case summaries, broader differential diagnoses, stronger analytical justification, more incisive probing questions, and more deliberate self-directed learning goals. The largest gains were observed in the probing-questions and self-directed-learning domains, both of which depend on the learner's ability to generate novel higher-order questions rather than reproduce previously acquired information — precisely the domains in which a well-supervised AI scaffold, by modelling alternative questions and unexplored angles, might be expected to exert its greatest educational leverage.

A particularly notable finding was the strong positive correlation between SCT and SNAPPS rubric scores ($r = 0.72$, $p < 0.001$), indicating that residents who presented cases in a more structured manner also demonstrated superior clinical reasoning.

This association supports the construct validity of SNAPPS as a proxy for underlying reasoning ability and is consistent with prior work linking structured clinical discussion to reasoning competence **[22,23]**.

It also implies that interventions which strengthen structured presentation — as AI-assisted SNAPPS did in this study — are likely to yield genuine gains in reasoning rather than superficial improvements in communication style alone.

Knowledge acquisition, measured by the MCQ test, improved significantly more in the AI-assisted phase than in the traditional phase. AI is sometimes criticized for potentially reducing independent thinking by supplying ready answers; the present findings, obtained under a protocol of mandatory independent reasoning before AI consultation and mandatory faculty validation afterward, suggest the opposite outcome when such safeguards are enforced. AI appeared to function primarily as a facilitator that organized concepts, clarified pathophysiological mechanisms, and prompted evidence-based review, rather than as a substitute for learner effort. This pattern aligns with Vygotsky's concept of scaffolded learning, in which guided support enables performance beyond a learner's unaided capability while progressively building toward independent mastery **[11]**, and with cognitive load theory, which holds that well-structured external support reduces extraneous cognitive load and frees working memory for schema construction and illness-script formation **[24]**.

Both resident and faculty perception data were strongly favourable, corroborating findings from recent reviews that AI-based educational tools enhance learner engagement, motivation, and personalized learning experience **[16,17]**. Faculty specifically noted more organized reasoning, clearer justification of differential diagnoses, and greater participation during discussions — observations that are consistent with the broader literature describing AI's potential to augment, rather than replace, experiential and mentorship-based teaching **[23]**. Importantly, faculty in this study uniformly emphasized that AI should complement, not substitute, traditional teaching; the mandatory faculty-validation step built into the protocol appears to have been central to translating AI assistance into genuine educational benefit while minimizing the risk of uncritical acceptance of hallucinated or inaccurate content **[14]**.

The present findings are also directly relevant to the aims of competency-based medical education, which prioritizes milestones in reasoning, communication, and lifelong learning over factual knowledge in isolation **[1,2]**. By promoting metacognitive reflection, hypothesis generation, evidence appraisal, and self-directed learning, AI-assisted SNAPPS aligns closely with CBME principles and offers a practical, scalable strategy for embedding responsible digital innovation within existing postgraduate curricula, without requiring additional faculty time beyond validation of AI-generated content **[3]**. Such scaffolded, deliberate-practice-oriented exposure is consistent with broader educational theory linking repeated, feedback-rich practice to the development of expertise, and echoes the wider simulation- and technology-enhanced learning literature, in which structured, low-stakes rehearsal with timely feedback — whether through simulation, virtual patients, or, as here, an AI-supported scaffold — has consistently been shown to accelerate skill acquisition **[25,26,27,30]**.

International literature increasingly supports this model of supervised AI integration. Recent scoping reviews of AI in pediatric and general postgraduate medical education emphasize AI's potential to enhance clinical reasoning while underscoring the continued necessity of faculty oversight, verification

of content accuracy, and adherence to ethical safeguards around patient data **[15,17]**. The World Health Organization has similarly called for governance frameworks that ensure AI in health and health education remains transparent, accountable, and subordinate to human judgement **[14]**. The present study operationalized these principles directly — through anonymization of case material, mandatory independent reasoning prior to AI consultation, and compulsory faculty validation — and its favourable outcomes provide empirical support for this governance-conscious approach to AI adoption in postgraduate pediatric education.

Limitations

Several limitations must be acknowledged, and addressing them helps to delineate the boundaries within which these findings should be interpreted. First, the study was conducted at a single institution with a modest sample of thirty residents, which limits statistical power for subgroup analyses (for example, by postgraduate year) and constrains generalizability to other training environments; multicentric replication is needed. Second, the within-subject, sequential design — while efficient and ethically appropriate for a promising teaching innovation — cannot fully exclude an order or practice effect, whereby residents' performance in the AI-assisted phase may have been partly attributable to the immediately preceding traditional-SNAPPS phase rather than to AI exposure alone; future studies should consider counterbalanced or randomized sequencing across cohorts to isolate this effect. Third, the intervention period was relatively short, and long-term retention of clinical-reasoning gains beyond the study period was not assessed; longitudinal follow-up incorporating delayed SCT administration would clarify durability. Fourth, large language models evolve rapidly, and the specific prompt protocol and AI platform used here may not generalize to future model versions without re-validation. Finally, faculty enthusiasm for a novel, well-supported intervention may have introduced a degree of positive-outcome bias in perception measures, which future studies could mitigate through blinded external assessment of case-presentation recordings.

Notwithstanding these limitations, the study has notable strengths. The use of a prospective, within-subject design allowed each resident to serve as their own control, reducing inter-individual variability that would otherwise complicate small-cohort comparisons. Multiple validated instruments — the Script Concordance Test, SNAPPS rubric, MCQ assessment, and structured perception surveys — provided convergent evidence across cognitive, behavioural and affective domains, and the significant positive correlation between SCT and SNAPPS scores strengthens confidence that the observed gains reflect genuine improvement in reasoning rather than assessment artefact. Mandatory faculty validation and a documented AI-use monitoring checklist embedded ethical safeguards directly into the educational protocol rather than treating them as an afterthought, offering a template that other departments could adopt with minimal modification **[14]**.

Taken together, these findings suggest several practical implications for departments considering similar innovations. First, the incremental faculty burden was limited chiefly to validation rather than content generation, suggesting that AI-assisted SNAPPS could be implemented without materially

increasing teaching workload once a standardized prompt protocol and monitoring checklist are in place. Second, the magnitude of improvement in the probing-questions and self-directed-learning domains suggests that AI scaffolds may be particularly useful for developing precisely those higher-order skills that postgraduate curricula struggle most to teach explicitly. Third, because the intervention relies on freely or widely available conversational AI platforms rather than bespoke educational software, the model is potentially scalable to resource-limited postgraduate programmes without significant additional infrastructure investment, provided that institutions establish clear data-protection and validation safeguards beforehand. Future multicentric studies with randomized or counterbalanced sequencing, larger sample sizes, blinded outcome assessment, and longer follow-up — including delayed post-testing to evaluate durability of reasoning gains and, ultimately, correlation with clinical performance and patient outcomes — are needed to consolidate this preliminary evidence into a generalizable recommendation for postgraduate pediatric curricula.

CONCLUSION

This prospective educational interventional study demonstrates that AI, when deployed as a faculty-supervised cognitive scaffold within the structured SNAPPS framework, significantly improves clinical reasoning, structured case-presentation quality, and knowledge acquisition among pediatric postgraduate residents, and is well received by both learners and faculty. These findings support the responsible incorporation of AI into competency-based postgraduate pediatric education while reaffirming that faculty mentorship, independent reasoning, and validation of AI-generated content remain indispensable safeguards. AI-assisted SNAPPS represents a feasible, scalable, and reproducible teaching innovation that warrants confirmation in larger, multicentric, and longer-duration studies, and may serve as a template for the broader integration of artificial intelligence into competency-based medical curricula.

Declarations

Ethical approval: Approval was obtained from the Institutional Ethics Committee from NKP Salve institute of Medical Sciences and Research Centre prior to commencement of the study. Written informed consent was obtained from all participants.

Conflict of interest: The authors declare no conflict of interest.

Consent for publication is not applicable as this is an orginal article and its ownership solely lies with the corresponding author.

Funding: This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability: De-identified study data are available from the corresponding author on reasonable request in excel format.

Author contributions: RA — conceptualization, data collection, analysis, manuscript drafting. JV — supervision, mentorship, manuscript review and editing.

ACKNOWLEDGEMENT

The authors sincerely thank Dr. Jayant Vagha for constant guidance, valuable suggestions, and expert mentorship throughout this study. The authors are grateful to the faculty and administration of Jawaharlal Nehru Medical College, Wardha, and the ACME team for providing an excellent learning environment during the Advanced Course in Medical Education (ACME). The authors acknowledge Dr. Abhishek Ingole, Co-convener of ACME 18A Batch, for his motivation and support, and thank the ACME faculty for their dedicated teaching and insightful discussions. The authors also thank their colleagues at NKP Salve Institute of Medical Sciences & Research Centre and Lata Mangeshkar Hospital, particularly Dr. Anjali Edbor and Dr. Girish Nanoti, for their valuable suggestions and support, and gratefully acknowledge all pediatric postgraduate residents who participated in this educational intervention.

References

1. National Medical Commission. Competency-Based Postgraduate Curriculum for MD Pediatrics. New Delhi: National Medical Commission; 2023.
2. Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, et al. Competency-based medical education: Theory to practice. Med Teach. 2010;32(8):638-645.
3. Harden RM. Outcome-based education: The future is today. Med Teach. 2007;29(7):625-629.
4. Gruppen LD. Clinical reasoning: Defining it, teaching it, assessing it, studying it. West J Emerg Med. 2017;18(1):4-7.
5. Croskerry P. Clinical cognition and diagnostic error: Applications of a dual process model of reasoning. Adv Health Sci Educ Theory Pract. 2009;14(Suppl 1):27-35.
6. Wolpaw TM, Wolpaw DR, Papp KK. SNAPPS: A learner-centered model for outpatient education. Acad Med. 2003;78(9):893-898.
7. Eva KW. What every teacher needs to know about clinical reasoning. Med Educ. 2005;39(1):98-106.
8. ten Cate O, Durning SJ. Dimensions and psychology of peer teaching in medical education. Med Teach. 2007;29(6):546-552.
9. Masters K. Artificial intelligence in medical education. Med Teach. 2019;41(9):976-980.
10. Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Differ. 2023;103:102274.
11. Vygotsky LS. Mind in Society: The Development of Higher Psychological Processes. Cambridge: Harvard University Press; 1978.

12. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on the USMLE: Potential for AI-assisted medical education. PLoS Digit Health. 2023;2(2):e0000198.
13. Sallam M. ChatGPT utility in healthcare education, research, and practice: A comprehensive review. Healthcare (Basel). 2023;11(6):887.
14. World Health Organization. Ethics and Governance of Artificial Intelligence for Health. Geneva: World Health Organization; 2021.
15. Nadarajah A, Malekzadeh G, Ratnapalan S, et al. The use of artificial intelligence in paediatric postgraduate medical education: A scoping review. Paediatr Child Health. 2025;30(8):694-704.
16. Hernández Rincón EH, Jimenez D, Chavarro LA, et al. Mapping the use of artificial intelligence in medical education: A scoping review. BMC Med Educ. 2025;25:526.
17. Li Q, Qin Y. AI in medical education: Medical student perception, curriculum recommendations and design suggestions. BMC Med Educ. 2023;23:852.
18. Tlili A, Shehata B, Adarkwah MA, Bozkurt A, Hickey DT, Huang R, et al. What if the devil is my guardian angel: ChatGPT as a case study of AI in higher education. Smart Learn Environ. 2023;10:15.
19. Luckin R, Holmes W, Griffiths M, Forcier LB. Intelligence Unleashed: An Argument for AI in Education. London: Pearson Education; 2016.
20. Sandars J, Patel R, Steele H, McAreavey M. Developmental student support in the era of artificial intelligence. Med Teach. 2023;45(2):119-126.
21. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten CPM. The Script Concordance Test: A tool to assess clinical reasoning. Med Educ. 2000;34:707-714.
22. Charlin B, Boshuizen HPA, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. Acad Med. 2007;82(Suppl 10):S4-S10.
23. Topol E. High-performance medicine: The convergence of human and artificial intelligence. Nat Med. 2019;25:44-56.
24. van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: Design principles and strategies. Med Educ. 2010;44(1):85-93.
25. Cook DA, Triola MM. Virtual patients: A critical literature review and proposed next steps. Med Educ. 2009;43(4):303-311.
26. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. Med Teach. 2005;27(1):10-28.
27. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: Theory and implication. Acad Med. 1990;65(10):611-621.
28. Norman G. Research in clinical reasoning: Past history and current trends. Med Educ. 2005;39(4):418-427.

29. Durning SJ, Artino AR Jr, Pangaro LN, van der Vleuten C, Schuwirth L. Context and clinical reasoning: Understanding diagnostic decision making. Med Educ. 2011;45(9):927-938.
30. Ericsson KA. Deliberate practice and acquisition of expert performance in medicine and related domains. Acad Emerg Med. 2008;15(11):988-994.

Figures

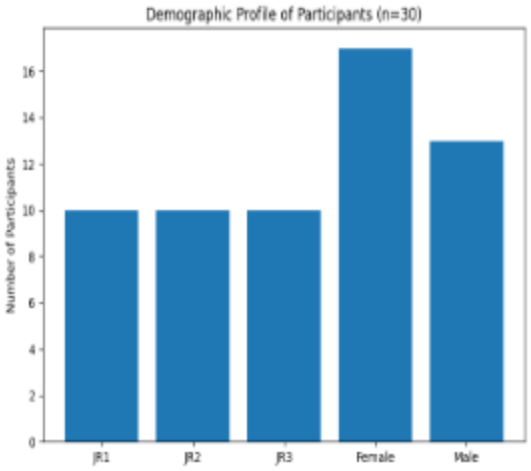


Figure 1

Demographic profile of the study cohort.

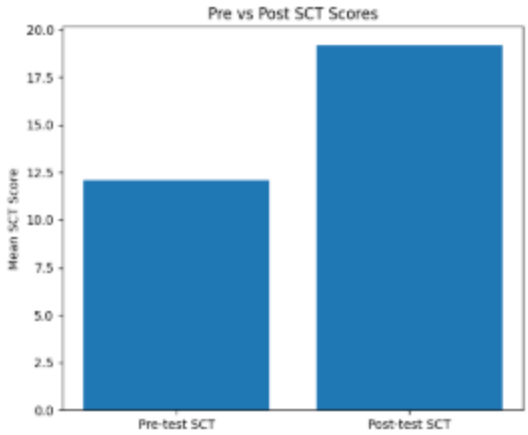


Figure 2

Pre- and post-intervention SCT scores by phase.

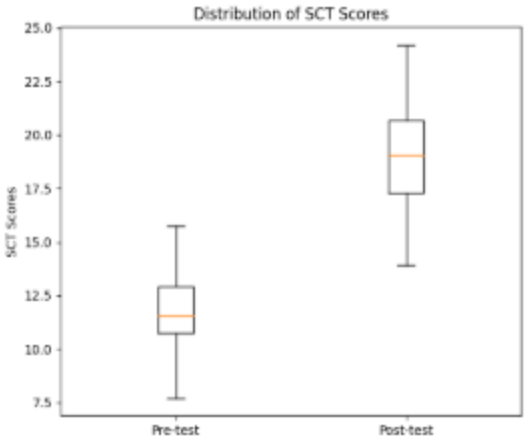


Figure 3

Box-plot comparison of post-intervention SCT scores between phases.

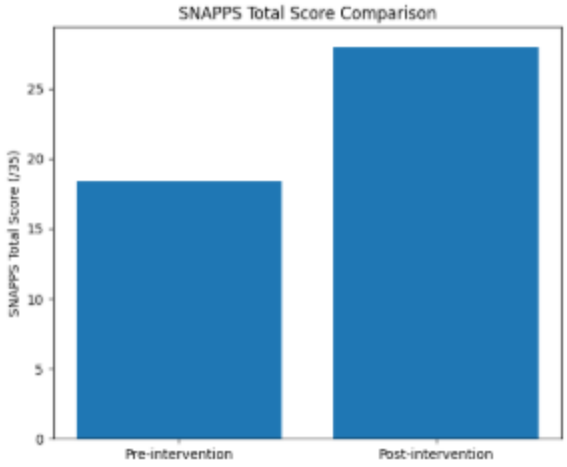


Figure 4

Total SNAPPS rubric score comparison.

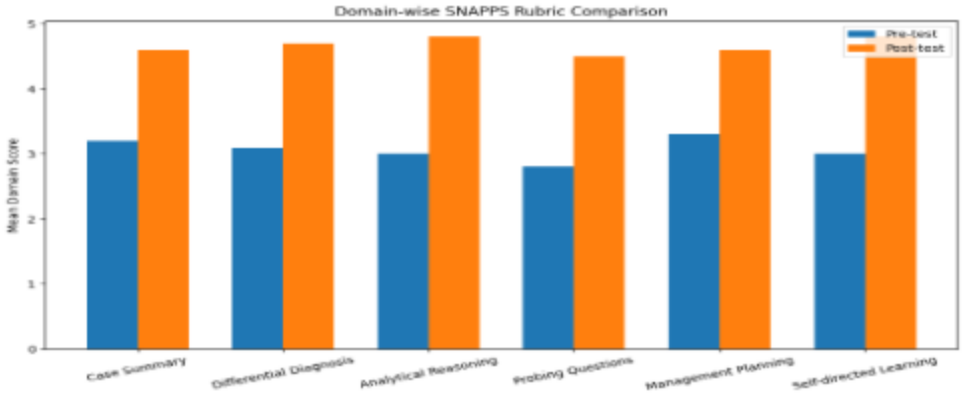


Figure 5

Domain-wise SNAPPS rubric comparison.

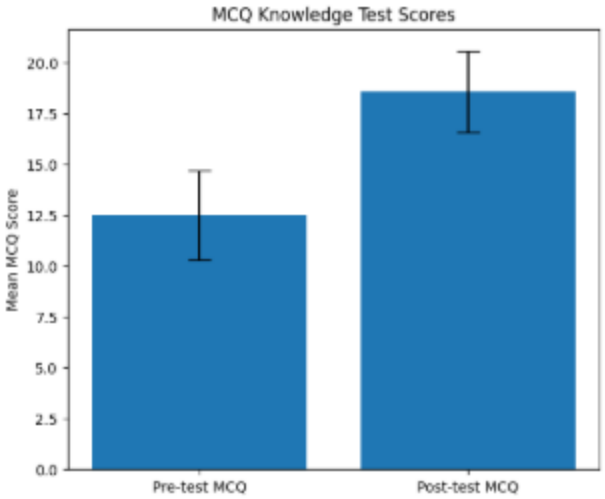


Figure 6

MCQ knowledge test scores across phases.

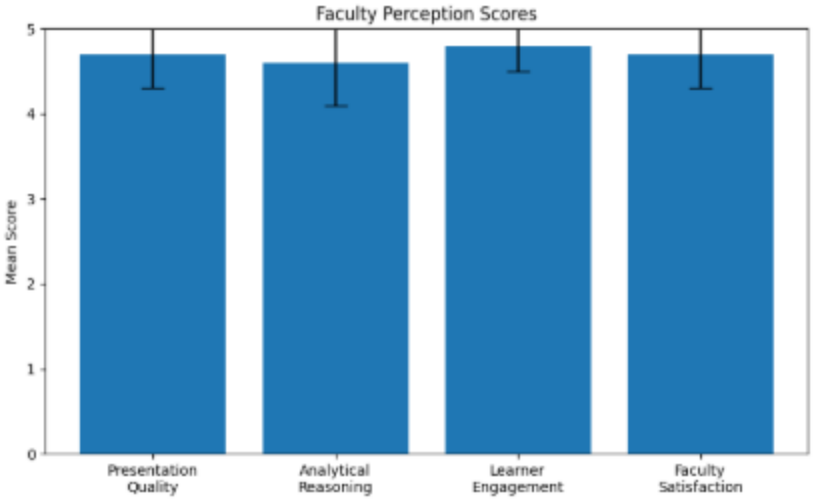


Figure 7

Faculty perception scores across domains.

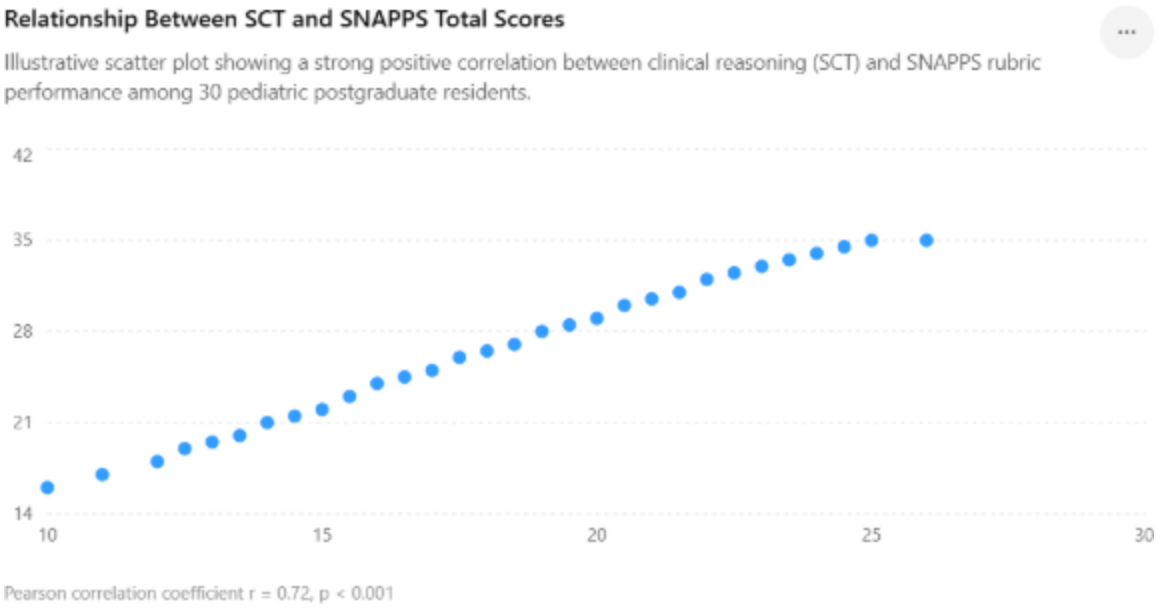


Figure 8

Relationship between SCT and SNAPPS total scores.

Supplementary Files

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

- SNAPPSAnnexures.docx