

Enhancing Cognitive Reflection in Medical Trainees and Physicians: An EBM–CBL–PBL Integrated Teaching Model

Xianzhi Chen¹, Lei Xu², Manman Xu², Deshun Liu³, Fang Yang¹, Guobao Xu¹, Qizhu Feng⁴, Huaicheng Yang⁴

¹Department of Thyroid and Breast Surgery, The First Hospital of Anhui University of Science & Technology, Huainan, Anhui, 232007, People's Republic of China; ²Department of Nuclear Medicine, The First Hospital of Anhui University of Science & Technology, Huainan, Anhui, 232007, People's Republic of China; ³Medical Imaging Department, The First Hospital of Anhui University of Science & Technology, Huainan, Anhui, 232007, People's Republic of China; ⁴General Surgery, The First Hospital of Anhui University of Science & Technology, Huainan, Anhui, 232007, People's Republic of China

Correspondence: Lei Xu, Department of Nuclear Medicine, The First Hospital of Anhui University of Science & Technology, Huainan, Anhui, 232007, People's Republic of China, Tel +86-0554-6616760, Email 619115747@qq.com

Background: Insufficient cognitive reflection in medical trainees and early-career physicians is a root cause of diagnostic errors, yet research on enhancing cognitive reflection in complex surgical settings remains limited. This study aimed to evaluate whether an integrated evidence-based medicine, case-based learning, and problem-based learning (EBM–CBL–PBL) teaching model can improve cognitive reflection.

Methods: A two-phase design was employed. Cognitive reflection was assessed using the Cognitive Reflection Test (CRT). A generalized linear mixed model analyzed associations between sex, clinical experience, and CRT performance. In the intervention phase, trainees were randomized to traditional teaching (n = 18) or an EBM–CBL–PBL group (n = 18). CRT response patterns (intuitive vs. reflective) and teaching satisfaction were compared between groups.

Results: 41.18% of medical trainees exhibited an “All-Reflective” response, whereas 11.76% exhibited an “All-Intuitive” response; among resident physicians, the corresponding rates were 51.85% and 14.81%, and among attending/consulting physicians, they were 57.14% and 8.57%, respectively. No significant overall gender difference was observed (P > 0.05); however, senior female physicians demonstrated significantly stronger analytical thinking (odds ratio [OR] = 5.92, P = 0.005). The “All-Reflective” response rate in the EBM–CBL–PBL group (61.11%) was significantly higher than that in the traditional teaching group (44.44%), with an OR of 3.39 (P = 0.003); furthermore, teaching satisfaction improved significantly in the EBM–CBL–PBL group (P < 0.001).

Conclusion: Cognitive reflection ability improves with increasing clinical experience in clinical practice, with senior female physicians demonstrating the most pronounced enhancement. The integrated EBM–CBL–PBL teaching model adopted in the surgical field shows promise for fostering cognitive reflection. Sustained educational exposure and long-term clinical practice appear essential for consolidating these cognitive gains.

Keywords: cognitive reflection test, clinical reasoning, medical education, generalized linear mixed model, dual-process theory

Introduction

The clinical clerkship phase represents a pivotal stage in medical education, where students transition from theoretical learning to real-world patient care and begin developing diagnostic reasoning. Accurate diagnosis enables timely testing and treatment, underpinning effective disease management and patient safety. Central to this process is clinical reasoning: the cognitive ability to collect and interpret patient information, generate differential diagnoses, and make informed management decisions.¹

Diagnostic decision-making involves two complementary thinking modes according to dual-process theory: intuitive (System 1) and analytical (System 2).^{2,3} Intuitive thinking is rapid, automatic, and pattern-based, well suited for routine

or time-pressured situations but prone to cognitive and emotional biases. Cognitive reflection is deliberate, effortful, and rule-based, advantageous in complex or unfamiliar cases. Clinicians shift between modes depending on case difficulty, expertise, and confidence. Evidence suggests that novice clinicians tend to rely disproportionately on intuitive reasoning.⁴ Approximately 75% of diagnostic errors have been attributed to deficiencies in clinical reasoning, including insufficient knowledge, incomplete data gathering, and inadequate hypothesis testing.⁵ However, the traditional view that novices rely heavily on intuition while experts primarily engage in analysis has been challenged.⁶ Diagnostic errors do stem from reasoning deficiencies, but these often reflect a lack of available knowledge rather than fundamental flaws in the reasoning process itself. The two modes of reasoning—intuition and analysis—should be seen less as independent processes and more as consequences of retrieved knowledge.⁷ Experts develop efficient pattern recognition from rich illness scripts, yet continue analytic reasoning when presentations are unfamiliar;⁸ neuroimaging shows that expertise enhances rather than diminishes intuitive processing.⁹ Thus, clinical reasoning develops as a progressive calibration between intuition and analysis, mediated by deliberate practice – not a linear shift.

Despite growing attention to patient safety, clinical reasoning is not consistently an explicit educational objective; many assume it develops naturally through experience.^{10–12} However, systematic reviews have demonstrated that instructional strategies explicitly integrating both analytical and non-analytical reasoning significantly improve diagnostic accuracy.^{13,14} Deeper understanding of cognitive patterns and reasoning pitfalls is therefore essential for optimising training and educational design.

The CRT is a brief, three-item scale that measures an individual's ability to override intuitive (System 1) responses and engage Cognitive reflection (System 2).¹⁵ As measured by the CRT, cognitive reflection is a validated indicator of System 2 (analytical) thinking within the dual-process framework.⁴ Standard CRT scoring distinguishes intuitive from reflective responses, and accumulating evidence suggests that the CRT predominantly measures reflective analytical reasoning capacity.¹⁶ Studies in early-stage clinical learners indicate that, although increasing expertise is associated with more efficient and automated Cognitive reflection, novices require structured training to deliberately strengthen this ability.¹⁷ However, surgical clinical reasoning is deeply embedded in domain-specific knowledge structures, including contextual awareness, procedural experience, anatomical visuospatial judgment, and tactile feedback. Patel et al found that clinicians rely on their own domain knowledge for diagnosis, and diagnostic accuracy declines significantly when they step outside their specialty.¹⁸

To address this gap, medical education has introduced “cognitive forcing strategies”, such as guided reflection—effectively improve diagnostic accuracy.^{19,20} The integrated EBM–CBL–PBL model acts as cognitive debiasing training: it requires students to construct differential diagnoses, reflect on assumptions, and appraise evidence, thereby reducing anchoring bias and premature closure.^{21–23} In surgical education, this model strengthens analytical ability and decision stability in complex clinical situations through case reasoning and evidence-based reflection. The “expert-novice paradox” suggests that while increased experience enhances pattern recognition efficiency, it may also lead experts to over-rely on intuition and increase the risk of cognitive bias; novices, in contrast, rely more on explicit analytical processes.²⁴ Therefore, it is crucial to establish reflective thinking and cognitive debiasing skills early in clinical training. The EBM–CBL–PBL integrated teaching model can serve as a specific cognitive forcing function, structurally pushing learners to go beyond their first intuition, verify and correct initial judgments, thereby consolidating reflective decision-making ability in early-career physicians.²³

The present study aims to evaluate the performance of intuitive (System 1) and Cognitive reflection (System 2) across different stages of clinical training using the CRT. In addition, we assess the impact of an integrated EBM–CBL–PBL teaching model on clinical interns' decision-making abilities. By analyzing trends across training levels and exploring potential sex-based differences in cognitive processing, we sought to provide evidence-based insights to inform educational strategies for strengthening clinical reasoning and improving diagnostic accuracy in medical education.

Material and Methods

Study Design and Participants

This study adopted a two-phase design to investigate cognitive characteristics and educational interventions in medical training at the first hospital of Anhui university of science and technology. The study population comprised medical

students undertaking clinical clerkships and hospital- employed clinicians at different career stages. Data collection was conducted between September and December 2024 ([Supplementary Figure 1](#)).

Phase I: Cross-Sectional Assessment of Cognitive Reflection

In Phase I, a total of 200 questionnaires were distributed, of which 93 were completed and included in the final analysis. Participants consisted of 34 clinical trainees (22 males and 12 females; mean age, 22.09 ± 1.62 years), 27 resident physicians (16 males and 11 females; mean age, 28.41 ± 2.12 years), 21 attending physicians (12 males and 9 females; mean age, 32.57 ± 3.03 years), and 11 consultant-level physicians (5 males and 6 females; mean age, 51.29 ± 8.82 years).

Participants voluntarily completed the CRT via questionnaires distributed by short message service (SMS). Completion of the questionnaire was considered to imply informed consent, and all responses were collected anonymously. Eligibility criteria included a minimum of two years of clinical practice for clinicians or active participation in clinical placements for students. Individuals with prior exposure to the CRT or unexplained absenteeism during the study period were excluded. Ethical approval for the study was obtained from the Ethics Committee of the First Affiliated Hospital of Anhui University of Science and Technology.

Phase II: Educational Intervention Study

In Phase II, 36 general surgery trainees (21 males and 15 females; mean age, 21.72 ± 1.04 years) were enrolled in a randomized educational intervention. Participants were allocated using a random number table method to either a conventional teaching group or an integrated EBM–CBL–PBL teaching group. The conventional group underwent standard case-based teaching throughout the entire period, whereas the EBM–CBL–PBL group replaced some passive or repetitive elements with structured small-group evidence appraisal and reflective exercises. Following completion of the instructional program, participants received an SMS link to complete both the CRT questionnaire and a teaching satisfaction survey.

Teaching Methods

Conventional Teaching Group

The conventional teaching group followed the standard clinical teaching curriculum. Instructors selected representative cases from hospitalized patients and guided students through bedside learning activities. Teaching focused on disease etiology, pathogenesis, diagnostic approaches, and treatment principles. Students were encouraged to integrate theoretical knowledge with clinical practice to gradually develop clinical reasoning and problem-solving skills. The total instructional time for this group was 18 credit hours.

EBM–CBL–PBL Integrated Teaching Group

The EBM–CBL–PBL integrated teaching group incorporated principles of evidence-based medicine (EBM), case-based learning (CBL), and problem-based learning (PBL). Instructors designed open-ended questions based on selected CBL cases, addressing disease etiology, pathogenesis, differential diagnosis, treatment strategies, and prevention. Common but initially missing positive clinical signs and investigation results were deliberately supplemented to stimulate diagnostic reasoning. Students worked in small groups to search and appraise relevant literature using textbooks and the “5S” pyramid model of evidence-based medicine.²⁵ The literature search followed a structured EBM evidence pathway, progressing from systematic reviews to clinical guidelines. Each teaching cycle concluded with group discussions and student presentations ([Supplementary Appendix 1](#)). The total instructional duration for this group was also 18 credit hours.

Questionnaires

Cognitive Reflection Test (CRT)

Cognitive reflective ability was assessed using an internationally validated three-item Cognitive Reflection Test administered via a QR-code-based electronic survey ([Supplementary Figure 2](#)). Each item elicited an intuitive but incorrect response and a reflective, analytical correct response ([Table S1](#)).

Teaching Satisfaction Survey

Teaching satisfaction was evaluated within 24 hours after the intervention using a Likert-scale questionnaire.²⁶ Students rated five domains: overall satisfaction with the teaching method, depth of content comprehension, proactive learning engagement, teacher–student interaction, and classroom atmosphere.²⁷ The satisfaction survey is provided in [Supplementary Figure 3](#).

Statistical Analysis

All statistical analyses were performed using SPSS Statistics software (version 22.0; SPSS Inc., Chicago, IL, USA). Continuous variables are presented as medians with interquartile ranges (IQRs) and were compared using the Mann–Whitney *U*-test. Categorical variables are expressed as counts and percentages and were analyzed using χ^2 -tests or corrected χ^2 -tests, as appropriate.

Using individual CRT items as the primary analytical units, generalized linear mixed models (GLMMs) were applied to examine associations between cognitive performance and participant characteristics. Correct CRT responses were treated as the dependent variable, while gender and participant group were included as explanatory variables. All statistical tests were two-tailed, and a P-value < 0.05 was considered statistically significant.

Results

Characteristics of Analytical Thinking Across Gender and Clinical Experience Levels

The three CRT items (CRT-Q1, CRT-Q2, and CRT-Q3) increased progressively in difficulty; all items' raw counts and proportions are available in [Tables S2](#) and [S3](#). We defined CRT-Qall as a composite outcome indicating consistently reflective (non-intuitive) responses to all three questions. Reflective responses (CRT-Reflective) were interpreted as indicators of analytical thinking. Distinct patterns of analytical reasoning emerged across levels of clinical experience ([Table 1](#)).

Table 1 Summary of CRT Questionnaire Results Across Different Clinical Experience Groups

Group		Trainees	Resident Physicians	Attending Physicians	Consultant Physicians	Pearson χ^2	P-value
N		34	27	21	14		
Age (years)		22.09±1.62	28.41±2.12	32.57±3.03	51.29±8.82		
Male (n, %)		22 (64.71)	16 (59.26)	12 (57.14)	6 (42.86)	1.969	0.579
Q1	CRT-Reflective (n,%)	22 (64.71)	17 (62.96)	14 (66.67)	8 (57.14)		
	CRT-Intuitive (n,%)	10 (29.41)	8 (29.63)	6 (28.57)	6 (42.86)		
	Incorrect (n,%)	2 (5.89)	2 (7.41)	1 (4.76)	0		
Q2	CRT-Reflective (n,%)	24 (70.59)	21 (77.78)	17 (80.94)	9 (64.29)		
	CRT-Intuitive (n,%)	6 (17.65)	5 (18.52)	2 (9.53)	3 (21.83)		
	Incorrect (n,%)	4 (11.76)	1 (3.70)	2 (9.53)	2 (13.88)		
Q3	CRT-Reflective (n,%)	25 (73.53)	20 (74.07)	16 (76.19)	10 (71.42)		
	CRT-Intuitive (n,%)	6 (17.65)	5 (18.52)	3 (14.29)	2 (14.29)		
	Incorrect (n,%)	3 (8.82)	2 (7.41)	2 (9.52)	2 (14.29)		
All-Reflective (n,%)		14 (41.18)	14 (51.85)	13 (61.90)	7 (50.00)		
All-Intuitive (n,%)		4 (11.76)	4 (14.81)	1 (4.76)	2 (14.29)		
CRT-Mixed (n,%)		16 (47.06)	9 (33.33)	7 (33.34)	3 (35.71)		

Notes: CRT-Reflective represents analytical thinking answers; CRT-Intuitive represents intuitive thinking answers; Incorrect indicates wrong answers; All-Reflective means that all three CRT questions were answered analytically; All-Intuitive means that all three CRT questions were answered intuitively; CRT-Mixed indicates mixed answer patterns.

Among early clinical trainees, 41.18% (n = 14) demonstrated consistent analytical responses across all three CRT items, whereas 11.76% (n = 4) relied exclusively on intuitive responses. Notably, 78.57% of male trainees (n = 11) achieved perfect CRT-Reflective scores. At the individual question level, analytical responses predominated (CRT-Q1: 64.71%, n = 22; CRT-Q2: 70.59%, n = 24; CRT-Q3: 73.53%, n = 25), with males accounting for 68.18–76.00% of analytical responders. Intuitive responses accounted for 17.65–29.41% across questions, with male representation ranging from 33.33% to 50.00%.

Resident physicians demonstrated improved analytical consistency. More than half (51.85%, n = 14) answered all CRT questions analytically, while 14.81% (n = 4) remained entirely intuitive. Males constituted 57.14% (n = 8) of residents with perfect analytical performance. Analytical response rates increased further at the question level (CRT-Q1: 64.71%, n = 17; CRT-Q2: 77.78%, n = 21; CRT-Q3: 74.07%, n = 20), with males comprising 58.82–65.00% of analytical responders. Intuitive responses declined to 18.52%–29.63%, with male representation between 40.00% and 50.00%.

Attending and consultant-level physicians exhibited the strongest analytical performance overall. A total of 57.14% (n = 20) achieved fully reflective CRT-Qall scores, whereas only 8.57% (n = 3) provided exclusively intuitive responses. Interestingly, male representation among perfect analytical scorers declined to 50.00% (n = 10). Although analytical response rates remained high across questions (62.86–74.29%), the proportion of males among analytical responders progressively decreased from CRT-Q1 (50.00%, n = 11) to CRT-Q3 (38.46%, n = 10). In contrast, males accounted for 58.33–80.00% of intuitive responses, particularly in CRT-Q3 (80.00%, n = 4 of 5).

Gender Differences in CRT-Reflective Performance

No statistically significant differences in CRT-Reflective performance were observed between male and female participants. Mann–Whitney *U*-tests and chi-square analyses for each CRT item (CRT-Q1 to CRT-Q3) and the composite outcome (CRT-Qall) showed no significant sex-based differences (all $P > 0.05$; Table 2).

Specifically, 64.3% of males (36/56) and 62.5% of females (25/40) demonstrated analytical thinking in CRT-Q1 ($P = 0.858$). Similarly, analytical responses in CRT-Q2 were observed in 78.6% of males (44/56) and 67.5% of females (27/40), with no statistically significant difference ($P = 0.223$). CRT-Q3 and CRT-Qall also showed no significant sex-related differences ($P = 0.455$ and $P = 0.492$, respectively). Although males demonstrated slightly higher analytical response rates in CRT-Q2 and CRT-

Table 2 R×C Chi-Square Test Results of CRT-Reflective by Gender

CRT Measure		Age Median (IQR)	Gender		Test Populations			
			Male (n=56)	Female (n=40)	Trainees (n=34)	Residents (n=27)	Attendings (n=21)	Consultant-Level (n=14)
CRT-Q ₁	CRT–Reflective	28.0(23.0;32.0)	36	25	22	17	14	8
	Non-Reflective	28.0(23;38.0)	20	15	12	10	7	6
	Mann Whitney U/X ²	999.000 ^a	0.032 ^b		0.360 ^b			
	P value	0.601	0.858		0.948			
CRT-Q ₂	CRT–Reflective	29.0(23.0;34.0)	44	27	24	21	17	9
	Non-Reflective	26.0(23.0;31.0)	12	13	10	6	4	5
	Mann Whitney U/X ²	876.500 ^a	1.485 ^b		1.618 ^b			
	P value	0.927	0.223		0.655			
CRT-Q ₃	CRT–Reflective	29.0(23.0;34.0)	43	28	25	20	16	10
	Non-Reflective	26.0(23.0;31.0)	13	12	9	7	5	4
	Mann Whitney U/X ²	840.000 ^a	0.558 ^b		0.104 ^b			
	P value	0.691	0.455		0.991			

(Continued)

Table 2 (Continued).

CRT Measure		Age Median (IQR)	Gender		Test Populations			
			Male (n=56)	Female (n=40)	Trainees (n=34)	Residents (n=27)	Attendings (n=21)	Consultant-Level (n=14)
CRT-Q _{All}	CRT-Reflective	29.0(23.0;34.5)	29	19	14	14	13	7
	Non-Reflective	26.0(23.0;32.5)	27	21	20	13	8	7
	Mann Whitney U/X ²	1046.000 ^a	0.473 ^b		2.286 ^b			
	P value	0.436	0.492		0.515			

Notes: ^aMann-Whitney U-test; ^bR×C chi-square test.

Q3 (approximately 8–11%), these differences were not statistically significant, indicating overall gender homogeneity in analytical thinking performance.

Interaction Effects Between Gender and Clinical Experience

A generalized linear mixed model (GLMM) was constructed with CRT-Reflective performance as the dependent variable and gender and clinical experience level as explanatory variables (Table 3). A significant interaction effect between gender and clinical experience was identified ($P = 0.005$).

Table 3 GLMM Analysis of Factors Associated with Analytical Thinking

Variable	β	SE	χ^2	P value	OR	95% CI
Fixed effects						
Intercept	-0.814	0.4618	3.108	0.078	0.443	0.179–1.059
Test populations (attending physicians or higher)						
Trainees	-0.721	0.6290	1.314	0.252	0.486	[0.142, 1.668]
Residents	-0.432	0.6588	0.429	0.512	0.649	[0.179, 2.362]
Gender (Male)						
Female	-0.545	0.5921	0.848	0.357	0.580	0.182–1.850
CRT test (CRT-Q3)						
CRT-Q1	0.662	0.5951	1.239	0.266	1.940	[0.604, 6.226]
CRT-Q2	-0.099	0.6226	0.025	0.874	0.906	[0.267, 3.070]
Interactions of test populations and gender (male trainees)						
Female attending and above	1.778	0.6306	7.951	0.005*	5.919	[1.720, 20.373]
Female residents	0.997	0.6683	2.228	0.036*	2.711	[0.732, 10.047]
Interactions of CRT test and gender (male answer CRT-Q3)						
Female answer CRT-Q1	-0.271	0.6462	0.175	0.675	0.763	[0.215, 2.707]
Female answer CRT-Q2	0.228	0.6753	0.114	0.736	1.256	[0.334, 4.718]
Interactions of test populations and CRT test (attending or higher physicians answer CRT-Q3)						
Trainees answer CRT-Q1	-0.101	0.7607	0.018	0.895	0.904	[0.204, 4.016]
Trainees answer CRT-Q2	0.150	0.7901	0.036	0.849	1.162	[0.247, 5.467]
Residents answer CRT-Q1	-0.022	0.7958	0.001	0.978	0.978	[0.206, 4.653]
Residents answer CRT-Q2	-0.216	0.8467	0.065	0.798	0.806	[0.153, 4.234]

Notes: *Bold p-values indicate statistical significance ($p < 0.05$).

Female attending or consultant-level physicians demonstrated the strongest tendency toward analytical thinking, with an odds ratio (OR) of 5.919 (95% CI, 1.720–20.373) compared with male interns as the reference group. Female residents also exhibited a significantly greater likelihood of reflective thinking (OR = 2.711, $P = 0.036$). In contrast, no significant gender difference was observed among interns (female vs male: OR = 0.580, $P = 0.357$).

No significant differences were detected across CRT question types (Q1–Q3) ($P = 0.65$), and no interaction effect was found between clinical experience level and question difficulty ($P = 0.79$). These findings suggest that senior female physicians may be more inclined to adopt analytical reasoning strategies in clinical decision-making, an advantage not observed at earlier training stages.

Value of the EBM–CBL–PBL Pedagogy in Developing Cognitive Reflection

Baseline characteristics, including age, gender, and pre-intervention CRT performance (CRT-Q1, CRT-Q2, CRT-Q3, and CRT-Qall), did not differ significantly between the conventional teaching group and the EBM–CBL–PBL integrated teaching group (all $P > 0.05$). However, overall teaching satisfaction was significantly higher in the EBM–CBL–PBL group ($P < 0.001$).

In the conventional teaching group, 44.44% of students demonstrated consistent analytical thinking across all three CRT items, whereas this proportion increased to 61.11% in the EBM–CBL–PBL group. The proportion of exclusively intuitive responders decreased from 22.22% in the conventional group to 16.67% in the integrated teaching group, indicating a reduced reliance on intuitive reasoning.

Analysis by question difficulty showed higher analytical response rates in the EBM–CBL–PBL group for CRT-Q1 (66.67% vs 61.11%), CRT-Q2 (77.78% vs 55.56%), and CRT-Q3 (83.33% vs 61.11%) compared with the conventional teaching group. Although these differences did not reach statistical significance ($P = 0.248, 0.080, \text{ and } 0.068$, respectively), a consistent trend favoring the integrated pedagogy was observed (Table 4).

A generalized linear model further demonstrated that students exposed to the EBM–CBL–PBL teaching model were significantly more likely to exhibit enhanced analytical thinking (OR = 3.39; 95% CI, 1.52–7.56; $P = 0.003$). No significant effects were found for gender or for interaction terms involving teaching method \times gender (all $P > 0.05$; Table 5). These findings suggest that while the EBM–CBL–PBL model effectively enhances reflective cognitive skills, it does not preferentially amplify gender-specific advantages in CRT-Reflective performance.

Table 4 Single Factor Analysis of EBM–CBL–PBL Pedagogy Outcomes

Variable	Traditional (n=18)	EBM–CBL–PBL (n=18)	Mann–Whitney U/ χ^2	P value
Age (years)	22.0 (21.0–23.0)	22.0 (21.0–22.25)	150.000 ^a	0.690
Male (%)	11 (61.11)	10 (55.56)	0.114 ^b	0.735
Analytical Responses				
CRT-Q1 (%)	12 (66.67)	15 (83.33)	1.33 ^b	0.248
CRT-Q2 (%)	9 (50.00)	14 (77.78)	3.06 ^b	0.080
CRT-Q3 (%)	10 (55.56)	15 (83.33)	3.33 ^b	0.068
CRT-QAll (%)	8 (44.44)	11 (61.11)	2.22 ^b	0.136
Satisfaction Score	69.0 (62.0–72.0)	74.0 (71.5–78.0)	58.500 ^a	<0.001

Notes: ^aMann–Whitney U-test; ^bR \times C chi-square test.

Table 5 Interactive Effects of Gender and Teaching Mode on CRT-Reflective Performance

Variable	β	SE	χ^2	p-value	OR	95% CI
Intercept	0.85	0.32	7.12	0.008*	2.34	[1.25, 4.38]
Teaching group [EBM–CBL–PBL]	1.22	0.41	8.84	0.003*	3.39	[1.52, 7.56]
Gender (Female)	−0.31	0.38	0.67	0.413	0.73	[0.35, 1.54]
Interactions of teaching group and gender	0.45	0.53	0.72	0.396	1.57	[0.56, 4.41]

Notes: *Bold p-values indicate statistical significance ($p < 0.05$). GLMMs: Generalized linear mixed models.

Abbreviations: CRT, Cognitive Reflection Test; GLMM, generalized linear mixed model; EBM, evidence-based medicine; CBL, case-based learning; PBL, problem-based learning; SMS, short message service; IQR, interquartile range; OR, odds ratio.

Discussion

The cultivation of clinical thinking during preclinical medical education is critically important. Existing evidence suggests that males may, on average, exhibit an early advantage in logical reasoning and analytical performance, potentially due to a greater propensity for rapid logical integration.²⁸ Conversely, females often demonstrate stronger potential for progressive and sustained learning,²⁹ differences that may be shaped by physiological maturation, cognitive development, and sociocultural factors. In medical education—especially regarding clinical diagnosis and decision-making—analytical ability plays a decisive role in the accuracy and reliability of clinical judgments.³⁰

Clinical experience appears to be a critical determinant in the maturation of analytical reasoning.

Novice trainees tend to rely predominantly on intuitive reasoning, often drawing conclusions through pattern recognition or rapid judgment.³¹ Although efficient, this intuitive approach is highly susceptible to cognitive biases, thereby increasing the risk of diagnostic errors.³² With prolonged clinical exposure, however, trainees gradually develop more structured cognitive frameworks involving deliberate information gathering, hypothesis generation and testing, logical inference, and critical appraisal,³³ thereby demonstrating excellent analytical reasoning abilities closely related to the accumulated clinical experience and reducing diagnostic errors.³⁴

Our findings are consistent with this progression. While male participants showed an early advantage in cognitive reflection—potentially reflecting innate or early-acquired logical reasoning tendencies—senior female physicians demonstrated particularly strong analytical performance during clinical decision-making, in some cases surpassing their male counterparts. Notably, this advantage among senior female physicians appears to derive primarily from long-term clinical experience rather than from formal medical school training alone. These results suggest that sustained clinical practice may compensate for or even override initial gender-related differences, substantially enhancing analytical capacity in female physicians. However, Due to the potential influence of prior exposure and familiarity on CRT records,³⁵ this led to a relatively small sample size of senior female doctors (n=11) obtained from a single center. Therefore, future studies need to collect samples from multiple centers for observation.

Although short-term educational intervention measures, such as the EBM–CBL–PBL model that integrates evidence-based medicine, case-based learning, and problem-based learning, can effectively enhance the overall analytical ability of preclinical interns, there is no significant difference in the improvement of analytical thinking between male and female participants. No persistent gender-specific advantages were observed. This indicates that although structured educational intervention can enhance analytical reasoning ability in the short term, it may not be sufficient to produce long-term, gender-related cognitive style differences. This discovery offers significant insights for the design of medical education. In addition to classroom teaching, more emphasis should be placed on experiential learning based on real clinical scenarios, such as clinical internships, structured case discussions, and simulation training. Educators should prioritize the cultivation of students' analytical thinking skills, enabling them to conduct more effective clinical reasoning and decision-making throughout their careers. Although the basic course teaching models such as EBM–CBL–PBL play an important role in developing analytical thinking, long-term clinical experience is particularly beneficial for improving the analytical abilities of female physicians.

Noteworthy, The CRT uses decontextualized mathematical or logical puzzles rather than specialized clinical scenarios, and its ecological validity is questionable in the medical context, as clinical reasoning involves knowledge organization, disease script activation, and context-dependent judgment. The general CRT can only partially capture.^{4,15,19} Although the CRT is an effective proxy indicator of the tendency to “suppress intuitive responses” in the dual-process theory, it mainly measures reflective analytical ability rather than general intelligence.¹⁶ Given that domain-specific experience can modulate the dual processing process,³⁶ future research should examine whether the improvement in reflective ability reflected by the CRT can translate into an improvement in domain-specific diagnostic accuracy.

Given that cognitive bias is a significant factor in diagnostic errors, researchers are increasingly seeking “de-biasing strategies” that can be incorporated into training courses. One influential approach is the cognitive forcing strategy, which requires clinicians to consciously apply structured steps to consider alternative diagnoses beyond their initial intuitive impressions.^{37,38} Lambe et al's systematic review found that the cognitive forcing strategy improved diagnostic accuracy and confidence judgments, with guided reflective intervention being the most consistently effective method.¹⁹ The EBM–

CBL–PBL integrated teaching model in this study actually incorporates similar forcing functions, combined with the “expert–novice paradox”, and this model promotes learners to go beyond initial intuitions through structured learning, consolidating reflective decision-making abilities in early training. In conclusion, the de-contextualized CRT cannot fully predict surgical clinical reasoning, but as an effective proxy indicator of analytical thinking tendencies, combined with the cognitive forcing strategy framework, it still provides a valuable assessment tool for this study; future research should adopt domain-specific tools and longitudinal follow-up designs to verify the transformation of reflective abilities into clinical outcomes.

Conclusion

Cognitive reflection in clinical reasoning strengthens gradually with accumulating clinical experience. Although no overall gender differences were observed, a small-sample analysis indicated that senior female physicians exhibited particularly strong cognitive reflection skills. Based on the CRT tool for non-clinical logical assessment, A teaching model integrating EBM–CBL–PBL instruction effectively enhances reflective thinking and learning satisfaction among pre-clinical and early-career trainees. Overall, A dual strategy in medical education: early structured training to cultivate cognitive reflection, combined with long-term, high-quality clinical experience to consolidate and maintain these skills. This approach may improve diagnostic accuracy, promote fairness in cognitive development, and ultimately enhance the quality and safety of clinical decision-making.

Abbreviations

CRT, Cognitive Reflection Test; GLMM, Generalized Linear Mixed Model; EBM, Evidence-Based Medicine; CBL, Case Based Learning; PBL, Problem Based Learning; SMS, Short Message Service; IQR, Interquartile Range; OR, Odds Ratio.

Data Sharing Statement

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request. No publicly archived datasets were generated.

Acknowledgments

We would like to extend our sincere gratitude to the Medical Administration Department and Education Office for providing researcher information. During the preparation of this work, the authors used DeepSeek (version V2) solely for language polishing and grammatical refinement, with no involvement in data analysis, scientific interpretation, or any other aspect that could affect the intellectual content or conclusions of the study. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the final manuscript.

Funding

This work has been funded by the Quality Engineering Program for Higher Education Institutions in Anhui Province (2023) (No. 2023jyxm1073); The Quality Engineering Program for Higher Education Institutions in Anhui Province (2024)(No. 2024jyxm0699) ; 2023 Medical Special Cultivation Project of Anhui University of Science and Technology (No. YZ2023H2C018).Scientific research project of the Health Commission of Anhui Province (No. AHWJ2024BAg30007).

Disclosure

The authors declare no conflicts of interest in this work.

References

1. Pelaccia T, Tardif J, Tribby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online*. 2011;16:5890. doi:10.3402/meo.v16i0.5890
2. Epstein RM, Hammond KR. Dual-process theories in clinical decision-making. *J Clin Psychol*. 2011;67(4):456–467. doi:10.1002/jclp.20790
3. Kahneman D. *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux; 2011.
4. Tay SW, Ryan P, Ryan CA. Systems 1 and 2 thinking processes and cognitive reflection testing in medical students. *Can Med Educ J*. 2016;7(2):e97–e103. doi:10.36834/cmej.36777

5. Thammasitboon S, Cutrer WB. Diagnostic decision-making and strategies to improve diagnosis. *Curr Probl Pediatr Adolesc Health Care*. 2013;43(9):232–241. doi:10.1016/j.cppeds.2013.07.003
6. Norman G, Pelaccia T, Wyer P, Sherbino J. Dual process models of clinical reasoning: the central role of knowledge in diagnostic expertise. *Med Educ*. 2024;58(8):900–908.
7. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ*. 2010;44(1):94–100. doi:10.1111/j.1365-2923.2009.03507.x
8. den Brink N V, Holbrechts B, Brand PLP, Stolper ECF, Van Royen P. Role of intuitive knowledge in the diagnostic reasoning of hospital specialists: a focus group study. *BMJ Open*. 2019;9(1):e022724. doi:10.1136/bmjopen-2018-022724
9. Durning SJ, Costanzo ME, Artino AR, et al. Neural basis of nonanalytical reasoning expertise during clinical evaluation. *Adv Health Sci Educ*. 2015;20(5):1209–1223.
10. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med*. 2006;355(21):2217–2225. doi:10.1056/NEJMra054782
11. Croskerry P. A universal model of diagnostic reasoning. *Acad Med*. 2009;84(8):1022–1028. doi:10.1097/ACM.0b013e3181ace703
12. Wartman SA. The empirical challenge of 21st-century medical education. *Acad Med*. 2019;94(10):1412–1415. doi:10.1097/ACM.0000000000002866
13. Schmidt HG, Mamede S. How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Med Educ*. 2015;49(10):961–973. doi:10.1111/medu.12775
14. Cutrer WB, Sullivan WM, Fleming AE. Educational strategies for improving clinical reasoning. *Curr Probl Pediatr Adolesc Health Care*. 2013;43(9):248–257. doi:10.1016/j.cppeds.2013.07.005
15. Frederick S. Cognitive reflection and decision making. *J Econ Perspect*. 2005;19(4):25–42. doi:10.1257/089533005775196732
16. Pennycook G, Cheyne JA, Koehler DJ, Fugelsang JA. Is the cognitive reflection test a measure of both reflection and intuition? *Behav Res Methods*. 2016;48(1):341–348. doi:10.3758/s13428-015-0576-1
17. Vinaykumar N, Gugapriya TS, Kalaiselvi S. Exploring knowledge of cognitive disposition to respond in clinical decision-making among early clinical learners. *Maedica*. 2023;18(2):317–322. doi:10.26574/maedica.2023.18.2.317
18. Patel VL, Groen GJ. Domain specificity and knowledge utilization in diagnostic explanation. In: *Proceedings of the Cognitive Science Society*. 1990.
19. Lambe KA, O'Reilly G, Kelly BD, Curristan S. Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. *BMJ Qual Saf*. 2016;25(10):808–820. doi:10.1136/bmjqs-2015-004417
20. Griffith PB, Doherty C, Smeltzer SC, et al. Education initiatives in cognitive debiasing to improve diagnostic accuracy in student providers: a scoping review. *J Am Assoc Nurse Pract*. 2020;33(11):862–871. doi:10.1097/JXX.0000000000000479
21. Liu X. The effect of EBM–PBL–CBL integrated teaching method in external urinary clinical practice teaching. *Chin J Sci Technol Database Med*. 2023;8:29–32.
22. Liu XX. Application of CPL–PBL–EBM integrated teaching method in standardized training of radiation oncology residents. *Sichuan J Physiol Sci*. 2023;45(12):2455–2458.
23. Chen X, Xu M, Liu D, et al. Enhancing analytical thinking in early-career physicians: evaluating the EBM–CBL–PBL integrated teaching model. *Res Square*. 2025.
24. Groen GJ, Patel VL. Medical problem-solving: some questionable assumptions. *Med Educ*. 1985;19:95–100. doi:10.1111/j.1365-2923.1985.tb01148.x
25. Haynes RB. Of studies, syntheses, synopses, summaries, and systems: the “5S” evolution of information services for evidence-based healthcare decisions. *Evid Based Med*. 2006;11(6):162–164. doi:10.1136/ebm.11.6.162-a
26. Likert R. A technique for the measurement of attitudes. *Arch Psychol*. 1932;140:1–55.
27. Qi Y, Li C, Sun P, et al. Application of PBL combined with 3D anatomy software in clinical internship teaching of traditional Chinese medicine traumatology based on questionnaire survey. *Chin Contin Med Educ*. 2024;16(12):152–156.
28. Chen CS, Knep E, Han A, Ebitz RB, Grissom NM. Sex differences in learning from exploration. *Elife*. 2021;10:e69748. doi:10.7554/eLife.69748
29. Halpern DF. *Sex Differences in Cognitive Abilities*. 4th ed. New York, NY: Psychology Press; 2012.
30. Norman GR, Monteiro SD, Sherbino J, Ilgen JS, Schmidt HG, Mamede S. The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Acad Med*. 2017;92(1):23–30. doi:10.1097/ACM.0000000000001421
31. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ*. 2005;39(1):98–106. doi:10.1111/j.1365-2929.2004.01972.x
32. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ Qual Saf*. 2013;22(Suppl 2):ii58–ii64. doi:10.1136/bmjqs-2012-001712
33. Norman GR, Grierson LEM, Sherbino J, Hamstra SJ, Schmidt HG, Mamede S. Expertise in medicine and surgery. In: Ericsson KA, Hoffman RR, Kozbelt A, Williams AM, editors. *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge, UK: Cambridge University Press; 2018:331–355.
34. Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educ Psychol Rev*. 1993;5:205–221. doi:10.1007/BF01323044
35. Stiegel S, Replus UD. A Limitation of the Cognitive Reflection Test: familiarity. *Peer Reviewed J*. 2016;4:e2395. doi:10.7717/peerj.2395
36. Thompson VA, Pennycook G, Trippas D, Evans JS. Do cognitive reflection tests measure cognitive reflection? A mathematical modeling approach. *Mem Cognit*. 2020;48(4):648–668.
37. Croskerry P. Cognitive forcing strategies in clinical decisionmaking. *Ann Emerg Med*. 2003;41(1):110–120. doi:10.1067/mem.2003.22
38. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 2: impediments to and strategies for change. *BMJ Qual Saf*. 2013;22(Suppl 2):ii65–ii72. doi:10.1136/bmjqs-2012-001713

Advances in Medical Education and Practice

Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>

Dovepress
Taylor & Francis Group