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.

Integrating Diagnostic and Management Reasoning in Medical Genetics Education: A Two-Session Database-Driven and Simulation-Based Intervention for Pre-Clinical Students

Xin Jin

jinxin@usx.edu.cn

Shaoxing University

Yan Zeng

Shaoxing Maternal Shaoxing Maternal and Child Health Hospital

Jieyuan Jin

Shaoxing University

Wenlong Jin

Shaoxing University

Xingliang Fang

Shaoxing University

Research Article

Keywords: clinical reasoning education, medical genetics, genetic counseling, cognitive conflict, database-driven inquiry, simulation-based learning, undergraduate medical students

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10510265/v1

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

Additional Declarations: No competing interests reported.

Integrating Diagnostic and Management Reasoning in Medical Genetics Education: A Two-Session Database-Driven and Simulation-Based Intervention for Pre-Clinical Students

Running head: Clinical reasoning teaching in medical genetics

Jin Xin[1*], Zeng Yan[2], Jin Jieyuan[1], Jin Wenlong[1], Fang Xingliang[1]

[1] School of Medicine, Shaoxing University, Shaoxing, Zhejiang, China

[2] Genetic Laboratory, Shaoxing Maternal and Child Health Hospital, Shaoxing, Zhejiang, China

***Corresponding author:** Jin Xin

School of Medicine, Shaoxing University

1077 Chengnan Road, Shaoxing, Zhejiang 312000, China

Email: jinxin@usx.edu.cn

Tel: +86-13957581518

ORCID: Jin Xin: 0000-0002-1074-4412

Abstract

Background: Clinical reasoning constitutes a core competency for health professions trainees. However, traditional medical genetics instruction faces three major challenges: disconnection from real clinical scenarios, passive student engagement, and insufficient translation of basic science knowledge into clinical reasoning practice. This study evaluated a two-session intervention with distinct but interconnected purposes—diagnostic reasoning in Session 1 and management reasoning in Session 2—to develop both dimensions within a coherent, learner-centered framework.

Methods: A mixed-methods study recruited 30 pre-clinical Chinese medical undergraduates. Session 1 adopted a dual-teacher model with a 1,152-case chromosomal abnormality database for group data inquiry to train diagnostic reasoning. Session 2 delivered simulated genetic counseling with real-time bullet-screen comments to develop management reasoning. Questionnaires, group assignments, counseling scripts, and end-of-course examination scores were analyzed.

Results: All six groups identified clinically significant karyotype patterns not found in standard textbooks. Two groups found distributions deviating substantially from textbook expectations, triggering cognitive conflict and shifting students from passive "answer-seeking" to active "questioning-verifying" reasoning. Discussion board questions progressed from factual recall to analysis (38.9%) and evaluation (22.2%). Diagnostic thinking from Session 1 was applied in Session 2 counseling scripts. In an exploratory comparison with a historical cohort taught by the same instructors using the same assessment, the intervention group scored significantly higher (91.03 ± 4.89

vs 85.25 ± 7.97; t(58) = 3.38, p = 0.0013; Cohen's d = 0.87), and the proportion scoring ≥90 increased from 43.3% to 76.7% (χ^2 = 5.33, p = 0.021).

Conclusions: The progressive two-stage intervention provides a replicable curriculum framework that integrates diagnostic and management reasoning within a single coherent instructional arc. Students demonstrated a shift from passive answer-seeking to active questioning-verifying reasoning, and preliminary evidence from historical cohort comparison suggests measurable gains in examination performance on content directly aligned with the intervention. The model is particularly relevant in the current era, where fostering independent clinical thinking and patient-centered communication has become an urgent educational priority.

Keywords: clinical reasoning education, medical genetics, genetic counseling, cognitive conflict, database-driven inquiry, simulation-based learning, undergraduate medical students

Background

Clinical reasoning—gathering, interpreting, and applying patient information to reach diagnoses and treatment plans—is a core competency in health professions education[1]. Yet it is notoriously difficult to teach in the early years, when students have limited clinical exposure and rely on pattern recognition over analytical thinking[2]. Moreover, clinical reasoning encompasses not only diagnostic reasoning (data interpretation and hypothesis generation) but also management reasoning (communicating findings and engaging in shared decision-making)—a distinction often underemphasized in curricula[3, 4].

Medical genetics, bridging basic and clinical medicine, offers an early opportunity to cultivate these competencies. The Association of Professors of Human and Medical Genetics has established core competencies specifying that students should "interpret the results of a cytogenetic report with respect to common numerical and structural chromosome abnormalities"[5]. Yet traditional instruction in this domain continues to face three persistent challenges[6-8]: disconnection from clinical practice, passive learning, and failure to transfer foundational knowledge to clinical reasoning tasks.

Recent genetics education innovations—including case-based active learning with OMIM navigation[9], three-touch curricula with online clerkship modules[10], and student experience-based approaches using personal phenotypes[11] —have advanced student engagement. However, these interventions predominantly target information acquisition or concept demonstration, leaving two distinct gaps. **First**, few designs require students to reason diagnostically from unfiltered clinical data; most provide pre-

interpreted cases, depriving learners of the authentic cognitive work of pattern discovery. **Second**, diagnostic reasoning and management reasoning are typically taught separately or sequentially without explicit integration, fragmenting the clinical reasoning arc[3]. Few curricula integrate diagnostic and management reasoning into sequential learning modules. This study addresses this gap via a two-session framework: Session 1 uses raw clinical datasets to foster diagnostic reasoning through cognitive conflict, while Session 2 translates those diagnostic findings into simulated genetic counseling practice to build management competencies. This study asked: (1) whether foundational-year students can develop diagnostic reasoning through database-driven inquiry; and (2) whether simulation-based counseling can cultivate preliminary management reasoning in the absence of clinical exposure.

Methods

Study design and participants

This study employed a single-group mixed-methods design[2] with 30 undergraduate students enrolled in the Clinical Medicine program at a Chinese university. To provide interpretive context for the single-group findings, de-identified archival assessment data from the immediately preceding cohort (2024 cohort, n = 30) who received traditional lectures were also analyzed. Importantly, the same instructor team delivered the course to both cohorts, and both cohorts completed the same end-of-course examination, thereby controlling for instructor-based and instrument-based confounding. All participants had completed foundational coursework in embryology, with the genetics course approaching its conclusion. The intervention consisted of two

90-minute sessions conducted in May 2026 in a smart classroom with individual computer terminals. Ethics approval was obtained from the Medical Ethics Committee of the Affiliated Hospital of Shaoxing University (Approval No. 2026(shen)-141-01); all participants provided written informed consent.

For mixed-methods exploratory studies in health professions education, a sample size of 30 is consistent with recommendations for achieving data adequacy when drawing on diversified qualitative sources (open-ended survey responses, discussion board posts, bullet-screen comments, and performance artifacts) rather than relying solely on in-depth interviews[12]. For the quantitative comparison with the historical cohort, a sample size of 30 per group achieves approximately 80% power to detect large effect sizes (Cohen's $d \geq 0.8$) at $\alpha = 0.05$ (two-tailed).

Session 1: Database-driven dual-teacher collaborative inquiry (90 min)

Session 1 was grounded in a social constructivist framework[13], with instructors scaffolding students' progression from knowledge recall to diagnostic reasoning. A clinical database—containing 1,152 authentic abnormal karyotypes specifically designed for this teaching session (Software Copyright Registration No. 17625435)—served as a cognitive tool mediating students' engagement with real clinical data. The database was constructed from 1,152 consecutive abnormal karyotypes identified at a regional genetics laboratory, covering all clinical indications for prenatal diagnosis. All cases were de-identified and reviewed by two clinical geneticists before inclusion. An accompanying knowledge graph website displayed cell morphology images.

A basic science faculty member and a clinical genetic counselor co-taught. The basic science teacher delivered theoretical instruction (25 min); the clinical teacher introduced the database through real cases, illustrating how karyotype patterns correlate with clinical phenotypes and highlighting common diagnostic pitfalls. Students were split into six collaborative groups. Each group received a data analytical task (e.g., "What proportion of high-risk screening patients receive a trisomy 21 diagnosis?" and "How do the prevalence rates of chromosomal inversions compare with trisomy 21 among women of advanced maternal age?"), then spent 20 minutes on database retrieval and group deliberation. Every group delivered a 5-minute presentation of their findings (30 minutes in total for all presentations), which was followed by instructor debriefing and the distribution of extended independent research tasks (15 min).

Session 2: Simulation-based genetic counseling (90 min)

One week before Session 2, students received a case vignette describing a 32-year-old pregnant woman with high-risk serum screening and a cousin with trisomy 21. Groups posted questions on a discussion board hosted on Shaoxing Xuexitong (a university smart-classroom platform) and prepared counseling scripts.

In-class, groups performed genetic counseling scenarios (5-8 min per group, approximately 40 min total). Audience members posted real-time comments via a bullet-screen system (Rain Classroom). The instructor provided feedback on each performance, revisited questions posted on the discussion board, and facilitated a

structured discussion of recurring thematic points (approximately 35 min). The session concluded with an in-class quiz and debriefing (15 min).

The overall teaching workflow across the two sessions, separated by a one-week intersession interval for literature review and script preparation, is illustrated in Figure 1.

Data sources and analysis

Data sources included: (1) a post-Session 1 questionnaire (6 items) administered immediately after the session; (2) Session 1 group assignments; (3) discussion board posts from the week preceding Session 2; (4) bullet-screen comments from Session 2; (5) observation notes; (6) submitted counseling scripts from Session 2; and (7) selected items from an end-of-course questionnaire referencing the two-session intervention activities.

The chromosome disorders assessment was a 100-point summative examination consisting of 40 multiple-choice questions (2.5 points each) covering numerical and structural abnormalities, recurrence risk calculation, and genetic counseling scenarios. The same examination was administered to both cohorts. Item-level correctness rates were available for core overlapping items across both cohorts.

To examine whether the intervention effect was consistent across different content areas, we identified seven core content areas that were assessed in both cohorts through questions with identical or highly similar testing focus. Item-level correctness rates for these areas were aggregated by content area rather than by individual question number, and compared between cohorts.

Quantitative data were analyzed using descriptive statistics. Qualitative data—including open-ended questionnaire responses, discussion board posts, and bullet-screen comments—were thematically analyzed by two independent researchers[14] using Braun and Clarke's six-phase framework. Inter-rater reliability for thematic coding was assessed using Cohen's kappa (κ = 0.82, 95% CI [0.74, 0.90]); disagreements were resolved through discussion. Discussion board questions were classified using Bloom's revised taxonomy by two independent coders. All statistical analyses were performed using SPSS 26.0 (IBM Corp., Armonk, NY, USA), with two-tailed tests and $P < 0.05$ considered statistically significant. Normality of assessment scores was confirmed via Shapiro-Wilk test before independent-samples t-test analysis. Cohen's d effect sizes and 95% confidence intervals were calculated for t-test comparisons. Categorical variables (e.g., proportion scoring ≥ 90) were compared using chi-square tests.

Results

Session 1: Diagnostic reasoning through database-driven inquiry

All six groups completed retrieval tasks. Table 1 summarizes the core findings—all generated by students' own database operations. Two groups (Groups 1 and 3) identified patterns that deviated substantially from textbook-based expectations. Group 1 found that only a minority of "high-risk" cases were actually trisomy 21, prompting students to shift from "what is the answer?" to "why is the false-positive rate so high?"—a move from information-seeking to diagnostic questioning. Group 3 found that chromosomal inversions outnumbered trisomy 21 in advanced maternal age,

requiring students to revise their understanding that advanced maternal age primarily signals trisomy 21 risk.

Session-specific questionnaire results. Table 2 presents student evaluations and thematic analysis from the post-Session 1 questionnaire. Notably, half of respondents (50.0%) identified the clinical database platform as the most impressive component, suggesting that direct engagement with authentic clinical data resonated more strongly than instructional format.

A small subset of respondents (n = 3, 10%) expressed initial hesitation regarding the database interface "It took me a while to figure out which filters to apply" and one group reported feeling overwhelmed by the volume of raw data, suggesting that unfiltered data engagement may impose a cognitive load barrier for some students—a finding that informs future scaffolding improvements.

Session 2: Management reasoning through simulation-based counseling

Discussion board questions. Of 18 pre-class questions, 27.8% were at remembering/understanding level, 38.9% at applying/analyzing level, 22.2% at evaluating/creating level, and 11.1% addressed humanities/ethics. Question content progressed from factual recall ("What are AFP, hCG, μE3?") to analysis ("Is recurrence risk mildly or significantly increased?") to evaluation ("Why suggest invasive diagnosis rather than NIPT?") to ethics ("How should trisomy 21 diagnosis be communicated?"). This progression reflected instructor scaffolding: when students posted factual questions, the instructor redirected toward deeper inquiry; when students posted humanities/ethics questions, the instructor affirmed these as "good questions".

Bullet-screen comments. Thirteen students (43.3%) sent 97 comments. Categories: performance evaluation (34.8%), character critique (25.8%), emotional expression (18.0%), professional discussion (12.4%), humor (5.6%) (Figure 2). An additional 4 comments were posted by the instructor during the debriefing phase, each directly reflecting a question that students had raised in the pre-class discussion board—creating a visible link between students' pre-class questions and the in-class debriefing. The emotional peak at 14:42 (9 comments within a 3-minute window vs a session average of 4.2) occurred during Group 3's counseling performance—specifically, at the moment when the simulated "physician" delivered the confirmatory amniocentesis result ("The fetus is diagnosed with trisomy 21"), followed by the "couple" exchanging reassuring words and the "healthcare team" offering empathic support. This scene triggered a surge of emotional comments (e.g., "So touching," "Family atmosphere is perfect"), drawing responses even from typically silent participants, suggesting that high-immersion, emotionally charged scenarios can engage learners who do not otherwise participate in digital platforms.

Platform participation patterns. Platform participation patterns at the group level revealed a distinct picture. Bullet-screen contributions were highly concentrated: Group 3 contributed 52 messages (53.6%) and Group 1 contributed 35 (36.1%), together accounting for 89.7% of all student bullet-screen comments; Groups 4 and 6 posted none. In contrast, discussion-board contributions were more evenly distributed across all six groups (2-4 posts per group). One group (Group 3) dominated bullet-screen engagement while maintaining moderate discussion board activity; another group

(Group 4) was active on the discussion board but silent on bullet-screen. Figure 3 illustrates these group-level participation patterns.

End-of-course feedback. In an end-of-course questionnaire, open-ended responses referencing the two-session intervention confirmed its impact: ten students explicitly mentioned the simulation/role-play activity as the most memorable aspect of the course; five referenced the database/laboratory activities; seven explicitly mentioned developing clinical thinking; and one referenced the instructor's real-time guidance during the inquiry phase. Notably, "role-play/simulation" emerged as a prominent theme (33.3% of open-ended responses), consistent with the intervention's intended emphasis on experiential learning.

Comparative findings from historical cohort data. To provide context for interpreting the single-group findings, we compared the present cohort with the immediately preceding cohort (2024 cohort, n = 30) who completed the same course under traditional lecture-based instruction. Baseline characteristics were comparable across cohorts (gender distribution, age, first-semester GPA), and the same instructor team delivered both cohorts using the same examination instrument. This comparison is limited to the chromosome disorders assessment—the only assessment instrument administered identically across both cohorts and directly aligned with the intervention content.

On this assessment, the present cohort outperformed the historical cohort on total score (91.03 ± 4.89 vs 85.25 ± 7.97; $t(58) = 3.38$, $p = 0.0013$; Cohen's $d = 0.87$, 95% CI [0.54,

1.20]). The proportion scoring ≥ 90 increased from 43.3% to 76.7% ($\chi^2 = 5.33$, $p = 0.021$).

Item-level analysis of seven core content areas that were assessed with identical or highly similar testing focus across both cohorts showed consistent improvement, with correctness rate gains ranging from 13 to 25 percentage points (Table 3). The largest gains (25 percentage points each) were observed for genetic counseling risk calculation (Robertsonian translocation), counterintuitive knowledge (inv(9) as normal variant), and inversion carrier genetics—all topics directly addressed in the intervention. Notably, all seven content areas showed improvement, with not a single area showing a decline (binomial test for directional consistency, $p < 0.01$). Additionally, the standard deviation decreased from 7.97 (2024 cohort) to 4.89 (2025 cohort), and the minimum score increased from 66.8 to 77.4, suggesting that the intervention not only raised average performance but also reduced the performance gap among students.

Discussion

The two-session intervention was designed as a coherent instructional arc: Session 1 positioned students as active inquirers engaging with authentic clinical data, while Session 2 provided a communicative context in which the diagnostic insights generated from that inquiry could be applied. This sequential logic addresses a persistent problem in clinical reasoning education—the fragmentation of diagnostic and management reasoning into separate curricular silos[2, 3]—by demonstrating that both dimensions can be integrated within a single foundational-year course.

Cognitive conflict as a trigger for knowledge restructuring

The instructional design of Session 1 deliberately created conditions for cognitive conflict. Students were given authentic, unfiltered clinical data and asked to derive patterns themselves, rather than receiving pre-interpreted textbook summaries. When their findings diverged from textbook expectations—as occurred in multiple groups' analyses—this discrepancy did not merely correct factual errors; it triggered a deeper shift in how students conceptualized the clinical meaning of test results. "High risk" was no longer equated with "likely diagnosis" but redefined as a screening concept with limited predictive value. This recalibration emerged not from direct instruction but from students' own data operations.

This design aligns with two complementary theoretical frameworks. First, cognitive load theory[15] provides strategies for tailoring clinical reasoning training for novice medical students. By providing unfiltered clinical data without pre-interpreted summaries, the task was designed to increase germane cognitive load—mental effort devoted to schema construction and automation—while reducing extraneous load through instructor scaffolding and structured group collaboration. The cognitive conflict experienced by Groups 1 and 3 thus represents a productive engagement with intrinsic complexity, not unnecessary confusion. Second, Kolb's experiential learning cycle offers a complementary lens. Students began with concrete experience (raw data retrieval), moved to reflective observation (identifying unexpected patterns), engaged in abstract conceptualization (reformulating their understanding of "high-risk"), and finally applied their revised mental models in active experimentation (Session 2 counseling). This cycle is typically absent in lecture-based or pre-interpreted case

formats, which provide students with ready-made concepts before they have encountered the experiential trigger that gives those concepts meaning.

This mechanism—learning through self-generated discovery—distinguishes our approach from conventional case-based or lecture-based genetics instruction[9, 10], where the interpretive work is largely pre-packaged by instructors. For instance, standard genetic counseling simulations often provide students with a pre-determined diagnosis (e.g., “This is a trisomy 21 case”) and focus exclusively on communication skills[9]. Likewise, traditional case-based learning presents curated vignettes with known outcomes, guiding students toward predetermined answers rather than placing them in an epistemically active role of generating and testing hypotheses from raw data. By contrast, our Session 1 required students to independently filter, aggregate, and interpret 1,152 authentic karyotypes—a task that introduces genuine uncertainty and the possibility of unexpected patterns. The resulting cognitive conflict was not scripted by instructors but emerged organically from the data.

A practical implication is that pre-clinical curricula can fruitfully incorporate authentic, unfiltered datasets in place of pre-interpreted cases, provided that appropriate scaffolding—structured group tasks, focused analytical questions, and real-time instructor debriefing—is in place to calibrate task complexity so that uncertainty remains productive rather than overwhelming.

This design is particularly consequential in the current era of generative AI. As AI tools become increasingly capable of producing coherent diagnostic conclusions and communication scripts, instructional design must deliberately target competencies that

316 resist algorithmic substitution: formulating questions from ambiguous data, tolerating
317 uncertainty when patterns deviate from expectations, and exercising judgment in
318 contexts that demand emotional attunement and humanistic care. The bullet-screen data
319 illustrate this last dimension—comment volume peaked during a student's empathetic
320 counseling response to a simulated patient facing an adverse diagnosis, an act of moral
321 imagination that no algorithm can replicate[8, 16]. Moreover, the observation that some
322 students instinctively turned to AI for ready-made answers—and subsequently reflected
323 on the value of independent reasoning when redirected—suggests that prohibitive
324 policies alone are insufficient. A more productive educational response is to design
325 tasks that cannot be outsourced to AI because they require situated judgment, pattern
326 discovery, and human communication, and for instructors to intervene not to police AI
327 use but to reframe students' attention toward what only they can discern.
328 The present model demonstrates that teaching can simultaneously achieve two goals:
329 cultivating the humanistic dimensions of clinical practice—warmth, empathy, and
330 ethical judgment—while also improving measurable learning outcomes, a combination
331 supported by both the examination data and student feedback. This finding resonates
332 with evidence from a parallel domain: Wen et al.[17] demonstrated that an integrated
333 CBL approach incorporating standardized patient encounters in surgical internship
334 significantly enhanced students' humanistic care performance, providing cross-
335 disciplinary convergence with our observation of emotional peaks during Session 2
336 role-play.
337 ***From diagnostic insight to management reasoning: evidence of transfer***

The sequencing of Session 1 and Session 2 was designed to enable cognitive transfer. Diagnostic insights generated under conditions of cognitive conflict in Session 1 were intended to become content for communication in Session 2. The evidence suggests this transfer occurred: concepts and framing from Session 1 reappeared in Session 2 counseling scripts and performances. Students who had personally discovered, for example, that screening tests have limited predictive value were able to communicate this probabilistic information to standardized patients in empathetic, accessible language. The insight was not merely recalled—it was adapted to a communicative context requiring clinical judgment and emotional attunement.

This finding extends the work of Weiler and Landa-Galindez[10], who placed management reasoning instruction in the clinical clerkship phase, and Parsons et al.[18], whose "management script" framework assumes clinical experience as a prerequisite. What enables this nascent form of management reasoning in the absence of clinical experience? We propose that two conditions are essential. First, students must possess a knowledge base that they themselves have actively constructed—rather than passively received—so that they have both ownership and deep understanding of the content to be communicated. Second, the communication context must be low-stakes yet authentic enough to evoke the interpersonal dimensions of clinical practice—empathy, uncertainty disclosure, and shared decision-making—without the performance pressure of real patient encounters. Session 1 provided the first condition through database-driven discovery; Session 2 provided the second through simulated counseling with peer audiences.

Our results suggest that preliminary management reasoning can begin to develop in the foundational years when instructional design provides both (a) a knowledge base worth communicating and (b) a low-stakes context for practicing that communication. The transfer observed here is not yet expert-level management reasoning—but it represents a nascent form of the competence that Cook et al.[3] identified as distinct from diagnostic reasoning—one that is likely to require reinforcement through repeated clinical exposure to fully mature, yet can be meaningfully initiated in the pre-clinical years under the right instructional conditions.

Recent evidence from a scoping review of simulation-based education underscores the importance of designing explicit transfer pathways: the review found that only 11% of studies assessed transfer of clinical decision-making learning to clinical practice, with the vast majority assessing transfer only from one simulation to another[19]. This gap underscores the value of our explicit two-session design, which deliberately created shared clinical content and aligned learning objectives between diagnostic reasoning (Session 1) and management reasoning (Session 2) to facilitate transfer across reasoning domains. The review also highlighted that transfer is not guaranteed by realism or authenticity alone; rather, it depends on learners' knowledge base, motivation, and opportunities for reflective practice[19]—conditions that our sequential design was intended to foster.

Platform participation: understanding learner engagement diversity

The bullet-screen and discussion board data, viewed together, offer a methodological observation about capturing engagement in clinical reasoning education. Synchronous

platforms (bullet-screen) captured reactive, affective responses and were characterized by concentrated participation; asynchronous platforms (discussion board) captured reflective, question-driven contributions and were more evenly distributed. The minimal overlap between contributors to the two platforms suggests that no single format captures all learners' engagement.

This pattern may reflect differences in learners' cognitive and affective orientations toward participation. Some students process information more effectively through reflective writing, which allows time for formulation and revision; others respond more readily to immediate, emotionally charged stimuli that demand spontaneous reaction. The bullet-screen platform afforded the latter, while the discussion board afforded the former. Importantly, neither platform captured the full spectrum of engagement; the students who were digitally silent were not necessarily disengaged, as evidenced by their detailed end-of-course reflections. This suggests that written and oral modalities tap into different dimensions of learning—deliberative reasoning versus affective responsiveness—and that comprehensive assessment of engagement requires triangulation across modalities.

Multiple participation formats thus serve complementary roles in reaching diverse learners. McConnell and Eva[16] found that affective engagement enhances learning transfer, while Roberts et al.[20] demonstrated the effectiveness of pre-clerkship simulation-based clinical reasoning instruction. Our platform complementarity finding extends these lines of evidence by showing that effective early reasoning instruction must consider both cognitive style differences and participation modality preferences—

offering both low-stakes synchronous commenting and high-immersion experiential formats, and recognizing that digitally silent students may become actively engaged during embodied simulation experiences.

Integrating diagnostic and management reasoning: the instructional arc

The two sessions were not designed as two independent pedagogical activities, but as two phases of a single learning process. The educational logic linking them is that the products of diagnostic reasoning become the raw materials for management reasoning—a design principle we term "content-communication continuity". In this model, knowledge is not transmitted and then later applied; rather, it is discovered, internalized through cognitive conflict, and immediately repurposed as content for communicative action. This contrasts with conventional curricula where diagnostic reasoning and communication are often taught in separate courses, leaving students to infer the connections for themselves. This sequential design is grounded in the principle of "learning transfer"—the extent to which knowledge acquired in one context can be applied in another—which is most effectively facilitated when instructional tasks are explicitly sequenced with shared content and aligned objectives.

The coherence of the two-session arc lies in its alignment with the epistemic demands of clinical practice itself. In real clinical work, diagnostic conclusions are not endpoints but starting points—they must be communicated, negotiated, and translated into management decisions. This reflects the broader argument that clinical reasoning is not a linear sequence of discrete steps—"diagnose then manage"—but an integrated, iterative process in which diagnostic hypotheses and management plans co-evolve as

new information emerges[3, 4]. By mirroring this structure in the classroom, the intervention exposes students to the full arc of clinical reasoning in a way that isolated diagnostic or communication exercises cannot replicate.

This integrative design is further supported by converging quantitative evidence. Compared to the historical cohort, the intervention cohort showed significantly higher examination scores ($p = 0.0013$, Cohen's $d = 0.87$), with consistent improvement across all seven core content areas and no decline in any area (binomial test, $p < 0.01$). These performance gains add to the growing body of evidence that integrated curricula—those that deliberately bridge foundational science and clinical application—can enhance learning outcomes when integration is understood as a cognitive activity occurring within learners, rather than merely as a logistical rearrangement of curriculum content[21, 22]. This distinction is critical: Kulasegaram and colleagues argue that the aim of integrating basic and clinical sciences is to achieve a conceptual, cognitive connection between different types of knowledge, and that this connection is most effectively fostered at the level of individual teaching sessions through causal explanations that link physiological mechanisms to clinical features[21]. The session-level design of our intervention—requiring students to derive diagnostic patterns from raw data and then apply those insights in counseling—exemplifies this principle. The reduction in score variability (SD: 7.97 to 4.89) and the elimination of the low-performance tail (minimum score: 66.8 to 77.4) further suggest that the intervention benefited the entire cohort, not only high-performing students. These findings, together

with the qualitative evidence of cognitive conflict and transfer, suggest that the score gains reflect genuine learning improvements rather than assessment artifacts.

Strengths and limitations

Strengths: The mixed-methods design captured both quantitative outcomes and qualitative processes. The two-session structure addressed both diagnostic and management reasoning. Authentic clinical data exposed students to real-world complexity. Triangulation of session-specific questionnaires, end-of-course feedback, and performance-based data strengthened validity. The use of the same instructor team and identical assessment instruments across cohorts strengthened the comparability of the historical comparison.

Limitations: First, the single-group, single-institution design without a concurrent control group limits causal inference and generalizability; multi-institutional studies with concurrent comparison groups are needed to confirm these findings. Second, although the same instructor team and identical assessment instruments were used across cohorts, and GPA, age, and sex matching was completed, unmeasured secular confounders (e.g., cohort-level differences in motivation, academic culture, and baseline genetics-specific prior knowledge) cannot be fully ruled out; inter-cohort score differences should therefore be interpreted as preliminary suggestive evidence rather than definitive causal effects. Third, standardized reasoning instruments (e.g., Diagnostic Thinking Inventory) were not used; assessment relied on a single end-of-course examination and self-report questionnaires, which may be subject to response bias. In addition, the examination contained fewer items directly assessing management

reasoning than diagnostic reasoning—a limitation of the assessment instrument rather than the instructional design—and qualitative data from counseling scripts and discussion boards provided complementary evidence for management reasoning outcomes, partially offsetting this assessment limitation. Fourth, cognitive conflict was observed cross-sectionally at Session 1; longitudinal follow-up (e.g., at clerkship entry) was not conducted to assess the durability of observed epistemic shifts. Fifth, the single-site, small-sample (n = 30) nature of this exploratory study limits statistical precision; effect sizes should be interpreted with correspondingly wide confidence intervals. Sixth, the seven content areas analyzed in Table 3 represent a subset of the 40 examination questions—specifically, those with identical or highly similar testing focus across both cohorts. This subset covers the core topics of the chromosomal disorders module but does not exhaust all content covered in the assessment.

Conclusions

This two-session intervention, with distinct but interconnected purposes, combined database-driven inquiry for diagnostic reasoning with simulation-based counseling for management reasoning. Students reported high satisfaction and perceived learning gains, with the majority endorsing the dual-teacher model and database inquiry approach. Qualitative evidence of transfer from Session 1 to Session 2—students applying data-derived insights in counseling contexts—suggests the coherence of the two-stage design. Quantitative comparison with a historical cohort on the chromosome disorders assessment showed significantly higher scores ($p = 0.0013$, Cohen's $d = 0.87$). The model offers a replicable framework for integrating diagnostic and management

reasoning into foundational curricula, with particular relevance in the current era where fostering independent clinical thinking and patient-centered communication has become an urgent educational priority.

List of abbreviations

AI: Artificial intelligence

OMIM: Online Mendelian Inheritance in Man

GPA: Grade-point average

AFP: Alpha-fetoprotein

hCG: Human chorionic gonadotropin

μE3: Unconjugated estriol (a placental estrogen marker)

NIPT: Non-invasive prenatal testing

Patient or Public Involvement

This study is a medical education research involving medical students as participants. No patients or members of the public were directly involved in the design, conduct, or reporting of this study.

Declarations

Ethics approval and consent to participate

This study was reviewed and approved by the Medical Ethics Committee of the Affiliated Hospital of Shaoxing University (Approval No. 2026(shen)-141-01). All participants provided written informed consent. The work was carried out in accordance with the Declaration of Helsinki, including, but not limited to, the anonymity of participants being guaranteed and the informed consent of participants being obtained.

Consent for publication

Not applicable. No individual person's data are presented in any identifiable form.

Availability of data and materials

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National First-Class Undergraduate Course (No. 2025-765) and the Shaoxing University 2025 Educational and Teaching Reform Research Project (No. 2025-15) awarded to JX (Jin Xin). Additional support was provided by the Zhejiang Provincial Teaching Reform Project (No. JGCG2025282) awarded to FXL (Fang Xingliang). The funders had no role in the study design, data collection, analysis, interpretation, or manuscript preparation.

Authors' contributions

JX (Jin Xin) conceptualized and designed the study, developed the curriculum, delivered the in-class teaching sessions, wrote the original draft, and was responsible for the final revision and submission of the manuscript. ZY (Zeng Yan) developed and provided the clinical karyotype database, contributed to the teaching delivery, and supervised the clinical data interpretation. JJY (Jin Jieyuan) and JWL (Jin Wenlong) assisted in teaching delivery and data collection. FXL (Fang Xingliang) provided

administrative support, supervised the project, and contributed to funding acquisition. All authors reviewed and approved the final manuscript.

Acknowledgements

The authors thank all participating students and the faculty who supported the curriculum reform and data collection.

References

1. Cooper N, Bartlett M, Gay S, Hammond A, Lillicrap M, Matthan J, Singh M: Consensus statement on the content of clinical reasoning curricula in undergraduate medical education. Medical Teacher 2021, 43(2):152-159.
2. Wagner FL, Sudacka M, Kononowicz AA, Elvén M, Durning SJ, Hege I, Huwendiek S: Current status and ongoing needs for the teaching and assessment of clinical reasoning - an international mixed-methods study from the students` and teachers` perspective. BMC medical education 2024, 24(1):622.
3. Cook DA, Sherbino J, Durning SJ: Management Reasoning: Beyond the Diagnosis. JAMA 2018, 319(22):2267-2268.
4. Norman G, Pelaccia T, Wyer P, Sherbino J: Dual process models of clinical reasoning: The central role of knowledge in diagnostic expertise. Journal of Evaluation in Clinical Practice 2024, 30(5):788-796.
5. Massingham LJ, Nuñez S, Bernstein JA, Gardner DP, Parikh AS, Strovel ET, Quintero-Rivera F: 2022 Association of Professors of Human and Medical Genetics (APHMG) consensus-based update of the core competencies for undergraduate medical education in genetics and genomics. Genetics in

Medicine 2022, 24(12):2400-2413.

6. Menon B, Ranabothu MR, Tarter KE, Layson TDO, Mukundan D, Laukka JJ: Reinforcement of core foundational science concepts during clinical rotations: a strategy to enhance medical student learning and diagnostic reasoning. Advances in Physiology Education 2024, 49(1):47-52.

7. Liang S, Fan F, Hu Y, Zhao J, Gao C, Wang L, Cao X, Feng L, Qin H: Integrating cases, enhancing interaction, and promoting learning through assessment: teaching reform and practical outcomes of medical genetics under the OBE concept. Hereditas(Beijing) 2026, 48(5):594-606.

8. Eva K: Medical Education and artificial intelligence: Responsible and effective practice requires human oversight. Medical Education 2024, 58(11):1260-1261.

9. Anderson H, Studer A, Holm K, Suzuki A: A Case-Based Active Learning Session for Medical Genetics Resources. MedEdPORTAL 2021, 17:11135.

10. Weiler T, Landa-Galindez A: Online interactive genetics education during internal medicine clinical clerkship. Genetics in Medicine 2022, 24(6):1362-1371.

11. Liang L, Yan X-C, Cao X-L, He L, Zheng M-H, Qin H-Y, Han H: A student experience-based teaching to improve the understanding of genotype-phenotype relationship in classroom teaching of medical genetics. Journal of Biological Education 2023, 57(3):555-565.

12. Guest G, Namey EE, Mitchell ML: Collecting Qualitative Data: A Field Manual for Applied Research. In. 55 City Road, London: SAGE Publications, Ltd; 2013.

13. Vygotsky LS: Mind in society: The development of higher psychological processes: Harvard University Press; 1978.
14. Braun V, Clarke V: Using thematic analysis in psychology. Qualitative Research in Psychology 2006, 3(2):77-101.
15. Si J: Using cognitive load theory to tailor clinical reasoning training for preclinical medical students. BMC medical education 2024, 24(1):1185.
16. McConnell MM, Eva KW: The Role of Emotion in the Learning and Transfer of Clinical Skills and Knowledge. Academic Medicine 2012, 87(10):1316-1322.
17. Wen Z, Chen P, Lyu J, Wu J, Qin B, Yang Y, Huang G, Wu C: CBL with integrated SDH versus CBL alone in esophageal surgery internship: a randomized controlled trial on clinical competence development. BMC medical education 2026. [Epub ahead of print]
18. Parsons AS, Wijesekera TP, Rencic JJ: The Management Script: A Practical Tool for Teaching Management Reasoning. Academic Medicine 2020, 95(8):1179-1185.
19. Lavoie P, Lapierre A, Maheu-Cadotte M-A, Fontaine G, Khetir I, Bélisle M: Transfer of Clinical Decision-Making–Related Learning Outcomes Following Simulation-Based Education in Nursing and Medicine: A Scoping Review. Academic Medicine 2022, 97(5):738-746.
20. Roberts JK, Weigle N, Fox JW, Natesan S, Gordon D, Chudgar SM: Validity Evidence for Using Virtual, Interactive Patient Encounters to Teach and Assess Clinical Reasoning for First-Year Medical Students. Academic Medicine 2025,

100(4):445-452.

21. Kulasegaram KM, Martimianakis MA, Mylopoulos M, Whitehead CR, Woods NN: Cognition Before Curriculum: Rethinking the Integration of Basic Science and Clinical Learning. Academic Medicine 2013, 88(10):1578-1585.

22. Albert FA, Seidu A-A, Mason HM, Anderson E, Alele FO, Heggarty P, Hollins A, Gupta TS, Hays RB, McArthur L et al: A systematic review of medical practitioners' retention and application of basic sciences to clinical practice. BMC medical education 2024, 24(1):997.

Table 1 Session 1 group inquiry tasks and core findings

Group	Clinical indication (n)	Core analytical question	Core finding	Cognitive conflict triggered
1	High risk for trisomy 21 (183)	What proportion are actually trisomy 21?	20.8% (38/183) were trisomy 21	Strong
2	NIPT high risk for trisomy 21 (99)	How does detection rate compare with serum screening?	94.9% (94/99) were trisomy 21	Moderate
3	Advanced maternal age (201)	What is the distribution of abnormality types?	Inversions (32.5%) outnumbered trisomy 21 (17.4%)	Strong
4	Trisomy 21 (271)	What are the proportions and clinical implications of mosaic ratios?	53 mosaic cases; ratios 3% to 98%	Moderate
5	NIPT sex chromosome aneuploidy (59)	Which syndromes do karyotypes represent?	Klinefelter (23), XYY (16), XXX (14), Turner (4)	Moderate
6	NT thickening (49)	What abnormalities can NT thickening detect?	12 abnormality types; trisomy 21 (16), Turner (4) most common	Moderate

Cognitive conflict intensity was coded based on: (a) presence of student-initiated diagnostic questioning in group discussions, (b) frequency of reported “surprise” in questionnaires, and (c) evidence of mental model revision in group presentations.

Table 2 Student evaluations and thematic analysis (N = 30)

Table 2 Part A: Quantitative student feedback (N = 30)

Evaluation Item	n (%)
Liked the dual-teacher co-teaching model	28 (93.3)
Reported substantial learning gains	29 (96.7)
Hoped to extend this teaching module	30 (100)

Table 2 Part B: Thematic analysis of open-ended responses

Core Theme	Representative Student Quote	Mention Frequency (%)
Dual-teacher model recognition	"Both teachers brought complex text to life"	15
Authentic database reduces learning barriers	"Real cases pop up immediately"	12
Raw data revises textbook biases	"Only about 20% of high-risk cases actually have trisomy 21"	8
Demand for more data-driven classes	"This self-directed inquiry approach is worth adopting"	7
Shift from answer-seeking to independent inquiry	"We need to ask boldly, guess boldly, verify boldly"	6

Table 3. Comparison of correctness rates on core chromosomal abnormality content areas between cohorts

Content Area	Teaching Focus	2024	2025	Gain
Structural aberration types	Session 1: Foundational	85.0%	97.5%	↑12.5%
Deletion karyotype interpretation	Session 1: Database screening	70.0%	95.0%	↑25%
Mosaic karyotype interpretation	Session 1: Group 4 inquiry	75.0%	90.0%	↑15%
Inversion: normal vs pathogenic	Session 1: Group 3 discovery	70.0%	95.0%	↑25%
Inversion heterozygote meiosis	Session 1: Cognitive conflict	80.0%	95.0%	↑15%
Robertsonian translocation counseling	Session 2: Simulation	60.0%	85.0%	↑25%
Inversion carrier genetic effects	Session 1 → Session 2	40.0%	65.0%	↑25%

Note: Each core content domain was measured via one matched test item with consistent testing focus in both cohorts. Significant uniform improvement across all seven domains was confirmed via binomial test ($p < 0.01$).

Figure legends:

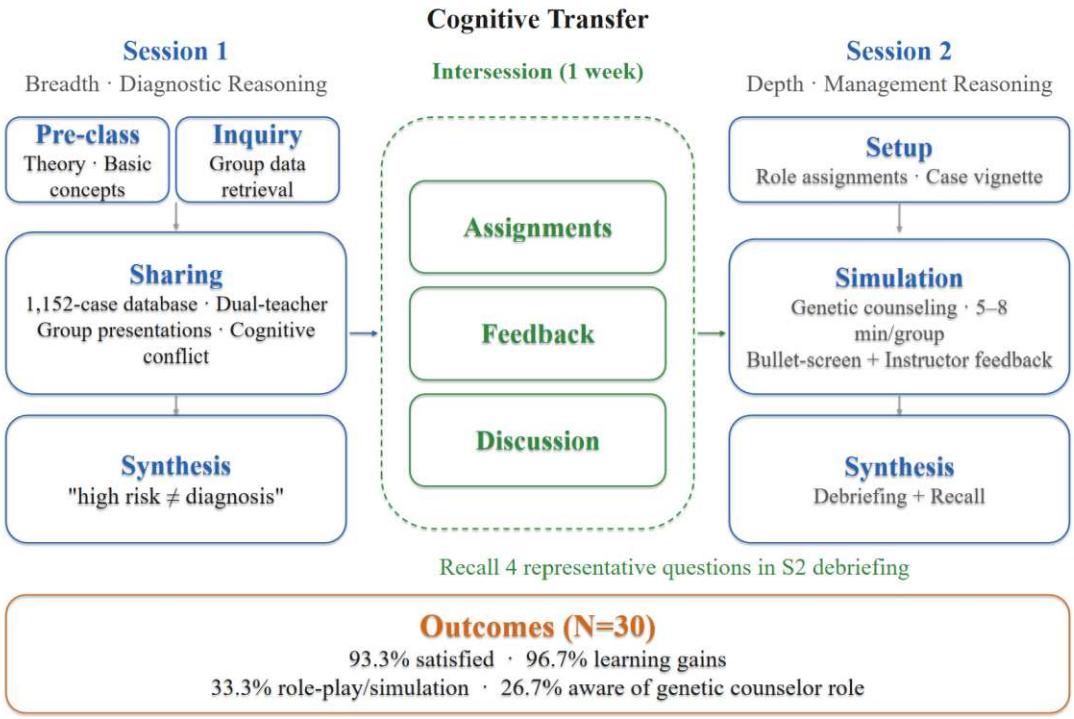


Figure 1. Sequential two-stage teaching framework for pre-clinical medical students (N = 30). Session 1 focuses on database-based diagnostic reasoning; Session 2 delivers simulated genetic counseling for management reasoning. A one-week intersession interval is reserved for literature review and counseling script drafting.

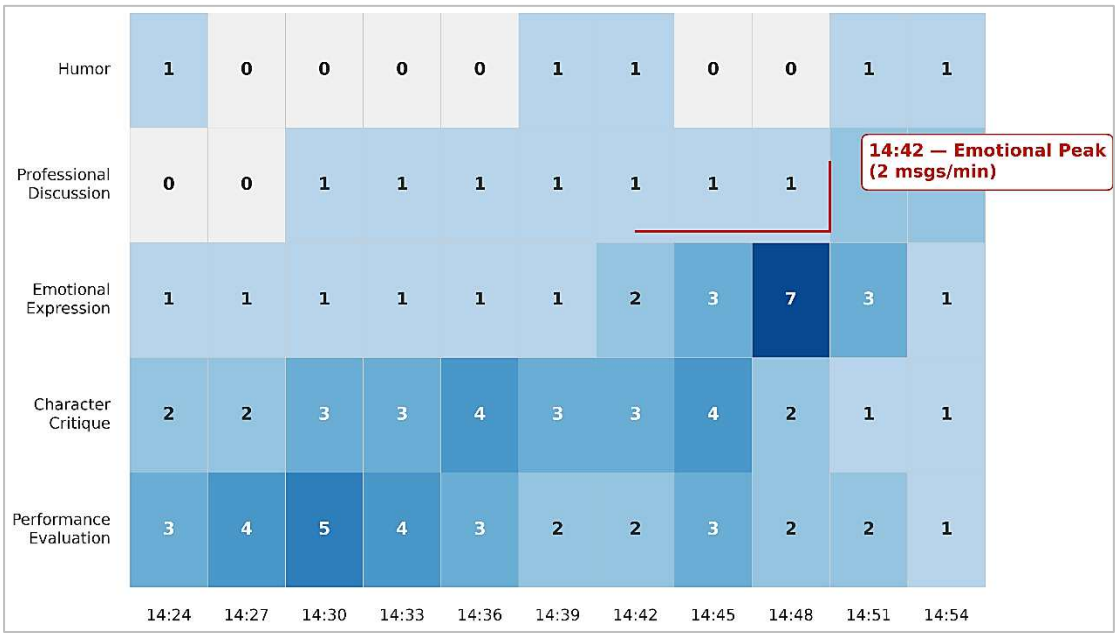


Figure 2. Temporal heatmap of real-time bullet-screen comments, grouped into five thematic categories. Each cell represents comment count within a 3-minute time bin; darker blue indicates higher comment volume. The 14:42 peak corresponds to the most empathetic patient counseling role-play segment.

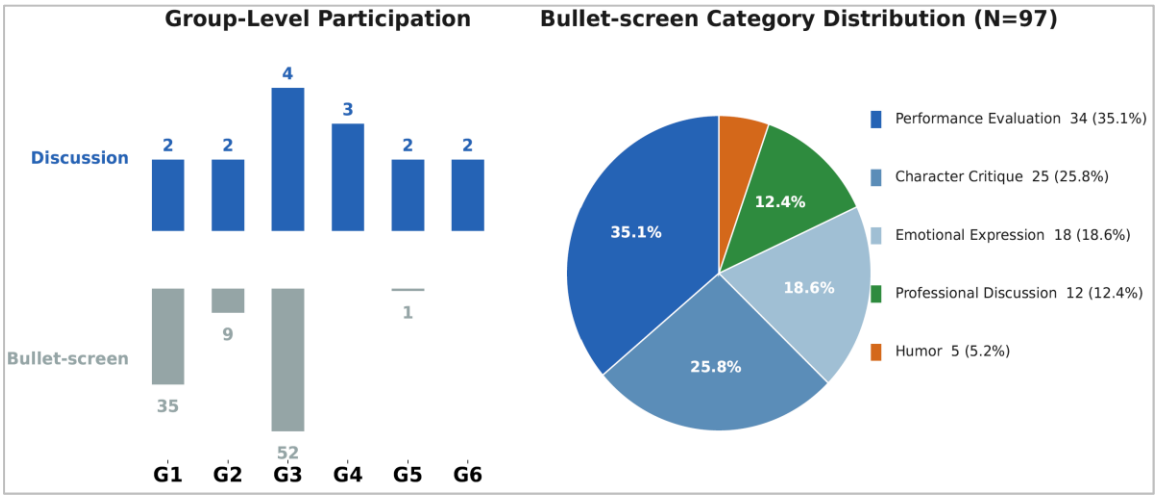


Figure 3. Group-level learning engagement comparison. Left: asynchronous discussion board posts vs synchronous bullet-screen messages across six student groups; Right: proportional distribution of all 97 bullet-screen comments by thematic category. Groups 4 and 6 submitted zero bullet-screen comments.