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.

Video-Based Patient Simulation Enhances Clinical Reasoning, Knowledge Retention, and Resource Utilisation in Undergraduate Ophthalmology: A Randomised Controlled Trial

Sibel Zorlu Ozturk

ozturkzorlusibel@gmail.com

Altınbaş University

Haci Omer TONTUS

Ondokuz Mayıs University

Research Article

Keywords: clinical reasoning, medical education, ophthalmology, red eye, randomised controlled trial, simulation, virtual patient, knowledge retention, CONSORT

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10007339/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Clinical reasoning skills are essential for diagnosing and managing common but heterogeneous presentations such as the red eye, where conditions range from benign self-limiting disease to vision-threatening pathology. Traditional lecture-based instruction provides limited opportunity for active reasoning. Screen-based patient simulations incorporating authentic patient video material may bridge this gap, but evidence in undergraduate ophthalmology education remains limited.

Methods

In this single-centre, parallel-group randomised controlled educational trial, 100 fifth-year medical students were randomised 1:1 to a screen-based simulation arm or an instructor-led interactive lecture arm covering the evaluation and management of red eye. The simulation incorporated real patient video material and required structured engagement across history-taking, physical examination, differential diagnosis, consultation decision-making, final diagnosis and treatment planning. The primary outcome was overall clinical performance assessed by a pre-specified six-domain weighted rubric (range 0–100) scored independently by two blinded evaluators. Secondary outcomes included step-based clinical performance, multiple-choice knowledge tests at four time points (pre-test, immediately post-intervention, 30 and 60 days), appropriateness of diagnostic resource utilisation, and a five-domain satisfaction questionnaire scored on a 5-point Likert scale. Analyses followed the intention-to-treat principle. Reporting adhered to CONSORT 2025.

Results

All 100 randomised participants (50 simulation; 50 lecture) completed the study with no attrition. Baseline knowledge scores were comparable ($p = 0.64$). The simulation group achieved markedly higher total clinical performance scores (84.6 ± 6.8 vs 69.2 ± 9.5; mean difference 15.4, 95% CI 12.1–18.7; $p < 0.001$; Cohen's $d = 1.85$). Step-based performance favoured the simulation group across every domain (d range 1.23–1.88). Knowledge advantages persisted at the immediate post-test ($d = 1.55$), 30-day, and 60-day retention assessments (all $p < 0.001$). The simulation group also requested fewer unnecessary laboratory tests (12% vs 38%, $p = 0.003$) and imaging studies (6% vs 24%, $p = 0.01$). Student satisfaction, engagement, perceived realism, and confidence were all significantly higher in the simulation group.

Conclusions

Video-based patient simulation produced large, durable improvements in clinical reasoning, knowledge retention and diagnostic resource utilisation compared with traditional lecture-based instruction in

undergraduate ophthalmology education. The magnitude of the observed effect should be interpreted alongside potential ceiling effects, intervention–outcome alignment, and novelty considerations discussed in Limitations. The findings support integration of structured, video-based simulation into clinical curricula for high-frequency, high-variability presentations such as the red eye.

Trial registration

ClinicalTrials.gov, NCT07679672 (https//clinicaltrials.gov/study/NCT07679672) date of registration 01 July 2026.

Background

Clinical reasoning is the cognitive process by which clinicians integrate information, generate hypotheses, weigh diagnostic possibilities and reach decisions. It is the central skill that converts factual knowledge into safe, accurate clinical practice and is foundational to undergraduate medical education. [1, 2] Contemporary cognitive accounts frame clinical reasoning within dual-process theory, in which fast pattern recognition (System 1) and slower analytical reasoning (System 2) interact to produce diagnoses and decisions, and in which both systems contribute to diagnostic accuracy and to error; [3–5] reflective practice can attenuate System-1 errors when learners are prompted to reconsider initial impressions. [6] The acquisition of clinical reasoning is a long developmental process and is widely regarded as one of the most challenging skills to teach during medical school, particularly for high-frequency presentations whose differential diagnosis spans benign and serious conditions.

The red eye is a paradigmatic example. It is among the most frequent ophthalmologic complaints encountered in primary care and emergency settings, yet encompasses conditions ranging from self-limiting conjunctivitis to vision-threatening keratitis, anterior uveitis and acute angle-closure glaucoma. Accurate evaluation requires not only factual recognition of clinical findings but also the integrated synthesis of history, examination signs, and risk-stratified prioritisation of management. Eye-related complaints account for an estimated 2–3% of consultations in primary care and emergency settings, with red eye the predominant presenting feature. [7, 8] Yet undergraduate ophthalmology instruction has been progressively compressed in many curricula, leaving graduates with limited confidence in the assessment and triage of acute eye conditions. [9] The mismatch between this cognitive demand and the limited reasoning practice provided by traditional lecture-based teaching is well recognised in the medical-education literature.

Simulation-based education has emerged as an effective response. By providing structured, controlled environments in which learners can engage actively in clinical decision-making, simulation supports deliberate practice [11] and feedback in ways that lectures cannot, with meta-analytic evidence indicating superiority over traditional clinical education when simulation is paired with deliberate practice. [17] Beyond deliberate practice specifically, simulation-based health-professions education is supported by a substantial cumulative evidence base. [12, 13] Systematic reviews have shown that

technology-enhanced simulation is associated with improvements in knowledge, skills and clinical behaviours across health professions education, [13, 14], and best-practice frameworks have summarised the design features that support effective learning. [15, 16] Cognitive load theory provides a complementary framework, suggesting that simulation environments can be deliberately designed to manage extraneous load and support germane processing during reasoning practice. [10]

Among simulation modalities, screen-based virtual-patient platforms have attracted increasing attention because of their scalability, reproducibility, and capacity to deliver standardised case content to large cohorts. Meta-analytic evidence supports their effectiveness for clinical reasoning outcomes, [18–20] and recent reviews have positioned them as a core component of digital and simulation-based medical education. [21] In ophthalmology specifically, randomised evidence has shown that interactive web-based virtual ophthalmology clinics outperform traditional teaching for medical-student learning, [24, 25] providing a precedent for screen-based simulation in this discipline.

However, many existing platforms rely on text-based vignettes or static images that cannot capture the dynamic visual and behavioural cues central to ophthalmology — conjunctival injection patterns, discharge characteristics, photophobia, and subtle corneal findings. Design frameworks for virtual patients emphasise that narrative fidelity and clinical realism are critical for transfer of learning to authentic practice. [22] Authentic patient video material may therefore enhance the educational utility of screen-based simulation in visually-driven specialties.

Despite this rationale, evidence specifically evaluating video-based simulation in undergraduate ophthalmology is limited. Most prior simulation studies focus on immediate post-intervention performance with limited follow-up, leaving the durability of any educational gains uncertain. [23] Few studies have examined the impact of simulation-based teaching on the appropriateness of diagnostic resource utilisation, an outcome of growing importance in value-based care.

Against this background, the present randomised controlled educational trial evaluated whether a screen-based, real patient video–supported simulation improves clinical reasoning, knowledge acquisition, knowledge retention at 30 and 60 days, and the appropriateness of diagnostic decision-making in undergraduate ophthalmology education, compared with an interactive instructor-led lecture covering identical content. By analysing both overall performance and step-based components of clinical reasoning, the study sought to provide a comprehensive, mechanism-informed assessment of the educational impact of video-based simulation in a high-frequency, high-variability clinical scenario.

Materials and Methods

This study was designed and reported in accordance with the CONSORT statement for parallel-group randomised trials, [27] updated in 2025. [26] A completed CONSORT 2025 checklist is provided as Additional file 1. The protocol and primary and secondary outcomes were prespecified and locked prior to participant enrolment. Ethical approval was obtained from the Istinye University Human Research Ethics Committee (Protocol No. 2025/147; Meeting No. 2025/04) on 15 May 2025. All participants

provided written informed consent prior to enrolment, including consent for educational data collection and use of anonymised aggregate data for publication.

The trial was registered retrospectively in the ClinicalTrials.gov registry (identifier NCT07679672; https://clinicaltrials.gov/study/NCT07679672; date of registration 01 July 2026). Prospective registration was not undertaken because the study exclusively evaluated an educational intervention involving medical students and did not involve clinical treatment of human patients or patient-level therapeutic outcomes; prospective registration in a public clinical trials registry was therefore not mandated by the institutional ethics framework or local regulatory guidelines. Nonetheless, to ensure methodological transparency and prevent bias, the study protocol, primary endpoints, and secondary outcomes were fully prespecified, documented, and locked prior to participant enrolment, and the final analysis plan was strictly finalised before data collection.

Study design and participants

This single-centre parallel-group randomised controlled educational trial was conducted within the ophthalmology clerkship of a single Turkish university medical school during the second half of 2025. Eligible participants were fifth-year medical students enrolled in the clerkship rotation; there were no exclusion criteria beyond non-enrolment in the relevant rotation. A total of 100 students were enrolled, providing written informed consent, and randomised 1:1 to either the simulation-based intervention (n = 50) or the active-comparator interactive lecture (n = 50). Both groups received structured, equivalent total instructional time on the same clinical content. Participation was voluntary and did not affect academic standing.

Sample size

The trial was designed to detect a between-group difference of clinically meaningful magnitude in overall clinical performance score. Drawing on prior simulation-based education studies in clinical reasoning, which have repeatedly reported large effects (Cohen's d ≥ 0.8) for active screen-based simulation versus lecture instruction, [13, 18] we anchored the calculation to d = 0.8. With a two-sided α of 0.05 and 80% power, an independent two-sample t-test requires 26 participants per group; for a more conservative effect size of d = 0.6 (medium-to-large), the requirement is 45 per group. The achieved enrolment of 50 participants per group (total n = 100) therefore exceeds the requirement under the conservative effect-size assumption and provides additional protection against attrition. No participants were lost to follow-up, so the analytic sample equalled the randomised sample. The total sample size was further constrained by the size of the available fifth-year cohort. In keeping with standard practice for educational trials at a single institution, no interim analyses were planned.

Randomisation and allocation concealment

Participants were randomly assigned 1:1 using a computer-generated random allocation sequence with permuted blocks (block size not disclosed to investigators conducting allocation, to preserve allocation concealment) to ensure approximately equal group sizes throughout enrolment, in line with established methodological guidance. [28, 29] The randomisation sequence was generated by an investigator not involved in the intervention or outcome assessment, using R version 4.3.0 (R Foundation for Statistical Computing, Vienna, Austria). No stratification factors were used because the participant pool was a single fifth-year cohort attending the same institution. Allocation concealment was maintained through sequentially numbered, opaque, sealed envelopes (SNOSE) prepared by the same investigator who generated the sequence; envelopes were opened only at the start of the intervention by an administrative coordinator independent of the research team. Although participants and instructors were aware of being part of an educational research project (consented voluntarily), they were blinded to which specific session would constitute the study intervention; instructors delivering the comparator session were similarly blinded to whether their session was being evaluated, to prevent ad-hoc modification of routine teaching. Outcome assessors were blinded to allocation (see Blinding).

Educational interventions

Educational content was standardised across both arms to ensure comparability. The clinical scenario, Management and Differential Diagnosis of Red Eye, was selected from the curricular content of the Turkish National Core Curriculum for Undergraduate Medical Education (UÇEP 2020), [39] which lists “red eye (ocular redness)” and “dry eye” among the core symptoms whose evaluation falls within the disciplines of family medicine, ophthalmology and emergency medicine. The scenario presented a 27-year-old female office worker (accountant; height 163 cm, weight 52 kg) attending an outpatient clinic with red eye, with a UÇEP-aligned differential diagnosis required across allergic reaction, glaucoma, keratitis, conjunctivitis, urticaria/angioedema and uveitis; the final diagnosis was conjunctivitis. The scenario was scripted from the clinical records of a consenting outpatient and used a written-consent–authorised authentic patient video recorded in a university-provided clinical environment. It was constructed to require structured progression through history-taking, focused physical examination, differential diagnosis, decision on specialist consultation, final diagnosis, and treatment and follow-up planning. Both arms received identical clinical content delivered in a single 50-minute teaching session matched in total instructional time; each participant completed the relevant intervention once during the study period, with no repeat attempts permitted.

Participants in the simulation group interacted with the scenario through the digital simulation platform, which integrates authentic patient video material with structured decision points across the seven reasoning steps listed above. The platform required active engagement at each step before advancing to the next, providing immediate structured feedback at the conclusion of the scenario. Participants in the active-comparator lecture group received the same clinical content through an instructor-led interactive lecture delivered by an experienced ophthalmologist (12 years of clinical experience) covering the evaluation and management of red eye. The lecture incorporated patient photographs, published clinical

vignettes, and exemplar cases drawn from textbooks and peer-reviewed series, ensuring that case-based and visual stimuli were present in both arms. Questioning and discussion were encouraged throughout to maintain engagement and parity in active learning time. The principal between-arm difference therefore lay in the mode of engagement — active learner-led decision-making with dynamic patient video in the simulation arm versus expert-led explanation supported by static visual material in the lecture arm — rather than the absence of clinical material in either arm.

Outcomes

Primary outcome — overall clinical performance

The primary outcome was the overall clinical performance score derived from the structured red eye scenario, scored using a pre-specified six-domain weighted rubric: anamnesis (25%), physical examination (20%), preliminary diagnosis (15%), consultation decision (5%), final diagnosis (15%) and medical treatment (20%). Total scores ranged from 0 to 100, with higher scores indicating better performance. The rubric was developed by the investigators on the basis of established frameworks for evaluating clinical reasoning steps, [1, 2] piloted on a small group of senior medical students prior to the trial, and refined for clarity. Content validity was reviewed by two ophthalmology faculty members not involved in the trial. Each performance was scored independently by two trained evaluators blinded to group allocation. Inter-rater agreement was assessed using the intraclass correlation coefficient (ICC; two-way random-effects model, absolute agreement, single rater). [30] Discrepancies of more than 5 points on any domain were resolved by consensus between the two evaluators with reference to a written scoring guide.

Secondary outcomes — step-based reasoning, knowledge and resource utilisation

Step-specific performance scores were captured for each reasoning component (anamnesis, physical examination, preliminary diagnosis, consultation, final diagnosis, treatment, and follow-up planning) using the same rubric. Knowledge outcomes were evaluated using a structured 25-item multiple-choice examination aligned with the explicit learning objectives of the red eye scenario; each item had four options with a single correct answer, scored 0–100 (4 points per item). The blueprint was developed by the investigators in collaboration with two ophthalmology faculty using a content-by-objective grid covering history features, examination findings, differential diagnosis, investigation, management, and follow-up; content validity was reviewed by these faculty before use. The blueprint was constructed in accordance with established guidance on content-valid examination development. [31] Internal consistency of the 25-item form was assessed using Cronbach's α, interpreted following standard recommendations. [32] The same blueprint and item bank were used at all four assessment time points (immediately before the intervention as a pre-test; immediately after; at 30 days; and at 60 days), with

item order randomised between administrations to limit memorisation effects. Higher scores indicated better knowledge performance.

Diagnostic resource utilisation was captured as an exploratory but pre-specified secondary outcome. Within the structured scenario, learners were able to request laboratory and imaging investigations from a fixed menu; investigations were classified a priori as appropriate or unnecessary by the two ophthalmology faculty using a consensus-derived list of indications for the specific differential diagnoses presented in the scenario. The proportion of participants requesting at least one unnecessary laboratory test and at least one unnecessary imaging study was compared between groups.

Student satisfaction was assessed using an investigator-developed 25-item questionnaire grouped into five conceptual domains (Satisfaction, Engagement, Realism, Educational Value, and Confidence; five items per domain). Each item was scored on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). A summary score for each domain was calculated as the median of its five constituent items, giving five domain scores per participant for analysis (Table 5). Items in the Confidence domain asked participants to assess their perceived change relative to their pre-session state (e.g., “I feel more confident in evaluating patients with red eye after this session”), thereby capturing within-individual change rather than absolute confidence; the corresponding between-arm comparison should accordingly be read as a difference in self-perceived gain rather than in absolute capability. Items were drafted in English in accordance with published guidance on developing questionnaires for educational research and were reviewed for face validity by two medical-education faculty external to the trial. [33] The questionnaire is reported as an investigator-developed instrument and not as a previously validated scale. Likert-scale data are presented as median (interquartile range, IQR), consistent with recommended practice for ordinal Likert items. [34]

Blinding

Three layers of blinding were implemented to mitigate performance and detection bias. First, although both groups had consented to participate in an educational research project, neither participants nor the instructor delivering the comparator session were informed which specific teaching session within the broader curriculum would constitute the study intervention; this preserved the routine character of the comparator session and prevented ad-hoc enrichment of teaching content. Second, the pre-test, post-test and retention assessments were administered as part of the routine institutional curriculum-evaluation procedure already familiar to students, further reducing test-taking salience that might otherwise drive a Hawthorne-type response. Third, all outcome scoring was performed under standard blinded conditions: the two evaluators scoring clinical performance from de-identified scenario records, the marker scoring the multiple-choice examinations, and the analyst conducting the primary statistical comparison were all blinded to group allocation. Knowledge tests and rubric forms were de-identified by an administrative coordinator independent of the research team prior to scoring.

Statistical analysis

All analyses were conducted using SPSS Statistics version 28.0 (IBM Corp., Armonk, NY) and followed the intention-to-treat principle, with all randomised participants analysed in their assigned groups. As there was no attrition, the per-protocol and intention-to-treat populations were identical and no per-protocol analysis is reported separately. Continuous variables are presented as mean ± standard deviation, and categorical variables as frequencies and percentages. Distributional assumptions were assessed with the Shapiro–Wilk test. Between-group comparisons used independent-samples t-tests or Mann–Whitney U tests as appropriate; categorical variables were compared with χ^2 or Fisher's exact tests. Mean differences for continuous outcomes are reported with two-sided 95% confidence intervals derived using Welch's method for unequal variances. Repeated measurements of knowledge scores across the four assessment points were analysed using mixed-design analysis of variance with group (simulation vs lecture) as the between-subjects factor, time as the within-subjects factor, and the group-by-time interaction as the test of differential trajectory; Greenhouse–Geisser correction was applied where appropriate. Effect sizes for between-group differences are reported as Cohen's d. A two-sided p value of less than 0.05 was considered statistically significant. No interim analyses were performed and no formal adjustment for multiplicity across secondary outcomes was applied; secondary findings should therefore be interpreted as supportive of, rather than independently confirming, the primary result.

Results

Participant flow and characteristics

All 100 enrolled fifth-year medical students were randomised (50 to simulation, 50 to lecture) and all 100 (100%) completed the entire study, including all four knowledge assessment time points (Fig. 1). There was no attrition, and no participant crossed over between arms. Baseline characteristics were balanced between groups (Table 1), with comparable age (23.8 ± 1.2 vs 23.6 ± 1.3 years, p = 0.42), sex distribution (52% vs 50% female, p = 0.84), and pre-test knowledge scores (46.8 ± 9.2 vs 45.9 ± 8.7, p = 0.64). Inter-rater reliability for the clinical performance rubric was excellent (ICC = 0.91). Internal consistency of the 25-item knowledge test was acceptable (Cronbach's α = 0.78).

Table 1
Baseline characteristics of randomised participants.

Variable	Simulation (n = 50)	Lecture (n = 50)	p
Age, years (mean ± SD)	23.8 ± 1.2	23.6 ± 1.3	0.42
Female sex, n (%)	26 (52)	25 (50)	0.84
Pre-test knowledge score, mean ± SD	46.8 ± 9.2	45.9 ± 8.7	0.64

Continuous variables compared with the independent-samples t-test; categorical variables compared with the χ^2 test.

Primary outcome — overall clinical performance

Students in the simulation group achieved markedly higher overall clinical performance scores than those in the lecture group (84.6 ± 6.8 vs 69.2 ± 9.5; mean difference 15.4 points on the 0–100 scale, 95% CI 12.1–18.7; $p < 0.001$; Cohen's $d = 1.85$). The separation between distributions was substantial, with limited overlap, indicating a difference unlikely to reflect marginal variation around a similar central tendency. Step-based performance favoured the simulation group across every domain, with effect sizes ranging from $d = 1.23$ (consultation decision) to $d = 1.88$ (final diagnosis) (Table 2). The pattern — favouring the simulation group at every reasoning step rather than only at terminal decision points — is consistent with an intervention effect on integrated reasoning rather than isolated knowledge recall, although alternative explanations are addressed in Limitations.

Table 2
Overall and step-based clinical performance by group.

Domain	Simulation (mean ± SD)	Lecture (mean ± SD)	Mean diff	p	Cohen's d
Total score (primary)	**84.6 ± 6.8**	**69.2 ± 9.5**	**15.4 (12.1–18.7)**	**< 0.001**	**1.85**
Anamnesis	88.2 ± 7.5	72.4 ± 10.1	15.8 (12.3–19.3)	< 0.001	1.77
Physical examination	85.6 ± 8.1	70.3 ± 9.8	15.3 (11.7–18.9)	< 0.001	1.68
Preliminary diagnosis	82.1 ± 9.4	66.7 ± 11.2	15.4 (11.3–19.5)	< 0.001	1.50
Consultation decision	79.4 ± 10.2	65.5 ± 12.3	13.9 (9.4–18.4)	< 0.001	1.23
Final diagnosis	87.9 ± 7.2	71.8 ± 9.6	16.1 (12.7–19.5)	< 0.001	1.88
Treatment	83.5 ± 8.6	67.1 ± 10.4	16.4 (12.6–20.2)	< 0.001	1.71
Follow-up planning	81.2 ± 9.1	66.8 ± 11.0	14.4 (10.4–18.4)	< 0.001	1.43

Mean differences are presented with two-sided 95% confidence intervals derived using Welch's method. Total clinical performance score is the primary outcome; remaining rows are pre-specified secondary outcomes.

Knowledge acquisition and retention

Knowledge scores diverged immediately after the intervention and the gap persisted across follow-up assessments (Table 3). Mixed-design ANOVA showed a significant group × time interaction ($p < 0.001$), indicating that the trajectory of knowledge scores differed between arms. Both groups exhibited a

modest decline from immediate post-test to the 60-day assessment, which is the expected forgetting curve in educational interventions; the decline was, however, less pronounced in the simulation group, supporting more durable consolidation of learning.

Table 3
Knowledge scores across four assessment time points.

Time point	Simulation (mean ± SD)	Lecture (mean ± SD)	Mean diff (95% CI)	p	Cohen's d
Pre-test	46.8 ± 9.2	45.9 ± 8.7	0.9 (− 2.6 to 4.4)	0.64	0.10
Immediate post-test	**83.1 ± 7.6**	**69.5 ± 9.3**	**13.6 (10.3–16.9)**	**< 0.001**	**1.55**
30-day retention	79.8 ± 8.4	66.2 ± 10.1	13.6 (9.9–17.3)	< 0.001	1.46
60-day retention	76.9 ± 9.1	63.5 ± 10.8	13.4 (9.4–17.4)	< 0.001	1.34

Mean differences with two-sided 95% confidence intervals (Welch's method). Group × time interaction p < 0.001 (mixed-design ANOVA).

Diagnostic resource utilisation

Within the structured scenario, the simulation group requested fewer a priori–defined unnecessary investigations than the lecture group: 12% vs 38% for laboratory tests (p = 0.003) and 6% vs 24% for imaging studies (p = 0.01) (Table 4). This finding suggests that the educational benefit of structured video-based simulation extends beyond diagnostic accuracy to the rational use of investigations, although the small absolute number of events warrants cautious interpretation.

Table 4
Unnecessary diagnostic resource utilisation.

Outcome	Simulation, n (%)	Lecture, n (%)	p
≥ 1 unnecessary laboratory test	6 (12)	19 (38)	0.003
≥ 1 unnecessary imaging study	3 (6)	12 (24)	0.01

Compared with χ^2 test. 'Unnecessary' classification was a priori, derived by consensus of two ophthalmology faculty for the differential diagnoses presented in the scenario.

Student satisfaction

Student-reported outcomes strongly favoured the simulation arm across all five domains (Table 5). The largest difference was in perceived realism, consistent with the use of authentic patient video material. Confidence in clinical decision-making, engagement, and overall satisfaction were also significantly

higher in the simulation group. These data are reported as median (IQR) on a 5-point Likert scale and were analysed using the Mann–Whitney U test.

Table 5
Student-reported outcomes by domain (median of five items per domain, 5-point Likert scale).

Domain	Simulation, median (IQR)	Lecture, median (IQR)	p
Overall satisfaction	5 (4–5)	4 (3–4)	< 0.001
Engagement	5 (4–5)	3 (3–4)	< 0.001
Perceived realism	5 (5–5)	3 (2–4)	< 0.001
Educational value	5 (4–5)	4 (3–4)	< 0.001
Confidence in decision-making	5 (4–5)	4 (3–4)	< 0.001

Each domain comprises five items (25 items total); domain scores reported as median (IQR) of constituent items. Between-group comparisons performed with the Mann–Whitney U test.

Discussion

In this single-centre, parallel-group randomised controlled educational trial of fifth-year medical students, video-based patient simulation produced markedly higher overall clinical performance scores than active-comparator interactive lecture instruction (mean difference 15.4 points; 95% CI 12.1–18.7; d = 1.85). The advantage extended consistently across every step of the clinical reasoning pathway and to knowledge acquisition and 60-day retention, and was accompanied by lower utilisation of unnecessary diagnostic investigations and higher student-reported satisfaction. Although the absolute magnitude of the observed effect is unusually large for an educational trial and warrants careful scrutiny, the pattern of findings — internal consistency across reasoning steps, persistence of advantage at 60 days, and concordant improvement in resource utilisation — supports a meaningful effect of the intervention on integrated clinical reasoning rather than a narrow gain limited to test-aligned recall.

The red eye is well suited to this kind of evaluation. It is among the most common ophthalmologic complaints encountered in primary care and emergency settings, yet differentiation between benign and vision-threatening causes depends on subtle visual cues — the pattern of conjunctival injection, the character of discharge, photophobia, and corneal findings — that are difficult to convey through text or static images. The integration of authentic patient video into a structured, decision-rich scenario is therefore plausibly aligned with the cognitive demands of this presentation. Our results are consistent with prior evidence that active engagement in simulated clinical environments improves diagnostic accuracy and reasoning performance, [13, 19] but extend that evidence by demonstrating that benefits propagate through every step of the reasoning pathway, including earlier processes such as anamnesis and physical examination — a finding compatible with theoretical accounts of clinical reasoning as a sequence of interdependent cognitive operations. [1, 2]

The persistence of the simulation group's advantage at the 30- and 60-day knowledge retention assessments is, in our view, the most clinically meaningful finding. Short-term post-intervention gains in educational trials are common but do not necessarily translate into durable clinical competence, and the basic-science forgetting curve has been well characterised across health-professions training. [36] Prior simulation literature has acknowledged that long-term retention evidence is more limited than immediate-effect evidence. [13, 14, 23] The mixed-design analysis showed a significant group × time interaction ($p < 0.001$), with both arms exhibiting the expected forgetting curve but the simulation arm declining less steeply. This pattern is consistent with experiential and elaborative-encoding accounts of learning, in which active engagement with clinically realistic material supports deeper consolidation that resists the typical decay trajectory of lecture-acquired knowledge. [23]

The reduction in unnecessary laboratory and imaging requests in the simulation group adds an instructive dimension to the results. In contemporary practice, accurate diagnosis is necessary but not sufficient: rational use of investigations is increasingly recognised as a competency in its own right and is associated with the depth of structured training received. [37] Active engagement with a structured scenario in which investigation requests must be justified appears to influence not only diagnostic accuracy but also the appropriateness of resource utilisation. This finding is exploratory and based on a small absolute number of events, and the classification of investigations as a priori appropriate or unnecessary inevitably involves judgement; nevertheless, the consistency of the direction of effect across both laboratory and imaging outcomes is supportive.

Interpreting the magnitude of effect

The consistent direction and large magnitude of the differences observed across all primary and secondary outcomes invite careful interpretation. Several methodological factors plausibly contribute to the observed effect size beyond the intervention itself, and we consider these in turn before assessing the residual evidence for a true educational benefit.

A particular point of interpretation concerns perceived realism, which showed the largest single-domain difference between arms in student-reported outcomes. Within this trial, realism reflects the experiential character of the learning environment rather than the clinical authenticity of the underlying material, which was comparable across arms (authentic patient video in the simulation arm; patient photographs, vignettes and exemplar cases in the lecture arm). The defining pedagogical distinction between the two interventions was that the simulation arm placed learners in an active decision-making position with an unfamiliar clinical case, whereas the lecture arm framed the same content through expert exposition. From a deliberate-practice perspective, [11, 17] the obligation to make sequential clinical decisions under uncertainty is itself a defining feature of an authentic learning environment and is one of the principal mechanisms by which simulation is hypothesised to confer benefit. The observed realism difference is therefore consistent with the intended intervention contrast rather than evidence of a methodological asymmetry favouring the simulation arm.

Several considerations are needed when interpreting the magnitude of effect. First, simulation, as an active learning modality, is intrinsically more aligned with the structured rubric used to score clinical performance than is a lecture, which delivers content didactically. Some intervention–outcome alignment is therefore expected, and may amplify measurable differences relative to what would be observed against a less reasoning-aligned reference. Second, ceiling effects at the upper end of the 0–100 rubric — simulation-group means approached 85 — can compress variance in one arm and inflate standardised differences. Third, although the session-blinding procedure described in Methods (in which neither participants nor the instructor delivering the comparator session knew which specific session constituted the study intervention) substantially reduces classical Hawthorne-type performance enhancement, residual novelty effects associated with first exposure to a new educational modality [38] cannot be excluded in a single-cohort design, although the persistence of advantage at 60 days argues against their being the dominant explanation. Fourth, instructor enthusiasm in the active-comparator lecture arm could not be standardised to the same degree as the digital platform; this is an inherent limitation of comparing a digital intervention with an instructor-delivered comparator. None of these considerations vitiates the qualitative finding that simulation outperformed lecture, but they do mean that the absolute magnitude of effect should be interpreted with the broader cautions that apply to standardised effect size measures generally, [35] and should be replicated in larger, multi-centre cohorts before being treated as a reference value for power calculations.

Limitations

The study was conducted at a single Turkish university medical school and focused on a single clinical scenario (red eye) within ophthalmology; generalisability to other institutions, curricula, and clinical domains will require replication. Because the educational nature of the intervention precludes blinding of participants and instructors, only outcome assessors and the analyst could be masked to allocation; this is intrinsic to educational trials and was mitigated through the procedural safeguards described in Methods, but cannot be fully eliminated. The clinical performance rubric and the multiple-choice knowledge assessment, although developed with content validity review and acceptable internal consistency ($\alpha = 0.78$) and inter-rater reliability (ICC = 0.91), were investigator-developed instruments rather than previously validated published scales; this is a recognised limitation of bespoke educational measurement and is reported transparently. The student satisfaction questionnaire is similarly investigator-developed and is reported as such. Knowledge retention was assessed up to 60 days; longer-term retention beyond a single semester remains uncertain. The effect size achieved ($d = 1.85$) considerably exceeded the $d = 0.8$ anchor used to derive the conservative power estimate, meaning the trial was likely over-powered for its primary outcome; this does not invalidate the findings but indicates that more conservative effect-size assumptions should guide future replication studies. Finally, no adjustment for multiplicity was made across the secondary outcomes; secondary findings should therefore be interpreted as supportive of, rather than independently confirming, the primary result.

Conclusion

Among 100 fifth-year medical students randomised to a video-based simulation intervention or to active-comparator interactive lecture instruction on the evaluation and management of red eye, simulation produced large and durable improvements in clinical reasoning, knowledge retention to 60 days, and the appropriateness of diagnostic resource utilisation. The findings support integration of structured, authentic-video–supported simulation into undergraduate ophthalmology curricula for high-frequency, high-variability presentations, while indicating that the magnitude of effect should be confirmed in larger, multi-centre cohorts using more conservative power assumptions.

Declarations

Ethics approval and consent to participate

This study was approved by the **Istinye University Human Research Ethics Committee (T.C. İstinye Üniversitesi İnsan Araştırmaları Etik Kurulu)** on **15 May 2025** (Protocol No. **2025/147;** Meeting No. **2025/04)**. All participants provided written informed consent prior to enrolment. The study was conducted in accordance with the Declaration of Helsinki.

Consent for publication

The patient who appears in the scenario video material provided written informed consent for the recording and educational and academic use of the recordings, including publication of associated educational research.

Availability of data and materials

The de-identified datasets generated and analysed during the current study are available from the corresponding author on reasonable request, subject to compliance with institutional data protection policies. The clinical performance rubric, knowledge-test blueprint, and satisfaction questionnaire used in this study are available from the corresponding author on reasonable request.

Competing interests

The digital simulation platform used in this study was co-developed by the second author (H.O.T.), who also serves as its medical director. To minimise the potential for bias, the following safeguards were implemented: (i) the second author had no direct contact with participants during the intervention or outcome assessment phase; (ii) all outcome scoring (clinical performance rubric and multiple-choice knowledge tests) was performed by trained evaluators who were not involved in platform development and who remained blinded to group allocation; (iii) the analysis plan and primary endpoint were prespecified and locked prior to data collection; and (iv) the platform is non-commercial and not currently marketed, with no licensing revenue, royalties, or commercial partnerships associated with its use in this study. The first author and corresponding author (S.Z.Ö.) declares no competing interests.

Funding

No external funding was received for this study. The institution provided the clinical environment for the recording of patient video material and the digital infrastructure necessary for the simulation intervention.

Authors' contributions

S.Z.Ö.: Conceptualisation, methodology, investigation, data curation, formal analysis, writing — original draft, project administration. H.O.T.: Conceptualisation, methodology, validation, formal analysis, resources, supervision, writing — review & editing. Both authors read and approved the final manuscript.

Acknowledgements

The authors thank the participating fifth-year medical students for their engagement, the patient who consented to authentic video recording, and the ophthalmology faculty who reviewed the rubric and knowledge-test content for validity.

References

1. Norman G. Research in clinical reasoning: past history and current trends. Med Educ. 2005;39(4):418–27.
2. Eva KW. What every teacher needs to know about clinical reasoning. Med Educ. 2005;39(1):98–106.
3. Croskerry P. Clinical cognition and diagnostic error: applications of a dual process model of reasoning. Adv Health Sci Educ Theory Pract. 2009;14(Suppl 1):27–35.
4. Croskerry P. A universal model of diagnostic reasoning. Acad Med. 2009;84(8):1022–8.
5. Pelaccia T, Tardif J, Triby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. Med Educ Online. 2011;16:5890.
6. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. Med Educ. 2008;42(5):468–75.
7. Frings A, Geerling G, Schargus M. Red eye: a guide for non-specialists. Dtsch Arztebl Int. 2017;114(17):302–12.
8. Cronau H, Kankanala RR, Mauger T. Diagnosis and management of red eye in primary care. Am Fam Physician. 2010;81(2):137–44.
9. Succar T, Grigg J, Beaver HA, Lee AG. A systematic review of best practices in teaching ophthalmology to medical students. Surv Ophthalmol. 2016;61(1):83–94.
10. van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. Med Educ. 2010;44(1):85–93.
11. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med. 2004;79(10 Suppl):S70–81.
12. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach.

2005;27(1):10–28.

13. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA. 2011;306(9):978–88.
14. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. Acad Med. 2013;88(6):872–83.
15. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003–2009. Med Educ. 2010;44(1):50–63.
16. Motola I, Devine LA, Chung HS, Sullivan JE, Issenberg SB. Simulation in healthcare education: a best evidence practical guide. AMEE Guide 82 Med Teach. 2013;35(10):e1511–30.
17. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review. Acad Med. 2011;86(6):706–11.
18. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ. 2009;43(4):303–11.
19. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: systematic review and meta-analysis by the Digital Health Education Collaboration. J Med Internet Res. 2019;21(7):e14676.
20. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med. 2010;85(10):1589–602.
21. Tudor Car L, Kyaw BM, Dunleavy G, Smart NA, Semwal M, Rotgans JI, et al. Digital problem-based learning in health professions: systematic review and meta-analysis. JMIR Med Educ. 2024;10:e51925.
22. Ellaway RH, Davies D. Design for learning: deconstructing virtual patient activities. Med Teach. 2011;33(4):303–10.
23. Yardley S, Teunissen PW, Dornan T. Experiential learning: AMEE Guide 63. Med Teach. 2012;34(2):e102–15.
24. Succar T, Zebington G, Billson F, Byth K, Barrie S, McCluskey P, et al. The impact of the Virtual Ophthalmology Clinic on medical students' learning: a randomised controlled trial. Eye (Lond). 2013;27(10):1151–7.
25. Succar T, Grigg J, Beaver HA, Lee AG. Advancing ophthalmology medical student education: international insights and strategies for enhanced teaching. Surv Ophthalmol. 2020;65(2):263–71.
26. Hopewell S, Chan A-W, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ. 2025;389:e081123.
27. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332.

28. Doig GS, Simpson F. Randomization and allocation concealment: a practical guide for researchers. J Crit Care. 2005;20(2):187–91.
29. Efird J. Blocked randomization with randomly selected block sizes. Int J Environ Res Public Health. 2011;8(1):15–20.
30. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63.
31. Bridge PD, Musial J, Frank R, Roe T, Sawilowsky S. Measurement practices: methods for developing content-valid student examinations. Med Teach. 2003;25(4):414–21.
32. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ. 2011;2:53–5.
33. Artino AR Jr, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide 87. Med Teach. 2014;36(6):463–74.
34. Sullivan GM, Artino AR Jr. Analyzing and interpreting data from Likert-type scales. J Grad Med Educ. 2013;5(4):541–2.
35. Sullivan GM, Feinn R. Using effect size — or why the p value is not enough. J Grad Med Educ. 2012;4(3):279–82.
36. Custers EJFM. Long-term retention of basic science knowledge: a review study. Adv Health Sci Educ Theory Pract. 2010;15(1):109–28.
37. Sirovich BE, Lipner RS, Johnston M, Holmboe ES. The association between residency training and internists' ability to practice conservatively. JAMA Intern Med. 2014;174(10):1640–8.
38. McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. BMC Med Res Methodol. 2007;7:30.
39. Council of Higher Education of the Republic of Türkiye (YÖK). Mezuniyet Öncesi Tıp Eğitimi Ulusal Çekirdek Eğitim Programı 2020 [National Core Curriculum for Undergraduate Medical Education 2020]. Ankara: YÖK; 2020. https://www.yok.gov.tr/Documents/Kurumsal/egitim_ogretim_dairesi/Ulusal-cekirdek-egitimi-programlari/mezuniyet-oncesi-tip-egitimi-cekirdek-egitimi-programi.pdf

Figures

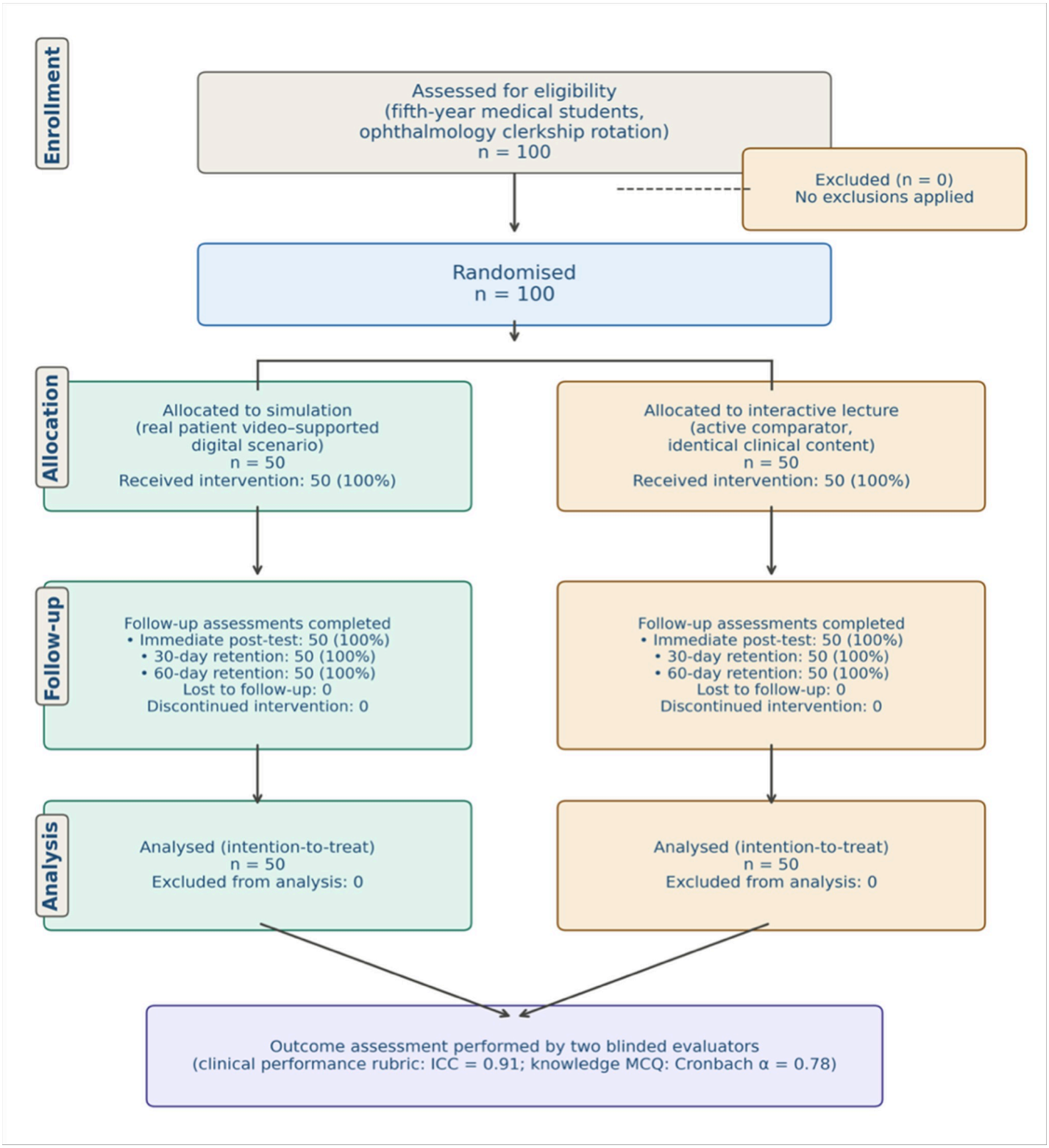


Figure 1

CONSORT 2025 flow diagram of participant recruitment, allocation, follow-up, and analysis. No participants were lost to follow-up; the analytic sample equalled the randomised sample (intention-to-treat).

Supplementary Files

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

- CONSORT2025ChecklistRedEye.docx