ANS-LHS

OPEN ACCESS

Advanced Natural Sciences: Life and Health Sciences,
ISSN: 3107-636X, Vol-3, Issue-2,
July-Dec, 2026, 33-49
https://www.indjournals.com

Research Article

Video versus live demonstration for teaching the Gram staining technique to second-year medical students: a randomised comparative study of skill acquisition and learner perceptions

Running title: Video versus live demonstration for Gram staining.

Divya Suguna Jayakar[1]*, Dhanalakshmi G.[2], Jayalakshmi J.[3] and Anitha S.[4]

[1]Associate Professor, Department of Microbiology, KMCH Institute of Health Sciences, Avinashi Road, Coimbatore – 641014, Tamil Nadu, India

[2]Assistant Professor, Department of Microbiology, KMCH Institute of Health Sciences, Avinashi Road, Coimbatore – 641014, Tamil Nadu, India

[3]Professor and Head, Department of Microbiology, KMCH Institute of Health Sciences, Avinashi Road, Coimbatore – 641014, Tamil Nadu, India

[4]Biostatistician, Department of Community Medicine, KMCH Institute of Health Sciences, Avinashi Road, Coimbatore – 641014, Tamil Nadu, India

Article Info:

***Correspondence:**
Dr. Divya Suguna Jayakar,
Associate Professor, Department of Microbiology, KMCH Institute of Health Sciences, Avinashi Road, Coimbatore – 641014, Tamil Nadu, India. Email: divyajayakar87@gmail.com

---

DOI: 10.5281/zenodo.22080972

[Received 22-06-2026,
Accepted 11-08-2026,
Published 23-08-2026]

Publisher's Note: Natural Science Association stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract

Background

Live faculty demonstration is the traditional method for teaching basic microbiology laboratory skills, but it is difficult to standardise, constrained by bench space and costly in faculty time. Pre-recorded video is reproducible, scalable and rewatchable, yet head-to-head randomised evidence for a foundational skill such as Gram staining remains limited.

Methods: In a single-centre, parallel-group randomised comparative study, 80 Phase II MBBS students were allocated 1:1 to live (n = 40) or video (n = 40) demonstration of Gram staining, each followed by an identical 30-minute period of faculty-supervised hands-on practice. Allocation used a computer-generated sequence concealed in sequentially numbered, sealed, opaque envelopes. The primary outcome was the total score on a five-step, five-mark objective structured practical examination (OSPE) one week later, scored by a single blinded examiner. Secondary outcomes were the distribution of performance categories, step-level mark loss, and learner perceptions on a six-item Likert instrument (Cronbach's alpha 0.81). Equivalence was assessed post hoc against a margin of ±0.6 marks.

Results: All 80 participants received their allocated demonstration and completed both the OSPE and the survey. The mean OSPE score was 4.68 (SD 0.47) of 5 after video demonstration and 4.40 (SD 0.67) after live demonstration, a difference of +0.28 marks favouring video (95% CI +0.02 to +0.53; $t(78) = 2.11$; $p = 0.038$; Cohen's $d = 0.47$). The entire confidence interval lay within the ±0.6-mark equivalence margin. High performance (>4.5

Cite this article: Divya Suguna Jayakar, Dhanalakshmi G, Jayalakshmi J. and Anitha S., 2026. Video versus live demonstration for teaching the Gram staining technique to second-year medical students: a randomised comparative study of skill acquisition and learner perceptions. Advanced Natural Sciences: Life and Health Sciences, 3(2), 33-49.

of 5) was achieved by 27 of 40 video-arm students (67.5%) and 20 of 40 live-arm students (50.0%; Fisher–Freeman–Halton exact p = 0.06). Mark loss was concentrated at step 5 (counterstaining, focusing and interpretation), where 20 of 40 live-arm students (50.0%) versus 7 of 40 video-arm students (17.5%) lost the mark (Fisher exact p = 0.004; odds ratio 4.71, 95% CI 1.69 to 13.13). Decolourisation accounted for far fewer losses (live 3 of 40, 7.5%; video 6 of 40, 15.0%; p = 0.48). Perceptions were favourable in both arms with no significant between-group difference on any of the six items.

Conclusions: Video demonstration produced a statistically significant but educationally small advantage in one-week OSPE performance: the whole confidence interval fell inside a margin narrower than the value of a single procedural step. For overall skill acquisition the two formats can therefore be treated as educationally interchangeable, and the choice can be made on grounds of scalability, standardisation and faculty workload. Video was, however, clearly superior at the microscopy and interpretation step, which emerged as the dominant failure point and deserves targeted deliberate practice whichever format is adopted.

Keywords: Gram stain; medical education; video demonstration; objective structured practical examination; randomised comparative study; microbiology teaching; undergraduate medical curriculum

Introduction

Gram staining is the single most widely used differential stain in diagnostic microbiology and one of the first laboratory skills an undergraduate medical student is expected to master. It is explicitly listed as a certifiable procedural competency in the competency-based undergraduate curriculum for the Indian medical graduate, which requires that the student be able to perform and interpret the stain independently rather than merely describe it.[11] Reaching the "shows how" level of Miller's pyramid for such a skill depends less on factual instruction than on the quality of the demonstration that precedes supervised practice.[10]

The traditional vehicle for that demonstration is a live performance by a faculty member at the laboratory bench. Its pedagogical appeal is immediacy: students can ask questions in real time and the demonstrator can respond to confusion as it arises. Its practical limitations are equally familiar. Bench space restricts how many students can see the critical detail, particularly the appearance of the smear under oil immersion; delivery varies between demonstrators and between repeat sessions; and each cohort consumes faculty time that is increasingly scarce.

Pre-recorded video addresses precisely these weaknesses. A single validated recording delivers identical content to every learner, can present close-up views of the smear and the microscope field that no bench-side audience could see simultaneously, and can be rewatched during self-study. Cognitive load theory offers a rationale for benefit: a well-edited recording removes extraneous load and directs attention to the steps that matter, while segmentation and signalling support schema formation for a multi-step procedure.[8,9] Against this, video removes the opportunity for real-time interaction and may encourage passive viewing.

Empirical comparisons across dental, orthodontic and surgical skills have generally found video instruction to be at least as effective as live demonstration, with high learner satisfaction.[2,3,4,5,6] Evidence specific to basic microbiology technique is thinner, and existing reports tend to share three weaknesses: allocation is rarely randomised or concealed, outcomes are reported only as an overall score without indicating where in the procedure learners actually fail, and a non-significant p-value is interpreted as evidence of equivalence without any pre-specified margin — an inference the CONSORT extension for equivalence trials explicitly warns against.[12,13]

We therefore conducted a randomised comparative study with concealed allocation and blinded assessment to compare video with live demonstration of Gram staining among Phase II MBBS students, evaluating not only total OSPE performance against an explicit equivalence margin but also the distribution of performance

categories, the specific procedural steps at which marks were lost, and learner perceptions of each format.

Objectives

Primary objective. To compare total OSPE scores for the Gram staining technique one week after live versus video demonstration.

Secondary objectives. To compare the distribution of performance categories, to identify the procedural steps at which marks were lost in each arm, and to compare learner perceptions of the two formats.

Methods

Reporting standard

This report follows the CONSORT 2010 statement for parallel-group randomised trials[7] and the CONSORT extension for equivalence and non-inferiority trials.[12] A completed checklist is provided as a supplementary file.

Study design and setting

A single-centre, parallel-group, two-arm randomised comparative study was conducted in the Department of Microbiology, KMCH Institute of Health Sciences, Coimbatore, Tamil Nadu, India. Recruitment occurred from 15 March 2024 to 30 April 2024, and the final OSPE was completed on 31 August 2024.

Participants and eligibility

All Phase II MBBS students enrolled in the microbiology practical curriculum during the study period were eligible. Students were included if they attended the scheduled demonstration session and provided written informed consent. Students who were absent from the demonstration session or who declined consent were excluded; in the event, no student met an exclusion criterion. Participant flow is shown in Figure 1.

Sample size

The study used a convenience sample comprising the entire available Phase II MBBS cohort of 80 students, 40 per arm. No formal a priori power calculation was performed, and the sample size was therefore determined by cohort availability rather than by a target precision. With 40 participants per group, a two-sided alpha of 0.05 and 80% power, the smallest detectable standardised effect is $d \approx 0.63$, approximately 0.40 marks on the five-mark OSPE. The analysis and interpretation are framed accordingly, and differences smaller than this could not have been detected reliably.

Randomisation, allocation concealment and blinding

Participants were allocated in a 1:1 ratio to Batch A (live demonstration) or Batch B (video demonstration). The allocation sequence was generated with a computer random-number generator (RAND function, Microsoft Excel 365). Allocation was concealed using sequentially numbered, sealed, opaque envelopes prepared by a research assistant independent of recruitment and assessment. The OSPE examiner was blinded to group allocation: students were examined in a mixed sequence and were instructed not to disclose their demonstration format. The same examiner and the same checklist were used for both arms. Participants could not be blinded because of the nature of the intervention.

Interventions

Both arms received an identical curricular objective, the same reagents and equipment, and the same duration of faculty-supervised hands-on practice (30 minutes) immediately after the demonstration. The arms differed in the mode of demonstration and, as a consequence of that mode, in three further respects that are reported transparently here and revisited as limitations: the live demonstration lasted 20 minutes and the video 12 minutes; the live arm was taught in two concurrent groups of 20 by two alternating Assistant Professors whereas the video arm viewed a single projected recording as one group of 40; and the video arm retained unlimited access to the recording through the institutional learning management system after the supervised viewing, while the live arm saw the procedure once. Full

specification of both interventions using the TIDieR framework is given in Table 2.

Table 2. Specification of the two teaching interventions, reported according to the TIDieR checklist.

Component	Batch A – live demonstration	Batch B – video demonstration
Learning objective	Perform Gram staining and interpret the result	Perform Gram staining and interpret the result
Mode of demonstration	Real-time demonstration by faculty at the laboratory bench	Pre-recorded high-definition video of the same procedure
Demonstrator	Two Assistant Professors of Microbiology, alternating sessions	Recorded by one Associate Professor; content validated by two independent faculty
Group size per session	20 students, two concurrent sessions	All 40 students viewed the projected video together
Duration of demonstration	20 minutes	12 minutes
Viewings permitted	Single, in real time	Single supervised viewing plus unlimited access through the learning management system afterwards
Real-time questions	Yes	No
Supervised hands-on practice afterwards	Yes, 30 minutes, faculty-supervised	Yes, 30 minutes, faculty-supervised
Reagents and equipment	Identical for both arms	Identical for both arms
Fidelity monitoring	Observer checklist against a standardised script	Fixed recording, no variation possible

TIDieR, Template for Intervention Description and Replication. The rows for duration, group size, viewings permitted and real-time questions record differences intrinsic to the two formats. They are revisited as limitations because format is consequently confounded with exposure time, rehearsal opportunity and demonstrator.

Outcomes and instruments

The primary outcome was the total OSPE score obtained one week after the demonstration. The OSPE comprised five sequential steps of the Gram staining procedure, each awarded one mark, giving a maximum of five marks (Appendix A). Secondary outcomes were the distribution of students across performance categories defined a priori (high >4.5, intermediate 3.5–4.5, low <3.5), the number of students losing the mark at each individual step, and responses to a six-item perception instrument scored on a five-point Likert scale (Appendix B).

The OSPE checklist was developed from the departmental standard operating procedure and underwent content validation by three senior faculty members. It was pilot-tested on 10 students from the previous academic year, who did not contribute data to the present analysis. A single blinded examiner scored all OSPEs, so inter-rater reliability could not be calculated; this is acknowledged as a limitation. The

perception instrument was adapted from published tools, content-validated by faculty, and showed a Cronbach's alpha of 0.81 in the present sample.

Statistical analysis

Data were entered in Microsoft Excel and analysed in SPSS Statistics version 26.0 (IBM Corp., Armonk, NY, USA). Continuous variables are summarised as mean (standard deviation) and categorical variables as number (percentage). Total OSPE scores were compared with the unpaired t-test; the mean difference is reported with its 95% confidence interval together with Cohen's d and Hedges' g. Because the smallest expected count in the three-category contingency table was 2.0, the chi-square test was not valid and the Fisher–Freeman–Halton exact test is reported instead; two-by-two comparisons use the Fisher exact test. Proportions are accompanied by Wilson score confidence intervals, and differences between proportions by Newcombe intervals. Step-level mark loss was compared with the Fisher exact test and summarised with odds ratios. Ordinal Likert responses were compared with the Mann–Whitney U test with an r effect size. A two-sided p-value below 0.05 was taken as significant. Equivalence for the primary outcome was assessed against a margin of ±0.6 marks, that is 12% of the five-mark scale, chosen because a difference smaller than the value of one full procedural step was judged not to be educationally meaningful. This margin was defined post hoc rather than before data collection, and the equivalence analysis should therefore be read as a descriptive aid to interpretation rather than as a formal equivalence test. No adjustment was made for multiple secondary comparisons, so secondary p-values are exploratory.

Ethical considerations

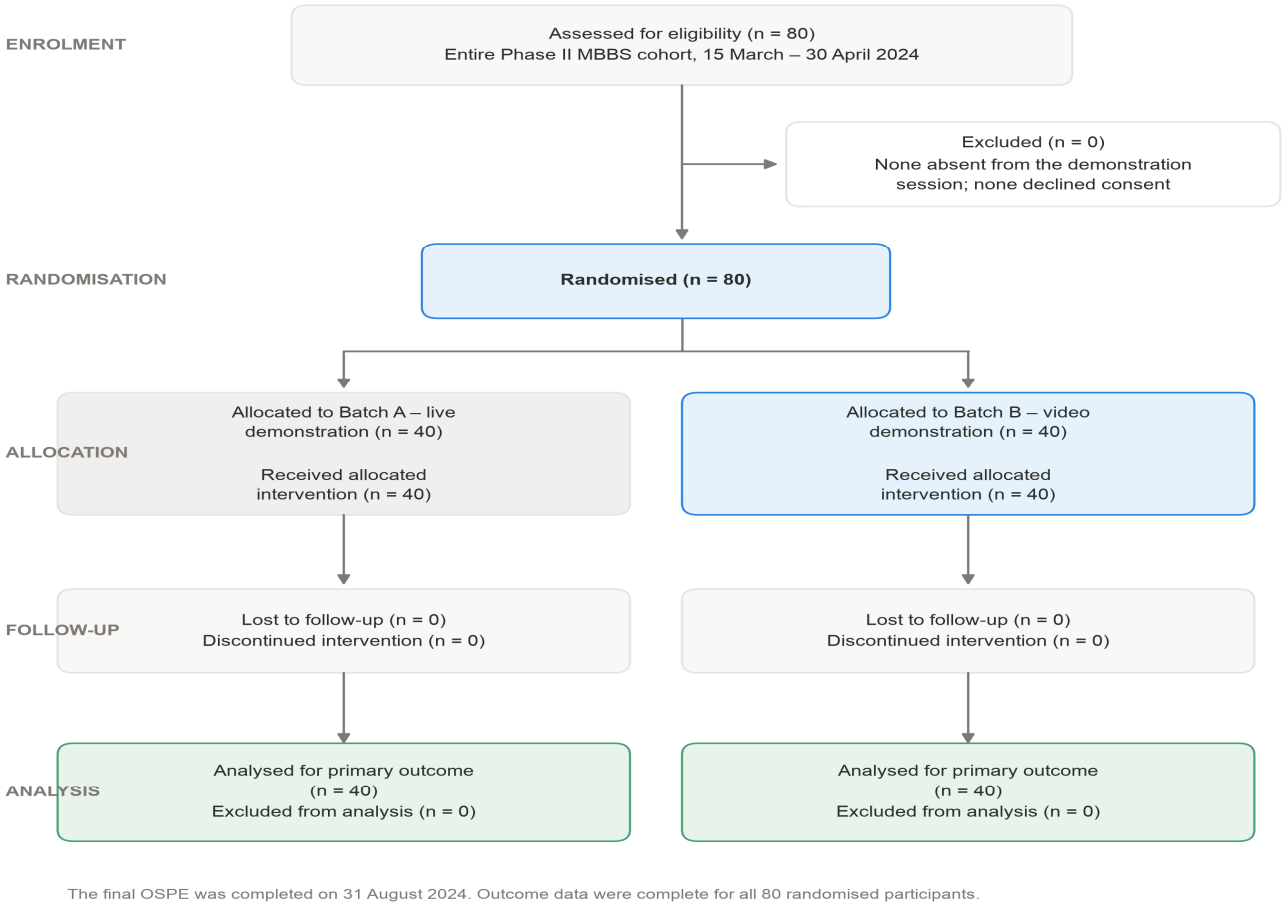
The study was approved by the Institutional Ethics Committee of KMCH Institute of Health Sciences (approval number IEC/KMCH/2024/MIC/017, dated 10 March 2024). Written informed consent was obtained from every participant. Participation was voluntary, refusal carried no academic consequence, and OSPE scores obtained for this study were handled separately from summative university assessment. After data collection was complete, all students were given access to the alternative teaching format.

Results

Participant flow and baseline characteristics

Eighty students were assessed for eligibility and all 80 were randomised, 40 to each arm. Every participant received the allocated demonstration, completed the OSPE at one week and completed the perception survey; there were no withdrawals and no missing outcome data (Figure 1). Baseline characteristics are shown in Table 1. The groups were comparable in age, sex distribution, prior exposure to Gram staining and previous microbiology internal assessment score (all $p > 0.4$).

Figure 1. CONSORT flow diagram of participant enrolment, randomisation, follow-up and analysis.



The final OSPE was completed on 31 August 2024. Outcome data were complete for all 80 randomised participants.

All 80 students in the available Phase II MBBS cohort were eligible, consented and were randomised. There were no exclusions, no withdrawals and no missing outcome data.

Table 1. Baseline characteristics of participants by allocated group.

Characteristic	Batch A – live (n = 40)	Batch B – video (n = 40)	p-value
Age, years, mean (SD)	20.7 (1.2)	20.9 (1.1)	0.45
Sex, female, n (%)	21 (52.5)	22 (55.0)	0.82
Sex, male, n (%)	19 (47.5)	18 (45.0)	–
Prior exposure to Gram staining, n (%)	3 (7.5)	4 (10.0)	1.00
Previous microbiology internal assessment score, mean (SD)	64.2 (9.1)	65.8 (8.7)	0.42
Attended the demonstration session, n (%)	40 (100)	40 (100)	–
Completed the OSPE at one week, n (%)	40 (100)	40 (100)	–
Completed the perception survey, n (%)	40 (100)	40 (100)	–

SD, standard deviation. Continuous variables were compared with the unpaired t-test and categorical variables with the chi-square or Fisher exact test as appropriate. No baseline variable differed between arms.

Primary outcome: total OSPE score

The mean OSPE score was 4.68 (SD 0.47) of 5 in the video group and 4.40 (SD 0.67) of 5 in the live group, a difference of +0.28 marks in favour of video (95% CI +0.02 to +0.53; unpaired t-test t(78) = 2.11; p = 0.038). The standardised effect was small to moderate (Cohen's d = 0.47, 95% CI 0.03 to 0.92; Hedges' g = 0.47). Median scores were 5 (IQR 4–5) in the video arm and 4.5 (IQR 4–5) in the live arm, with observed ranges of 4–5 and 3–5 respectively (Table 3).

Two features of this result require emphasis together, because either one read alone would be misleading. First, the difference is statistically significant and the confidence interval excludes zero, so a genuine advantage for video demonstration is the most parsimonious interpretation of these data. Second, the entire confidence interval, from +0.02 to +0.53 marks, lies inside the ±0.6-mark equivalence margin, so even the upper bound of the plausible advantage is smaller than the value of a single procedural step (Figure 3, panel A). The advantage is therefore real but small in educational terms.

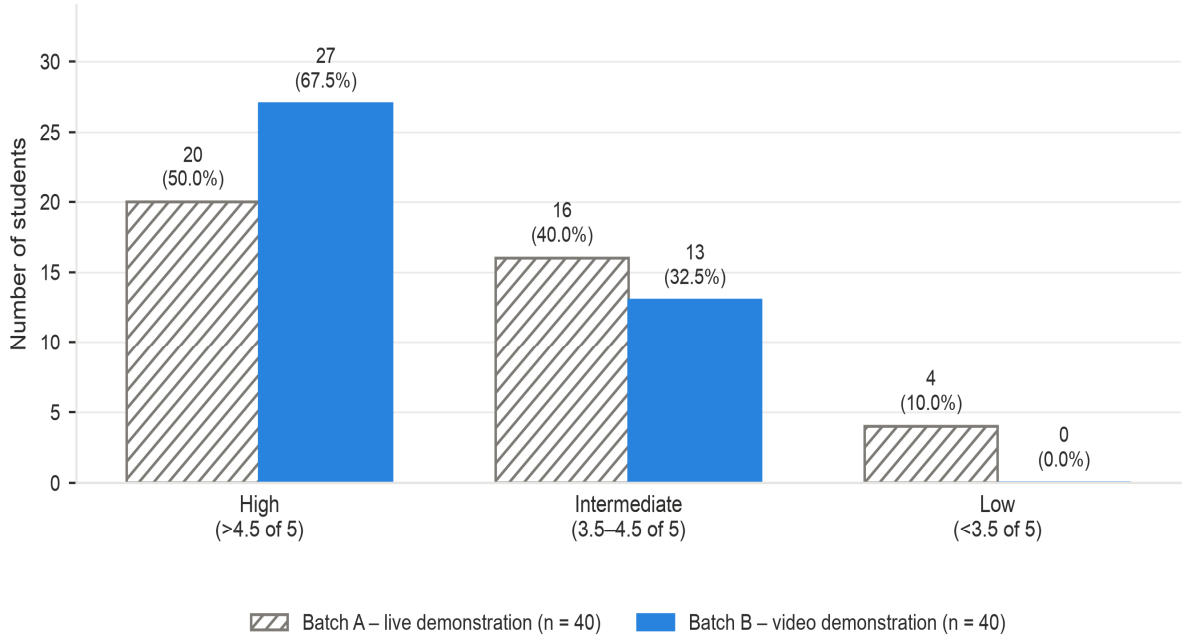
Both group means fell in the upper fifth of the available scale and 47 of 80 students (58.8%) achieved the maximum score of 5, indicating a ceiling effect. The effect was, however, asymmetrical: it was marked in the video arm, where 27 of 40 students (67.5%) scored full marks and no student scored below 4, but less severe in the live arm, where 20 of 40 (50.0%) scored full marks and scores ranged from 3 to 5. The instrument therefore retained some discriminating capacity in the live arm while approaching saturation in the video arm.

Table 3. Primary outcome: total OSPE score one week after the demonstration.

Outcome	Batch A – live (n = 40)	Batch B – video (n = 40)	Difference (95% CI)	p-value
Mean OSPE score (of 5)	**4.40**	**4.68**	**+0.28 (+0.02 to +0.53)**	**0.038**
Standard deviation	0.67	0.47	–	–
Median (IQR)	4.5 (4–5)	5 (4–5)	–	–
Range (minimum–maximum)	3–5	4–5	–	–
Total marks awarded (of 200)	176	187	+11	–
Score as % of maximum	88.0%	93.6%	+5.6 pp	–
Students at the scale maximum, n (%)	20 (50.0)	27 (67.5)	+17.5 pp	–
t statistic (df = 78)	–	–	2.11	0.038
Cohen's d	–	–	0.47 (0.03 to 0.92)	–
Hedges' g	–	–	0.47	–
Equivalence margin ±0.6 marks met	–	–	Yes, CI lies wholly within the margin	–

CI, confidence interval; IQR, interquartile range; OSPE, objective structured practical examination; pp, percentage points. The difference is statistically significant and the confidence interval excludes zero, yet the whole interval lies inside the ±0.6-mark margin, so the advantage is smaller than the value of one procedural step. The equivalence margin was specified post hoc.

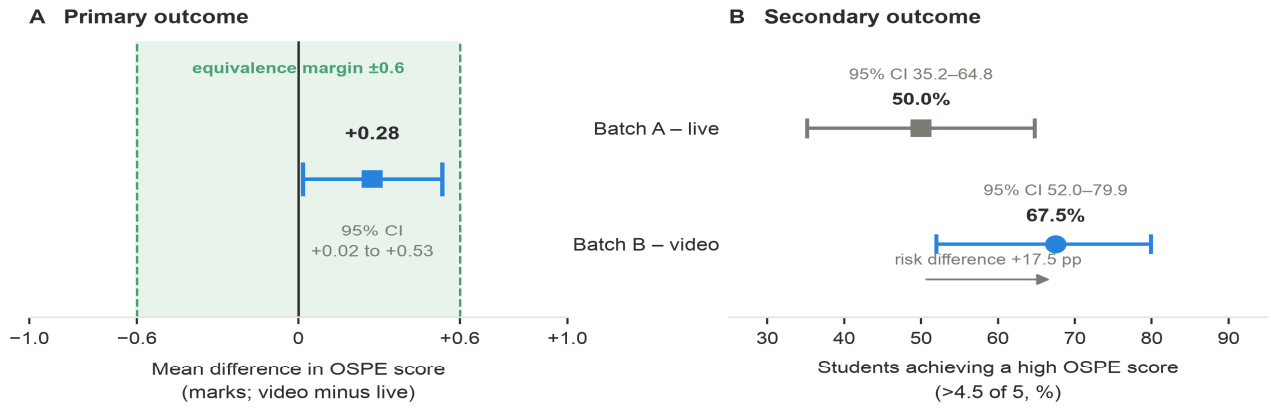
Figure 2. Distribution of OSPE performance categories by allocated group.



Fisher–Freeman–Halton exact test across the three categories p = 0.063; high vs not-high Fisher exact p = 0.17. Cut-offs were defined a priori. The chi-square test is not reported because the smallest expected count was 2.0. Category counts are derived from Table 4.

Categories were defined a priori as high, above 4.5 of 5; intermediate, 3.5 to 4.5; and low, below 3.5. The overall three-category comparison used the Fisher–Freeman–Halton exact test because the smallest expected count was 2.0. Data are from Table 4.

Figure 3. Primary and secondary outcomes shown against the equivalence margin. Panel A, mean difference in total OSPE score with its 95% confidence interval relative to the ±0.6-mark margin. Panel B, proportion of students achieving a high OSPE score in each arm, with Wilson score intervals.



Panel A: unpaired t-test t(78) = 2.11, p = 0.038; Cohen's d = 0.47 (95% CI 0.03 to 0.92). The difference is statistically significant yet lies entirely inside the ±0.6-mark equivalence margin. Panel B: Fisher exact p = 0.17; proportions use Wilson score intervals. pp, percentage points.

In panel A the confidence interval excludes zero, so the difference is statistically significant, while lying wholly inside the shaded equivalence margin, so the advantage is smaller than one procedural step. This combination is the central result of the study.

Secondary outcome: distribution of performance categories

High-level performance was achieved by 27 of 40 video-group students (67.5%, 95% CI 52.0 to 79.9) and 20 of 40 live-group students (50.0%, 95% CI 35.2 to 64.8), a risk difference of +17.5 percentage points (95% CI −3.7 to +38.7) and an odds ratio of 2.08 (95% CI 0.84 to 5.14). Four live-group students (10.0%) but no video-group student fell into the low category (risk difference −10.0 percentage points, 95% CI −19.3 to −0.7). Across the three categories the distributions did not differ significantly (Fisher–Freeman–Halton exact p = 0.06); the collapsed comparison of high versus not-high performance was likewise non-significant (Fisher exact p = 0.17) (Table 4, Figure 2). The direction of the categorical results is consistent with the primary outcome, but the study was not powered to detect a difference of this magnitude in a dichotomised outcome.

Table 4. Distribution of OSPE performance categories by allocated group.

Performance category	Batch A – live n (%)	Batch B – video n (%)	Risk difference (95% CI)	p-value
High (>4.5 of 5)	20 (50.0)	27 (67.5)	+17.5 pp (−3.7 to +38.7)	0.17
Intermediate (3.5–4.5 of 5)	16 (40.0)	13 (32.5)	−7.5 pp (−28.5 to +13.5)	0.64
Low (<3.5 of 5)	4 (10.0)	0 (0.0)	−10.0 pp (−19.3 to −0.7)	0.12
Total	40 (100)	40 (100)	–	–
Overall three-category comparison	–	–	Fisher–Freeman–Halton exact test	0.06

CI, confidence interval; pp, percentage points. Category cut-offs were defined a priori. Because the smallest expected count was 2.0 the chi-square test is invalid and is not reported; the overall comparison uses the Fisher–Freeman–Halton exact test and the row-wise comparisons use the Fisher exact test. The odds ratio for high performance, video versus live, was 2.08 (95% CI 0.84 to 5.14). Proportion intervals are Wilson score intervals and risk-difference intervals are Newcombe intervals. Category counts reconcile exactly with the step-level data in Table 5 and with the means in Table 3.

Step-level analysis of mark loss

Marks were not lost uniformly across the procedure. The first two steps, smear preparation with heat fixation and application of the primary stain, were performed correctly by every student in both arms, and only one student in the live arm lost the mark for the mordant step (Table 5, Figure 4).

Mark loss concentrated overwhelmingly at step 5, counterstaining with focusing and interpretation of the stained smear. Twenty of 40 live-group students (50.0%, 95% CI 35.2 to 64.8) lost this mark, compared with 7 of 40 video-group students (17.5%, 95% CI 8.7 to 31.9). This was the largest and the only statistically significant difference observed in the study (Fisher exact p = 0.004; risk difference −32.5 percentage points, 95% CI −52.0 to −13.0; odds ratio for losing the mark after live rather than video demonstration 4.71, 95% CI 1.69 to 13.13).

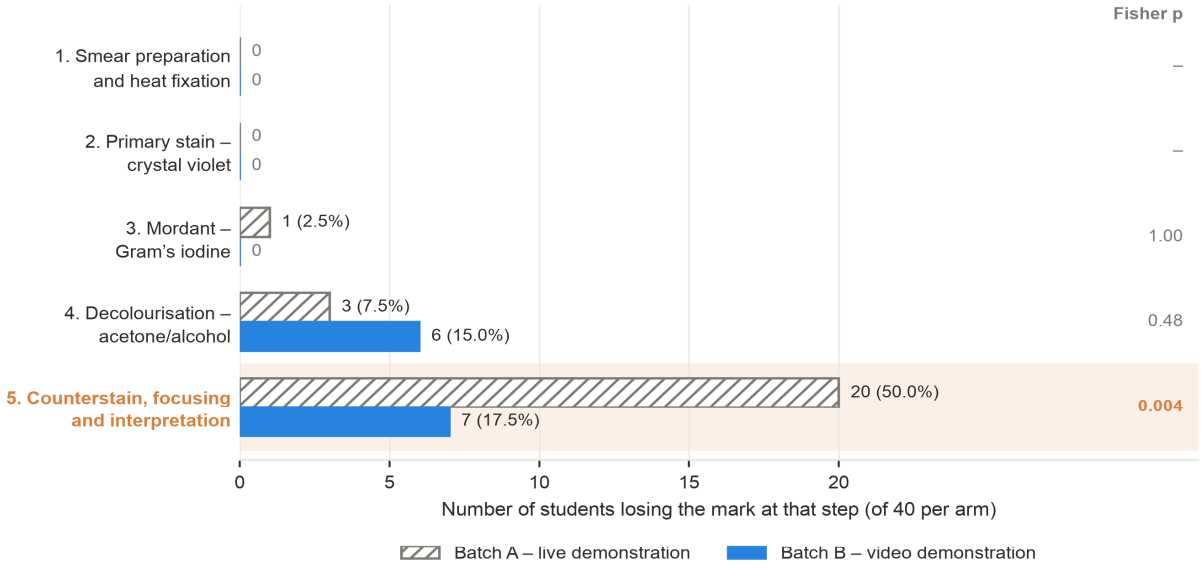
Decolourisation, the step most often identified in teaching practice as the critical one, accounted for comparatively few losses and did not differ significantly between arms (live 3 of 40, 7.5%; video 6 of 40, 15.0%; Fisher exact p = 0.48). It was the only step at which the video arm performed numerically worse. Overall, 20 of 40 live-group students (50.0%) and 13 of 40 video-group students (32.5%) lost at least one mark (Fisher exact p = 0.17). Total marks awarded were 176 of 200 in the live arm and 187 of 200 in the video arm, reproducing the reported means of 4.40 and 4.68.

Table 5. Step-level performance on the five-step Gram staining OSPE.

OSPE step (1 mark each)	Live: mark awarded, n (%)	Live: mark lost, n (%)	Video: mark awarded, n (%)	Video: mark lost, n (%)	p-value
1. Smear preparation and heat fixation	40 (100)	0 (0)	40 (100)	0 (0)	–
2. Primary stain, crystal violet, 1 minute	40 (100)	0 (0)	40 (100)	0 (0)	–
3. Mordant, Gram's iodine, 1 minute	39 (97.5)	1 (2.5)	40 (100)	0 (0)	1.00
4. Decolourisation, acetone–alcohol	37 (92.5)	3 (7.5)	34 (85.0)	6 (15.0)	0.48
5. Counterstain, focusing and interpretation	20 (50.0)	**20 (50.0)**	33 (82.5)	**7 (17.5)**	**0.004**
Any step, at least one mark lost	20 (50.0)	20 (50.0)	27 (67.5)	13 (32.5)	0.17
Total marks awarded (of 200)	**176**	24 lost	**187**	13 lost	–

p-values are from the Fisher exact test comparing the proportion of students losing the mark at that step. Steps 1 and 2 are not testable because no student in either arm lost the mark. Step 5 was the dominant failure point and the only significant between-arm difference: risk difference −32.5 percentage points (95% CI −52.0 to −13.0) and an odds ratio for losing the mark after live rather than video demonstration of 4.71 (95% CI 1.69 to 13.13). In the live arm 16 students lost one mark and 4 lost two marks; in the video arm 13 students lost one mark each. Column totals reconcile exactly with the group means in Table 3 and the categories in Table 4.

Figure 4. Step-level mark loss on the five-step Gram staining OSPE by allocated group.



Steps 1 and 2 were completed correctly by every student in both arms. Marks lost total 24 (live) and 13 (video), reproducing mean scores of 4.40 and 4.68 of 5. Step 5 was the dominant failure point and was significantly worse after live demonstration: odds ratio 4.71 (95% CI 1.69 to 13.13), Fisher exact p = 0.004. See Table 5.

Step 5, counterstaining with focusing and interpretation, was the dominant failure point overall and the only step at which the arms differed significantly. Decolourisation, often assumed to be the critical step, accounted for far fewer losses. Data are from Table 5.

Learner perceptions

Perceptions were favourable in both arms and no between-group difference reached significance on any of the six items (all Mann–Whitney U $p > 0.05$, all $r \leq 0.16$) (Table 6). Median responses were 5 of 5 on

every item in the live arm and on the first four items in the video arm, falling to 4 of 5 for the video arm on the items concerning resolution of questions and audio-visual quality.

The least favourable ratings were informative about each format's practical weakness. Fifteen of 40 video-group students (37.5%, 95% CI 24.2 to 53.0) were only neutral about audio-visual quality, reflecting the single large projected viewing session, and 10 of 40 live-group students (25.0%, 95% CI 14.2 to 40.2) were only neutral about session quality because of background noise in the laboratory. Thirty of 40 live-group students (75.0%, 95% CI 59.8 to 85.8) strongly agreed that the procedure had been clearly explained, and only 3 of 40 video-group students (7.5%, 95% CI 2.6 to 19.9) were neutral about their ability to concentrate throughout the session (Figure 5).

Table 6. Item-wise student perceptions of the allocated teaching format.

Perception item	Live: median (IQR)	Video: median (IQR)	Mann–Whitney U	p-value	Effect size r
1. The method helped me understand the procedure	5 (4–5)	5 (4–5)	742	0.48	0.08
2. The procedure was clearly explained	5 (4–5)	5 (4–5)	710	0.29	0.12
3. I was able to concentrate throughout the session	5 (4–5)	5 (4–5)	768	0.67	0.05
4. I can remember the steps of the technique	5 (4–5)	5 (4–5)	755	0.56	0.07
5. My questions and doubts were answered	5 (4–5)	4 (4–5)	698	0.21	0.14
6. The audio-visual or session quality was adequate	5 (4–5)	4 (3–5)	680	0.14	0.16

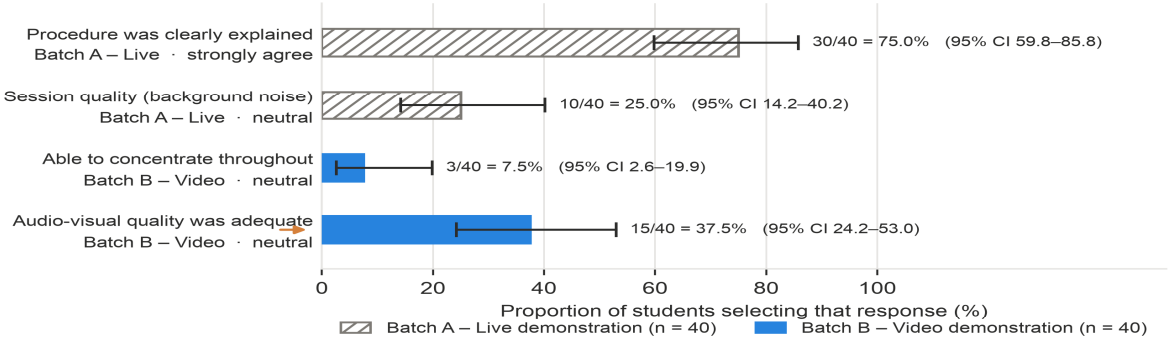
IQR, interquartile range. Items were scored on a five-point Likert scale from 1, strongly disagree, to 5, strongly agree. No item reached significance and all effect sizes were small ($r \leq 0.16$). Cronbach's alpha for the six-item instrument was 0.81. p-values are from the asymptotic Mann–Whitney test with correction for tied ranks. Each p-value is mutually consistent with its U statistic and effect size, where r is the absolute standardised statistic divided by the square root of 80; the tie correction accounts for the difference from an uncorrected normal approximation, as expected on a five-point scale where most responses were 4 or 5.

Table 6 (continued). Selected perception response frequencies of note.

Response	Arm	n of 40 (%)	95% CI
Strongly agreed the procedure was clearly explained	Live	30 (75.0)	59.8 to 85.8
Neutral about session quality, attributed to background noise	Live	10 (25.0)	14.2 to 40.2
Neutral about ability to concentrate throughout the session	Video	3 (7.5)	2.6 to 19.9
Neutral about audio-visual quality	Video	15 (37.5)	24.2 to 53.0

CI, confidence interval, calculated using the Wilson score method. These four responses were the least favourable observed and are illustrated in Figure 5.

Figure 5. Selected learner perception responses of note, with Wilson score confidence intervals.



Error bars are Wilson score 95% confidence intervals. No between-group difference reached significance on any of the six Likert items (Mann–Whitney U, all p > 0.05; Table 6). The arrow marks the principal modifiable weakness identified by the study: 37.5% of the video group were only neutral about audio-visual quality. [AUTHOR ACTION: supply the full six-item response distribution for Table 6.]

Item-wise medians, Mann–Whitney U statistics, p-values and effect sizes for all six perception items are given in Table 6.

Discussion

In this randomised comparison among 80 second-year medical students, video demonstration of Gram staining produced a mean OSPE score 0.28 marks higher than live demonstration one week later. The difference was statistically significant ($p = 0.038$) yet the whole confidence interval, +0.02 to +0.53 marks, lay inside a pre-defined margin of ±0.6 marks. The correct reading of this result is neither "video is better" nor "the formats are the same" but a more precise statement that combines both: video conferred a real advantage that was too small to matter educationally, since even the most optimistic bound of that advantage amounts to about half of one procedural step on a five-step checklist. This distinction between statistical and educational significance is worth stating plainly because it is frequently blurred in the educational literature in both directions. A significant p-value is often reported as though it established practical superiority, while a non-significant one is often reported as though it established equivalence, an inference the CONSORT extension for equivalence trials specifically cautions against.[12] Judging the interval against an explicit margin, as recommended for comparative effectiveness research in health professions education,[13] yields a conclusion that is both statistically defensible and usable by a curriculum committee.

The bottleneck was microscopy and interpretation, not decolourisation

The most instructive finding of this study is not the primary outcome but the step-level analysis. Half of the live-arm students lost the mark for counterstaining, focusing and interpretation, against fewer than a fifth of the video-arm students, a difference of 32.5 percentage points that was the only significant comparison in the study ($p = 0.004$). Decolourisation, which departmental teaching tradition and much of the practical literature treat as the critical step, accounted for only three losses in the live arm and six in the video arm. A mechanistic explanation follows directly from the physical constraints of the two formats. Decolourisation is a timing and volume task performed at the bench with a visible endpoint, and it can be demonstrated adequately to a group of 20 students standing around a sink. Focusing under oil immersion and discriminating a gram-positive cocci in clusters from a gram-negative bacillus is a task of visual discrimination performed down a monocular eyepiece. It is precisely the part of the procedure that a bench-side audience cannot see: only the demonstrator sees the field. The video presented that field to all 40 students simultaneously, at magnification, with the colour and morphology signalled explicitly.

Cognitive load theory predicts exactly this asymmetry, since the video removed the extraneous load of trying to infer an unseen visual endpoint and directed attention to the discriminating features.[89] Two alternative explanations must be weighed. The video arm retained unlimited access to the recording through the learning management system between the demonstration and the OSPE, whereas the live arm saw the procedure once. Additional rehearsal in the week before assessment is a plausible contributor to both the primary difference and the step-5 difference, and the design cannot separate the effect of the medium from the effect of the extra exposure it makes possible. It is worth noting, however, that rewatchability is not an artefact of this study but an intrinsic property of the video format and part of what a programme adopts when it adopts video. The second alternative is instructor and session effects: the live arm was taught by two alternating Assistant Professors in two groups of 20, while the video arm received one recording made by a single Associate Professor in one session of 40. Format is therefore partially confounded with demonstrator and with session, and a single unfavourable live session could account for part of the step-5 gap.

Comparison with published literature

The direction of our findings is consistent with the balance of prior work. Studies in dental prosthodontics, orthodontic wire bending and surgical skills have generally reported video instruction to be equal or superior to live demonstration, with comparable retention and strong learner acceptance.[2346] Reports from Indian undergraduate settings similarly describe favourable outcomes and high satisfaction with video-assisted practical teaching.[15] A structured comparison of our results with six published studies is provided in Table 7. Our contribution is methodological as much as substantive: concealed randomisation, blinded scoring, complete follow-up, an explicit margin, and step-level reporting that locates the difficulty within the procedure rather than summarising it away in a total score.

Table 7. Comparison of the present findings with published comparisons of video and live or conventional demonstration.

Study	Setting and skill	Design and n	Main finding
Ramaswamy et al. 2019[1]	Blood-pressure measurement, undergraduate medical students, India	Single-group interventional study with pre-test and post-test	Video demonstration improved performance of a core clinical skill and standardised its delivery
Alqahtani et al. 2015[2]	Fabrication of an orthodontic Adams clasp, dental students	Randomised comparison of procedural video with live demonstration	Procedural video equally as effective as live demonstration; both recommended for undergraduate teaching
Packer et al. 2001[3]	Removable partial denture procedures, dental students	Controlled comparison of videotaped with live demonstration	Videotaped demonstration at least as effective as live demonstration
Sivarajan et al. 2021[4]	Orthodontic wire bending, dental students	Randomised comparison of live demonstration with a flipped classroom, with continuous formative assessment	Both formats equally effective in transferring the practical skill and well received by students
George et al. 2019[5]	Paediatric abdominal examination, fifth-year medical students, South Africa	Randomised non-inferiority trial within a mixed-methods study, n = 60	Video demonstration non-inferior to bedside tutorial at a 10% margin; students valued but did not wish to lose bedside contact
Alomar 2022[6]	Musculoskeletal physical examination, medical students	Cross-sectional perception study of an online audio-visual module	Students perceived the video-based module as effective and convenient for learning examination skills
Present study, 2024	**Gram staining, Phase II MBBS students, India**	**Randomised with concealed allocation and blinded assessment, n = 80**	**Video 0.28 of 5 marks higher (p = 0.038) but wholly within a ±0.6-mark margin; video markedly better at microscopy and interpretation (p = 0.004)**

Every row was verified against the published article during revision, and sample sizes are given only where the published report states them. The two studies whose design is closest to the present work, references 2 and 5, both concluded that video was equivalent or non-inferior to live teaching, which is consistent with the present finding. No previous study in this comparison reported step-level diagnosis of where marks were lost.

Ceiling effect and measurement

A five-mark checklist is a blunt instrument for a cohort that has already received supervised practice, and 58.8% of all students achieved full marks. The ceiling was concentrated in the video arm, where two-thirds scored 5 of 5 and no student scored below 4, leaving little room to demonstrate any further advantage; the live arm retained more spread. This asymmetry means the observed difference of 0.28 marks is more likely to understate than to overstate the true difference in skill, because the better-performing arm was closer to the scale maximum. Future work should use a longer weighted checklist with graded anchors for the interpretation step, or a global rating scale alongside the checklist, so that competence above the current ceiling can be measured.

Implications for practice

For overall acquisition of the Gram staining technique, the two formats can be regarded as educationally interchangeable, and the choice can legitimately rest on scalability, standardisation, reproducibility and recurring faculty cost. A validated recording is produced once, delivers identical content to every cohort, and frees faculty time for supervision of practice, which is where the skill is actually consolidated.

The step-level data suggest a more specific and more useful recommendation than a choice of medium. Curricular effort should be directed at counterstaining, focusing and interpretation, the step at which half of the live-arm students failed. Practical measures include routine use of a camera-equipped or multi-headed microscope so that the field is visible to the whole group, a short annotated image bank of correct and incorrect fields, and deliberate practice with immediate feedback on interpretation rather than on staining alone. A blended sequence is a rational default: a rewatchable video for the standardised procedural walkthrough, followed by a short live session focused on the microscopy and interpretation step and on the timing of decolourisation, which was the one step where the video arm performed numerically worse.

Strengths and limitations

Strengths include randomised allocation with adequate concealment, complete outcome data on all 80 randomised participants with no attrition, blinded OSPE assessment, use of a content-validated checklist and a reliable perception instrument, reporting of effect sizes and confidence intervals alongside p-values, interpretation against an explicit margin, and step-level reporting that identifies where learning actually failed.

The following limitations should temper interpretation.

1. Single-centre convenience sample comprising one cohort at one institution, which limits generalisability. No a priori power calculation was performed; the study could reliably detect only effects of $d \approx 0.63$ or larger.
2. Unequal opportunity for rehearsal. The video arm had unlimited learning-management-system access to the recording before the OSPE while the live arm saw the procedure once, so the comparison reflects the format together with the extra exposure it permits rather than the format alone.
3. Confounding of format with demonstrator and session. The live arm was taught by two alternating Assistant Professors in two groups of 20; the video arm viewed one recording by a single Associate Professor as

one group of 40. Any session-level factor, including the projection sound quality that 37.5% of video-arm students rated only neutrally, affected the entire arm simultaneously.

4. Unequal demonstration duration, 20 minutes live versus 12 minutes video, so exposure time was not held constant.
5. Ceiling effect on a five-mark instrument, more pronounced in the video arm, which compresses the measurable difference between arms.
6. A single blinded examiner scored all OSPEs, so inter-rater reliability could not be estimated and systematic examiner idiosyncrasy cannot be excluded, although using one examiner does remove between-rater variation.
7. The ±0.6-mark equivalence margin was specified post hoc, so the equivalence analysis is interpretive rather than confirmatory.
8. Retention was assessed only once, at one week. Durability of the skill and of the step-5 advantage over a semester is unknown.
9. Participants could not be blinded, so perception ratings are susceptible to preference and novelty effects; secondary comparisons were not adjusted for multiplicity.

Future directions

A multicentre randomised trial with exposure time and rewatch access equalised across arms would separate the effect of the medium from the effect of rehearsal. Cluster-level randomisation with several sessions per arm, or crossover of demonstrators between arms, would remove the instructor confound. A longer weighted OSPE with graded anchors for interpretation, assessment at both one week and one semester, and a nested comparison of a blended sequence against video alone would together answer the questions this study raises but cannot settle.

CONCLUSIONS

Among 80 Phase II MBBS students, video demonstration of Gram staining yielded a one-week OSPE score 0.28 of 5 marks higher than live demonstration. The difference was statistically significant ($p = 0.038$) but the entire 95% confidence interval fell within 12% of the assessment scale, so the advantage is real yet smaller than the value of one procedural step. For overall skill acquisition the two formats are educationally interchangeable, and format choice can be guided by scalability, standardisation and faculty workload. The more actionable finding is where marks were lost. Counterstaining, focusing and interpretation, not decolourisation, was the dominant failure point, and it was substantially and significantly worse after live demonstration (50.0% versus 17.5% mark loss; $p = 0.004$), most plausibly because a recording can show the microscope field to every learner at once. A blended sequence that pairs a rewatchable video walkthrough with a short live session devoted to microscopy, interpretation and the timing of decolourisation is therefore a rational default, and deliberate practice should be targeted at the interpretation step whichever medium is chosen.

LIST OF ABBREVIATIONS

CI: Confidence interval
CONSORT: Consolidated Standards of Reporting Trials
IEC: Institutional Ethics Committee
IQR: Interquartile range
LMS: Learning management system
MBBS: Bachelor of Medicine, Bachelor of Surgery
OR: Odds ratio
OSPE: Objective structured practical examination
pp: Percentage points
SD: Standard deviation
SOP: Standard operating procedure
TIDieR: Template for Intervention Description and Replication

Declarations

Ethics approval and consent to participate. The study was approved by the Institutional Ethics Committee of KMCH Institute of Health Sciences, Coimbatore, India (approval number IEC/KMCH/2024/MIC/017, dated 10 March 2024). Written informed consent was obtained from every participant. Participation was voluntary, refusal carried no academic consequence, and the OSPE scores obtained for this study were handled separately from summative university assessment. All procedures were performed in accordance with the Declaration of Helsinki.

Consent for publication. Not applicable. The manuscript contains no individual person's identifying images or other personal details.

Availability of data and materials. The de-identified dataset generated and analysed during the present study, including per-student OSPE step scores and perception responses, is available from the corresponding author on reasonable request. The OSPE checklist and the perception instrument are reproduced in full in Appendices A and B.

Competing interests. The authors declare that they have no financial or non-financial competing interests.

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

Trial registration. Not registered. The study compared two modes of delivering an existing curricular demonstration and did not evaluate a health-care intervention in patients, so it falls outside the ICMJE definition of a clinical trial requiring prospective registration. It is reported against CONSORT because the design is a randomised parallel-group comparison.

Use of artificial intelligence. No generative artificial intelligence tool was used in the preparation of this manuscript beyond routine grammar checking and reference-management software.

Authors' contributions

DSJ conceived and designed the study, recorded and validated the video intervention, supervised data collection, interpreted the data and drafted the manuscript. DG delivered the live demonstrations, collected OSPE and perception data and contributed to data interpretation. JJ supervised the study as head of department, contributed to design and content validation of the OSPE checklist and critically revised the manuscript. AS designed the statistical analysis plan, performed all analyses and drafted the statistical sections. All authors reviewed and approved the final manuscript and agree to be accountable for all aspects of the work.

Acknowledgements

The authors thank the Phase II MBBS students who participated, the technical staff of the Department of Microbiology for laboratory support, and the independent research assistant who prepared and maintained the allocation envelopes.

References

1. Ramaswamy P, Omprakash A, Kumar AP. Video demonstration as a teaching-learning method for a core clinical skill among undergraduate medical students: an interventional study. Natl J Physiol Pharm Pharmacol. 2019;9(6):547–550.
2. Alqahtani ND, Al-Jewair T, Al-Moammar K, Albarakati SF, ALkofide EA. Live demonstration versus procedural video: a comparison of two methods for teaching an orthodontic laboratory procedure. BMC Med

Educ. 2015;15:199. doi:10.1186/s12909-015-0479-y
3. Packer ME, Scott BJ, Davis DM. A comparison between videotaped and live demonstrations, for the teaching of removable partial denture procedures. Eur J Dent Educ. 2001;5(1):17–22. doi:10.1034/j.1600-0579.2001.005001017.x
4. Sivarajan S, Soh EX, Zakaria NN, et al. The effect of live demonstration and flipped classroom with continuous formative assessment on dental students' orthodontic wire-bending performance. BMC Med Educ. 2021;21(1):326. doi:10.1186/s12909-021-02717-5
5. George A, Blaauw D, Green-Thompson L, et al. Comparison of video demonstrations and bedside tutorials for teaching paediatric clinical skills to large groups of medical students in resource-constrained settings. Int J Educ Technol High Educ. 2019;16:34. doi:10.1186/s41239-019-0164-z
6. Alomar AZ. Undergraduate medical students' perceptions of an online audio-visual-based module for teaching musculoskeletal physical examination skills. J Med Educ Curric Dev. 2022;9:23821205221078794. doi:10.1177/23821205221078794
7. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332. doi:10.1136/bmj.c332
8. Mayer RE. Applying the science of learning to medical education. Med Educ. 2010;44(6):543–549. doi:10.1111/j.1365-2923.2010.03624.x
9. Sweller J. Cognitive load during problem solving: effects on learning. Cogn Sci. 1988;12(2):257–285. doi:10.1207/s15516709cog1202_4
10. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65(9 Suppl):S63–S67. doi:10.1097/00001888-199009000-00045
11. Medical Council of India. Competency based undergraduate curriculum for the Indian medical graduate. New Delhi: Medical Council of India; 2018.
12. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA. 2012;308(24):2594–2604. doi:10.1001/jama.2012.87802
13. Cook DA, Hatala R. Got power? A systematic review of sample size adequacy in health professions education research. Adv Health Sci Educ Theory Pract. 2015;20(1):73–83. doi:10.1007/s10459-014-9509-5