Revista Española de Educación Médica

UNIVERSIDAD DE MURCIA

Comparison of AI-generated and human-made animated videos for medical education: experts and students preferred AI over humans

Abdullah Bedir Kaya[1,2], Yavuz Selim Kıyak[3]*, Özlem Coşkun[4], and Işıl İrem Budakoğlu[5]

[1]Department of Computer Technologies, Hitit University, Çorum 19030, Turkiye, abedirkaya@hitit.edu.tr, https://orcid.org/0000-0003-0124-2127

[2]Health Sciences Institute, Gazi University, Ankara 06500, Turkiye.

[3]Department of Medical Education and Informatics, Faculty of Medicine, Gazi University, Ankara 06500, Turkiye, yskiyak@gazi.edu.tr, https://orcid.org/0000-0002-5026-3234

[4]Department of Medical Education and Informatics, Faculty of Medicine, Gazi University, Ankara 06500, Turkiye, ocoskun@gazi.edu.tr, https://orcid.org/0000-0001-8716-1584

[5]Department of Medical Education and Informatics, Faculty of Medicine, Gazi University, Ankara 06500, Turkiye, isilirem@gazi.edu.tr, https://orcid.org/0000-0003-1517-3169

*Correspondence: yskiyak@gazi.edu.tr

Recibido: 9/1/25; Aceptado: 9/22/25; Publicado: 9/23/25

Abstract:

Objective: This study compared medical students and experts, and evaluated a frames-to-video AI-generated problem-based learning (PBL) trigger against its scene-matched human-made animated counterpart in terms of evaluations and preferences. **Study Design:** A mixed-methods study was conducted at a medical school. Two scene-matched videos were used: an AI-generated video and an animated (human-made) video. Students (n=210; Years 2–5) viewed both videos in counterbalanced order and rated eight 5-point Likert items for each; they also indicated their preferred video for engagement, emotional impact, and PBL use. A multidisciplinary expert panel (n=104) evaluated only the AI video on comparable items and provided open-ended comments. Mann–Whitney-U tests compared experts with students on the AI video; Wilcoxon signed-rank tests compared students' ratings across videos. Qualitative data underwent thematic analysis. **Results**: Students rated the AI-generated video significantly higher than the animated video on all eight items (all $p \leq .026$) and preferred it for engagement (83.8%), emotional impact (81.0%), and PBL use (79.0%). Experts' ratings of the AI video were also high and exceeded students' ratings on visual quality, distraction avoidance, and visual consistency ($p \leq .001$). Qualitative themes highlighted realism, suitability for PBL sessions, and strong engagement, while suggested improvements included micro-continuity, pronunciation, and body language. **Conclusion**: Within the PBL context, a frames-to-video AI workflow produced a fully synthetic trigger that was preferred by students and endorsed by experts. AI-generated triggers appear feasible, acceptable, and educationally promising, provided attention is given to fine-grained audiovisual continuity and communication cues.

Keywords: artificial intelligence, video generation, problem-based learning, medical education.

1. Introduction

Problem-based learning (PBL) is a small-group, self-directed learning approach in which students identify learning objectives based on a clinical scenario, and it has traditionally been conducted with paper-based evolving cases (1). In PBL sessions, video-based triggers are used frequently (2); they enrich the clinical context (3) and are preferred by both students and facilitators (4). With advancing technology, animated video (5) implementations have begun to be adopted.

Advances in artificial intelligence (AI) that enable video generation offer a rapid, iterative, and relatively low-cost pathway for developing educational materials. For example, Google Veo 3 released in May 2025, can produce lip-synced synthetic scenes and offers a frames-to-video workflow to assemble reference frames into a consistent audiovisual video. Generating reference frames with Google Whisk further accelerates production. However, these advantages introduce new uncertainties regarding content accuracy, scene continuity, realism, clinical appropriateness, and copyright/licensing.

AI-generated videos hold substantial potential for health professions education (6–8). Although these videos have limitations, they can be suitable for educational use (9,10) and can be comparable to human-produced lecture videos (11). For example, they have been used in neurosurgical education (12), to create virtual narrators (13), in nursing simulation training (14), and to simulate doctor–patient interaction processes (15). As the literature indicates, the use of AI-generated videos in medical education is limited. The existing studies either discuss the potential at a conceptual level or remain confined to talking-head video formats without reflecting the complex real world. We found no study that employs a frames-to-video technique to realistically depict scenes such as a traffic accident, an emergency department setting, or two-way dialogues.

This study provides a comparative evaluation in the PBL context by producing two scene-by-scene matched versions of a clinical scenario currently used in PBL sessions— AI-generated and human-made animated. The AI-generated video was created entirely from synthetic visual and audio assets using Veo 3 (frames-to-video) and Whisk workflows; no real persons, patient data, or identifiers were used. Quantitative and qualitative data were collected from medical students and a diverse group of experts. The study addressed two research questions: Within the PBL context, first, how do medical students evaluate an animated video (human-made) versus an AI-generated video, and which do they prefer?, and how do experts from different disciplines evaluate the educational value of the AI-generated video?

2. Methods

Study Design

This mixed-methods study, comprising quantitative and qualitative components, examined perceptions and preferences regarding two video formats of the same clinical scenario. Quantitative and qualitative data were collected from both students and experts. The quantitative data consisted of five-point Likert-type items and preference questions, whereas the qualitative data were obtained through open-ended questions. The qualitative data were analyzed using thematic content analysis (16). This study evaluated the potential use of an AI-generated video for PBL sessions.

Participants and Procedure

The study was conducted at Gazi University Faculty of Medicine, Ankara, Turkiye. Out of the approximately 1,500 students (second-year students to fifth-year students) enrolled in the faculty, 210 participated through convenience sampling. The student sample was divided into two groups, ensuring that each group included one preclinical and one clinical year (Years 2 and 3 = preclinical; Years 4 and 5 = clinical). In Group 1, Year-2 and Year-4 students first watched the animated video and then the AI-generated video; in Group 2, Year-3 and Year-5 students first watched the AI-generated video and then the animated video. This constituted a counterbalanced viewing order across year-level groupings. All students met the inclusion criteria of being enrolled in the relevant year of study, providing online informed consent, and completing the questionnaire in full; there were no missing data.

The expert sample comprised 104 participants across three subgroups: medical faculty (n=59), educators who holds PhD degree or are PhD candidate in medical education (n=25), and non-medical education sciences or informatics experts (n=20). Among the faculty subgroup, 26% (15/59) were from the basic sciences departments, 42% (25/59) from the internal sciences departments, and 32% (19/59) from the surgical sciences departments. Furthermore, 95% (56/59) had previously facilitated a PBL tutorial session. The non-medical expert subgroup included 12 experts (all working on AI) and 8 education sciences experts specialized in instructional materials development. Experts evaluated only the AI-generated video.

Materials

Both videos were audiovisual adaptations of the first session of the same written PBL scenario and have identical durations (1 minute 43 seconds). The script and scene-by-scene progression were matched across the two formats; the spoken language was Turkish and no subtitles were included. The AI-generated video, all components—including visuals and narration—were generated by using AI tools. The human-made animated video has been used since 2018 in PBL sessions at Gazi

University Faculty of Medicine (developed in-house as part of a Project without any AI involvement) and has been shown to positively influence students' knowledge, skills, and attitudes (5). Building on this institutional legacy of innovation, the same scenario was re-created as an AI-generated video, and the two versions were empirically compared.

AI-generated Video: Production Workflow

The AI-generated video was designed to be an exact, scene-by-scene match of the previously used animated video. Production employed a frames-to-video model (Google Veo 3). The reference frames were created by using Google Whisk, and the static image displaying laboratory results was produced with ChatGPT-4o. Whisk is an experimental generative media tool developed by labs.google/fx that creates synthetic images from a few example images or prompts. While preparing the reference frames with Whisk, stock images of doctor, nurse, and paramedic from Freepik were used under a royalty-free license and were refined with prompts. All visuals and audio in the video are synthetic; no real persons, patient data, or official logos were used.

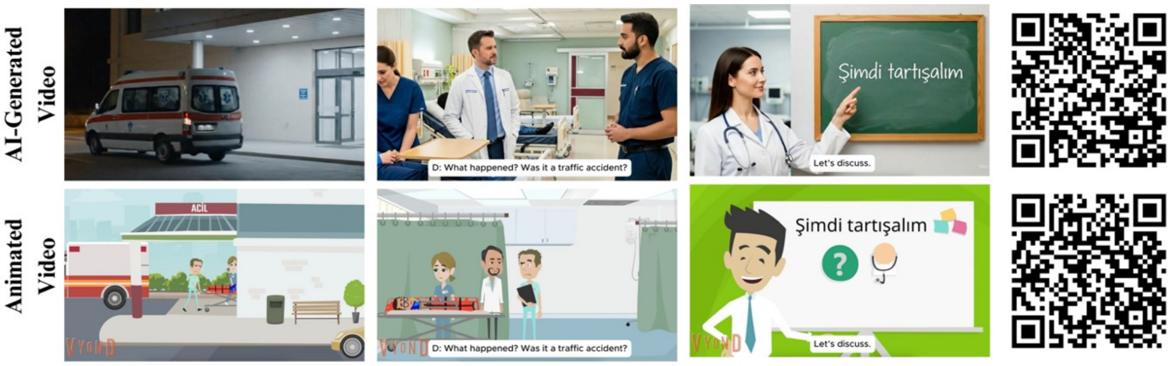

Figure 1. Scene-by-scene matching of AI-generated and animated videos. (QR codes link to the full videos)*.

*Students watched these videos but without subtitles:
AI-generated video link: https://www.youtube.com/watch?v=SAdAH-v7FXo
Animated video link: https://www.youtube.com/watch?v=LDQLr39pvgM

Scene consistency and continuity were ensured, wherever possible, by using locked-off camera angles and the frame-grab tool in the scene editor. Since Veo 3 is capable of generating only 8-second videos, the "meaningful last frame" of the mini video was captured and used as a reference frame for the next scene. To reduce misattribution of dialogue to the wrong character, the speaker's role—title, in-frame position, attire, and action—was specified explicitly in the prompts. To reinforce vocal continuity, tone-of-voice adjectives such as "gentle," "serious," or "authoritative" are used. Due to the model's tendency to generate nonsensical filler sounds in dialogs shorter than 8 seconds, the scripts were segmented into semantic units and simplified. In cases where dialogues longer than 8 seconds were not generated in full, scenes were produced in 3–4-second micro-scenes, and extraneous portions were trimmed.

When a reference frame contains a child's face, Veo 3 refused to generate the content to protect minors' rights. Accordingly, the child's face in the accident scenario was cropped out of the reference frames. However, when the body/limb proportions of the patient on the stretcher indicate a child, the Veo 3 model may generate a child's face under varying camera angles. In such cases, alternative prompts were used to keep the child out of frame, or angles that feature only adults were selected. Expressions such as 'undressing the patient' were flagged by the Veo 3 model as sexual content. To prevent this, examination narratives were rewritten in clinical, neutral terms (e.g., "preparing the patient for examination").

Across sequential video generations with Veo 3's 'Fast' option (20 credits per scene), a gradual loss of detail and sharpness was observed. To mitigate this, the "Quality" option (100 credits per scene) was used for critical shots. Additionally, when sharpness declines, the initial reference frame was reused. These measures substantially improved visual consistency. For these reasons, video editing skill was treated as a critical component in assembling AI outputs into a consistent, watchable whole. Background music and transitions were balanced through light post-production solely to enhance technical quality. No interventions were made that would alter the nature of the AI-generated content. The Veo 3 production process took approximately 40 hours and consumed a total of 2,380 AI credits (approximately USD 20). Sample prompts can be found in the supplementary material (Google Veo 3 and Google Whisk Prompt Samples).

Data Collection Instruments

Within this study, the researchers developed two instruments: an expert evaluation form and a student evaluation form. Both instruments were developed with reference to multimedia learning theory and the video-based learning literature. The instruments were also tailored to assess the specific affordances and constraints of AI-generated videos.

The item pool was built from multimedia principles (attention, duration/pace, signaling, weeding of extraneous elements, and segmenting), findings on educational-video engagement, and PBL implementation notes. The researchers selected 10 Likert-type and 1 open-ended item for experts, and for students 8 Likert-type, 4 preference-type, and 1 open-ended item. The procedure was administered via Google Forms with a target completion time of ≤ 10 minutes. Experts rated the AI-generated video on the following dimensions: visual quality; absence of distracting elements; naturalness and appropriateness of character voices (tone of voice); characters' perceived professional competence; realism of clinical settings; visual consistency; using it as an educational material; and its capacity to stimulate critical thinking and discussion in PBL sessions. Students evaluated each video separately on the following dimensions: medical appropriateness; visual quality; absence of distracting elements; naturalness of audio and appropriateness of character voice tones; perceived professional competence of the characters; realism of clinical settings; suitability for the preclinical level; and consistency of visual content. They also indicated which video they find more engaging, which evokes stronger emotion, and which they would prefer to be used in PBL sessions. The data collection instruments can be found in the supplementary material (Expert and student evaluation forms).

Statistical Analysis

Quantitative Data Analysis

Analyses were conducted in jamovi (version 2.6.45) (17). First, for the same eight Likert items pertaining to the AI-generated video, scores from students and experts were compared. Because normality assumptions were not met, the Mann–Whitney U test was applied at the item level. Second, students' scores for the animated video and the AI-generated video on the same eight Likert items were compared, using the Wilcoxon signed-rank test for each item. Descriptive statistics were used to summarize items not included in the two primary comparisons.

Qualitative Data Analysis

Experts' open-ended responses were analyzed using thematic analysis in accordance with AMEE Guide No. 131 (16). The analysis proceeded in six steps: familiarization with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and reporting. In the familiarization phase, technical quality, content appropriateness, pedagogical features, and others stand out, and recurring motifs include realism, liking, pronunciation, the emergency department context, and suitability for PBL. Next, initial codes were generated using data-proximal labels (e.g., "rapid transitions," "pronunciation issue," "realistic," "suitable for PBL sessions"). The codes were then clustered by conceptual proximity into four themes: "technical details," "medical content," "pedagogical elements", and "others". During the theme review stage, consistency and fit with the dataset were checked; overlapping subthemes were merged (e.g., flow/tempo; attention–interest–retention). Notes that could not be classified were collated under "general recommendations," and expressions of praise/thanks were grouped under "overall appreciation" and were placed under the "others" theme. Thus, a thematic structure was established that approached the data with minimal a priori assumptions and was aligned with the multimedia nature of the material and the PBL context. All stages were conducted by the researchers through discussions and consensus. Representative quotations for the identified subthemes are presented in the Results section.

Ethical Considerations

The study was approved by the Gazi University Ethics Committee (Decision No. 2025-1432). Online informed consent was obtained from all participants, and data were anonymized and used solely for research purposes. The datasets generated and analyzed during the current study are available in the Zenodo repository at https://zenodo.org/records/16926141 .

3. Results

The student sample was n=210: Year-2, 18.6% (n=39); Year-3, 41.4% (n=87); Year-4, 15.7% (n=33); and Year-5, 24.3% (n=51). Gender distribution was female 58.1% (n=122), male 40.5% (n=85), and prefer not to say 1.4% (n=3). Among students, 68.6% (144/210) reported believing that AI technologies will radically transform healthcare, whereas 31.4% (66/210) did not. Paid subscriptions to AI services such as ChatGPT were reported by 42.3% of experts (44/104) and 15.2% of students (32/210).

Questionnaire Reliability

The Likert-type scale used to evaluate the AI-generated video demonstrated excellent internal consistency, with Cronbach's α = .922.

Quantitative Results

The study's quantitative results are presented in Table 1. Compared with students, experts' ratings of the AI-generated video were statistically significantly higher on three items (M2, M3, M8). Although experts' ratings exceeded students' on all other items, these differences were not statistically significant. Students' ratings of the AI-generated video were higher than those of the animated video across all items.

In the preference items, students clearly favored the AI-generated video: "engagement" (83.8%), "emotional impact" (81.0%), and "preference for use in PBL sessions" (79.0%) were all in favor of the AI-generated video.

For the AI-generated video, experts reported Mean=4.38, Median=5 or "well structured as an instructional material," and M=4.26, Median=4.5 for "stimulates critical thinking/discussion." indicating strong endorsement of the video's pedagogical structuring and its capacity to stimulate discussion.

Qualitative Results

The study's qualitative results are presented in Table 2 with representative quotations for each subtheme. Table 2 shows that the most prominent subtheme were realism (n=45) under technical details; suitability for PBL (n=26) and level of engagement (n=25) under pedagogical elements; the recommendation to improve the scenario (n=26) under medical content and overall appreciation (n=48) under others. Several subthemes emerged as areas for improvement, such as, scenario improvement (n=26), inappropriate body language (n=12), inappropriate emergency department atmosphere (n=11). Because both videos were based on the same scenario, issues noted under "scenario improvement," "request for additional scenes," and "patient preparation for physical examination" were not modifiable within this study. Rapid transitions were flagged as an improvement area to be handled by pausing at designated moments during PBL sessions and therefore did not pose an implementation problem. Flow continuity and pronunciation issues stemmed from Veo 3's technical constraints, whereas the items under inappropriate body language, an inappropriate emergency department setting, and voice-over for written scenes are problems that can be addressed.

Table 1. Items and quantitative results

Item		Experts						Students (n=210)							
		AI- generated video						AI- generated video			Animation (human-made) video				
		n	Mean	Mdn	SD	Effect Size	p	Mean	Mdn	SD	Mean	Mdn	SD	Effect Size	p
M1	The video content is medically appropriate	84	4.31	5.00	0.918	-0.042	0.530	4.27	4.00	0.874	3.93	4.00	1.074	0.485	<.001
M2	The video's visual quality is satisfactory	104	4.63	5.00	0.727	-0.230	<.001	4.24	5.00	0.985	3.16	3.00	1.133	0.707	<.001
M3	The video is structured to avoid distraction	104	4.42	5.00	0.952	-0.328	<.001	3.75	4.00	1.27	3.44	4.00	1.26	0.199	0.026
M4	The audio is natural, and character voices tones are appropriate	104	4.09	4.00	0.956	-0.105	0.056	3.83	4.00	1.17	3.20	3.00	1.19	0.412	<.001
M5	The characters appear professionally competent	104	4.30	4.00	0.823	-0.070	0.092	4.10	4.00	1.07	3.30	3.00	1.14	0.616	<.001
M6	The clinical settings depicted are realistic	84	4.29	5.00	0.939	0.0145	0.781	4.25	5.00	1.06	2.86	3.00	1.19	0.818	<.001
M7	The video is appropriate for the preclinical level	84	4.49	5.00	0.768	-0.073	0.156	4.32	5.00	0.997	3.99	4.00	1.137	0.618	<.001
M8	The visual content is consistent	104	4.44	5.00	0.680	-0.173	0.001	4.06	4.00	1.06	3.70	4.00	1.14	0.385	<.001

Notes: Likert-type items were rated on a 1–5 scale (higher = stronger agreement). Left p column: **Mann–Whitney U** comparing **experts vs. students** on the **AI-generated video** only. Right p column: **Wilcoxon signed-rank** comparing **AI vs. animated video** within students.

Table 2. Qualitative results

Theme	Subtheme	n	Representative quotation
Technical details	Realism	45	"The environment and people were very realistic."
	Perceived quality improvement over animation	18	"Compared to animations, I think it is more realistic and motivating."
	Rapid transitions	10	"The test results pass by too quickly."
	Pronunciation issues	9	"The word -ambulance- is not pronounced correctly."
	Voice-over for written scenes	9	"Rather than showing the questions on the board, voicing them may be more appropriate."
	Request for additional scenes	9	"Scenes of initial post-accident intervention and transfer to the emergency department should be shown."
	Presentation of findings	6	"Examination findings should be shown more clearly."
	Flow continuity	4	"There are breaks in the visual flow."
Medical content	Scenario improvement	26	"The scenarios will get better as they are discussed with experts."
	Inappropriate emergency department atmosphere	11	"For an emergency department, the atmosphere was calm and slow."
	Medical terminology	4	"The term examination should be used instead of care."
	Patient preparation for physical examination	3	"How will the nurse prepare the patient for examination?"
Pedagogical elements	Suitability for PBL	26	"It is a good material for PBL sessions."
	Level of engagement	25	"It will capture students' interest."
	Inappropriate body language	12	"The doctor didn't need to keep their hands in their pockets."
	Communication skills	8	"I found the doctor–paramedic dialogue successful."
Others	Overall appreciation	48	"I liked it very much."
	General recommendation	23	"It should be improved a bit further."
	Improvement over time	8	"I believe these videos will further improve over time."
	Alignment with the era	6	"In the age of AI, proceeding this way is the right approach."
	Waste prevention	2	"No paper waste."

4. Discussion

In this study, students clearly preferred the AI-generated video over the human-made animated video on every aspect. A broad panel of experts from medical faculty, educators who hold PhD degree in medical education, and non-medical education sciences and informatics experts also rated the video positively for educational appropriateness, visual consistency, and its capacity to stimulate discussion in PBL sessions. One expert remarked, "It is hard to believe it was prepared entirely with AI. The environment and characters are very realistic."

The literature includes studies that discuss the potential of AI-generated videos at a conceptual level or rely on the comparison of traditional lectures and or "talking-head" AI (9). In contrast, within the PBL context, this study implements a clinical flow featuring multiple actors, cross-setting transitions, and dynamic scenes. To our knowledge, the current study represents the first reported application of a frames-to-video (Veo 3) model in medical education—and one of the earliest within the broader education literature. In addition to demonstrating this novel application, the quality of the AI-generated videos was evaluated by students and a diverse group of experts.

The frames-to-video model used in this study offers several advantages. Text-to-video models can be prone to technical and clinical errors (8). By contrast, in a frames-to-video workflow, pre-generating context-appropriate reference frames (via Whisk) can reduce such errors. Unsuitable or erroneous shots can be regenerated iteratively until the desired result is achieved. For talking-head scenarios, there is evidence that human-produced videos are, albeit slightly, superior to AI-generated videos (18). Our study shows that videos far more complex and dynamic than talking-head formats can be generated by Veo 3 with relative ease. While some argue that it is still premature for AI-generated videos to fully replace conventional methods (10), both students and experts in this study prefer and endorse the AI-generated video over the human-made animation. As one student noted, “I believe the AI-generated video will bring a new breath to PBL sessions.” With advances in generative AI, AI-generated content was expected to resemble traditional content in the near future (11). This expectation appears to be materializing, and at a faster pace than anticipated.

From an implementation standpoint, AI-generated videos can be produced economically (the present video cost approximately 20 USD) and in a democratized manner (by anyone who can articulate the target content), without requiring advanced technical expertise. Because all visual and audio assets are synthetic, a realistic clinical context can be constructed without using personal data, thereby helping to reduce copyright and privacy risks. However, producing a 1 min 43 s video required ~40 hours including operator learning time, which underscores a non-trivial upfront investment. With operator experience and model improvements, this production time is expected to decrease. Once these conditions are met, AI-generated videos can be deployed not only in PBL sessions but also in any teaching and learning activities. As one expert in the study observed, “The ability to prepare this type of educational material with AI is promising from an educational standpoint,” pointing to broader use cases. For example, an educator who wishes to run a case-based discussion in class can, in principle, generate a relevant short video with a brief text prompt, thereby creating a rich environment for discussion. However, current AI technology has not yet matured to fully support this level. While the frames-to-video model reduces errors, the generated video still presents limitations—pronunciation, body language, micro-continuity, and within-scene synchronization. Moreover, its ability to depict complex medical symptoms or findings remains uncertain. Generative AI may hallucinate (19); therefore, all generated videos must undergo expert review prior to use.

Strengths and Limitations

The study’s strengths include obtaining evaluations from students across different years of study; sampling both students who have previously encountered the same scenario and those who have not; including a large, multidisciplinary expert panel; and employing a newly released frames-to-video model.

Limitations of this study include the incomplete coverage of the PBL scenario (only a portion of the first session was included), the absence of an implemented PBL activity to assess potential effects on educational outcomes, and the lack of an a priori sample size or power calculation although there was a strong diversity in terms of participants. Although the counterbalanced design was employed to minimize potential order effects, residual learning or carryover effects cannot be fully excluded. Future studies should examine this possibility more directly to ensure that observed differences are not influenced by sequence-related factors. In addition, scenario coverage, clinical complexity, and symptom realism were not directly assessed, which may limit the generalizability of the findings to other clinical contexts. In addition, the output quality remains operator-dependent in the current generation of models.

5. Conclusions

- Students preferred the AI-generated video over the animation created by humans, and experts endorsed its educational suitability, visual consistency, and discussion-triggering capacity. In contrast to the predominantly potential- or talking-head–oriented examples in the literature on AI-generated video, a multi-actor, cross-setting, dynamic clinical process was implemented within the PBL context using a frames-to-video (Veo 3) model.

- The frames-to-video model—together with Whisk-generated reference frames and iterative generation—facilitated the creation of context-appropriate scenes by reducing the medical/technical errors common in text-to-video.

- Although small advantages for human-produced videos have been reported in talking-head settings, AI-generated video has been found to be reliable and preferred even in more complex, dynamic narratives. By virtue of fully synthetic assets that reduce copyright and privacy risks—and by offering a low-cost, accessible, and rapid production route—the method can be disseminated beyond PBL sessions to address other needs in medical education.

- Nevertheless, future studies that implement complete PBL tutorials and directly compare learning outcomes are needed. The present findings provide a helpful foundation upon which such investigations can be built.

Funding: No funding received for this study.

Declaration of conflict of interest: The authors declare that they have no conflict of interest.

Data availability: The datasets generated and analyzed during the current study are available in the Zenodo repository at https://zenodo.org/records/16926141 .

Author contributions: Conceptualization: ABK, YSK, ÖC, IİB; Data collection: ABK, YSK, ÖC; Formal analysis: ABK, YSK; Methodology: ABK, YSK, ÖC, IİB; Roles/Writing: - original draft: ABK, YSK; and Writing - review & editing: ABK, YSK, ÖC, IİB.

6. References.

1. Davis MH, Harden RM. Amee medical education guide no. 15: problem-based learning: a practical guide. *Medical Teacher*. **1999,** 21, 130–40. https://doi.org/10.1080/01421599979743.
2. Basu Roy R, McMahon GT. Video-based cases disrupt deep critical thinking in problem-based learning. *Medical Education*. **2012**, 46, 426–35. https://doi.org/10.1111/j.1365-2923.2011.04197.x
3. De Leng B, Dolmans D, Van De Wiel M, et al. How video cases should be used as authentic stimuli in problem-based medical education. *Med Educ*. **2007**, 41, 181–8. https://doi.org/10.1111/j.1365-2929.2006.02671.x.
4. Chan LK, Patil NG, Chen JY, et al. Advantages of video trigger in problem-based learning. *Medical Teacher*. **2010**, 32, 760–5. https://doi.org/10.3109/01421591003686260.
5. Budakoğlu Iİ, Coşkun Ö, Özeke V. e-PBL with multimedia animations: a design-based research. *BMC Med Educ*. **2023,** 23, 338. https://doi.org/10.1186/s12909-023-04298-x.
6. Dağcı M, Dost A, Çam F. Potential impacts of ai-generated videos on nursing care. *Nurse Educator*. **2025**, 50(5), E254-E258, https://doi.org/10.1097/NNE.0000000000001880.
7. Liu Y, Zhang K, Li Y, et al. Sora: a review on background, technology, limitations, and opportunities of large vision models. *arXiv preprint*. **2024,** arXiv:2402.17177 https://doi.org/10.48550/arXiv.2402.17177.
8. Waisberg E, Ong J, Masalkhi M, et al. Concerns with openai's sora in medicine. *Ann Biomed Eng*. **2024**, 52, 1932–4. https://doi.org/10.1007/s10439-024-03505-0.
9. Contreras I, Hossfeld S, De Boer K, et al. Revolutionising faculty development and continuing medical education through ai-generated videos. *Journal of CME*. **2024**, 13, 2434322. https://doi.org/10.1080/28338073.2024.2434322.
10. Shu J, Kou H, Zhang J, et al. AI-generated instructional videos: a systematic review of learning impacts, applications, and research perspectives. **2025** *7th International Conference on Computer Science and Technologies in Education (CSTE)*. Wuhan, China: IEEE 2025:877–81. 10.1109/CSTE64638.2025.11091997.
11. Netland T, Von Dzengelevski O, Tesch K, et al. Comparing human-made and ai-generated teaching videos: an experimental study on learning effects. *Computers & Education*. **2025**, 224, 105164. https://doi.org/10.1016/j.compedu.2024.105164.

12. Lawson McLean A, Gutiérrez Pineda F. Application of transformer architectures in generative video modeling for neurosurgical education. *Int J CARS*. **2024,** 20, 797–805. https://doi.org/10.1007/s11548-024-03266-0.
13. Yeung JWY, Tang FWK, Cheng HY. Tips for incorporating ai-generated video in disaster nursing education. *Nurse Educator*. **2025,** 50, E151–E151. https://doi.org/10.1097/NNE.0000000000001840.
14. Chan MMK, Wan AWH, Cheung DSK, et al. Integration of artificial intelligence in nursing simulation education: a scoping review. *Nurse Educator*. **2025**, 50, 195–200. https://doi.org/10.1097/NNE.0000000000001851.
15. Sardesai N, Russo P, Martin J, et al. Utilizing generative conversational artificial intelligence to create simulated patient encounters: a pilot study for anaesthesia training. *Postgraduate Medical Journal*. **2024**, 100, 237–41. https://doi.org/10.1093/postmj/qgad137.
16. Kiger ME, Varpio L. Thematic analysis of qualitative data: Amee guide no. 131. *Medical Teacher*. **2020,** 42, 846–54. https://doi.org/10.1080/0142159X.2020.1755030.
17. Leppink J. Data analysis: more expensive does not imply better. *Rev Esp Edu Med*. **2022**, 3. https://doi.org/10.6018/edumed.516731.
18. Arkün-Kocadere S, Çağlar Özhan Ş. Video lectures with ai-generated instructors: low video engagement, same performance as human instructors. *IRRODL*. **2024,** 25:350–69. https://doi.org/10.19173/irrodl.v25i3.7815.
19. Masters K. Medical teacher ’s first chatgpt’s referencing hallucinations: lessons for editors, reviewers, and teachers. *Medical Teacher*, **2023**; 45, 673–5. https://doi.org/10.1080/0142159X.2023.2208731.