George *et al. BMC Medical Education* (2025) 25:1626
https://doi.org/10.1186/s12909-025-08177-5

BMC Medical Education

RESEARCH Open Access

Evaluation of the impact of an online introductory paediatric physical examination skills course on students' OSCE performance

Ann George[1*], Duane Blaauw[2], Ziyaad Dangor[3,4] and Sanjay G. Lala[5]

Abstract

Background Videos demonstrating clinical skills are considered powerful educational tools. However, there is limited evidence on the effectiveness of comprehensive multimedia skills courses on knowledge and skills acquisition. We used cognitive load theory to develop an online course to introduce medical students to core paediatric physical examination concepts and skills, within a resource-constrained learning environment with high student-to-clinical teacher ratios. This study evaluated the impact and acceptance of the online course.

Methods We used an observational quasi-experimental analysis for the impact evaluation. The main outcome was students' marks for the paediatrics station in the final paper-based observed structured clinical examination (OSCE). We measured the frequency of viewing core course content and used the median number of views to divide 351 fourth-year students into low and high users. Propensity score matching analysis was used to adjust for potential selection biases of course use. Low- and high-users were matched on gender, age, type of course entry, previous academic performance, and OSCE group before calculating the average treatment effect (ATE). The matched ATE was also calculated for different sub-groups of the class. A voluntary, online, cross-sectional survey about the course was administered post-OSCE. Closed-ended survey responses were analysed using descriptive statistics, and open-ended questions using content analysis.

Results Increased course usage significantly increased performance in the paediatrics component of the OSCE: high course users scored an average of 7.0% points [95%CI: 3.8–10.2] higher. The impact of the course was greatest for the middle 60% of the class, in terms of previous academic performance, with a matched ATE of 8.2% [95%CI: 4.3–12.2]. The survey response rate was 74.9%. The respondents expressed positive perceptions about the overall value of the course, including the clearly presented content that fostered an appreciation for paediatrics and prepared them for hospital rotations and the OSCE.

Conclusions This study demonstrates that an online paediatric physical examination skills course can significantly improve students' examination performance, particularly for those in the middle academic performance range. The findings support the integration of multimedia clinical skills courses in medical curricula, especially in resource-constrained settings.

*Correspondence:
Ann George
Ann.George@wits.ac.za
Full list of author information is available at the end of the article

Keywords Impact evaluation, Propensity score matching analysis, Online clinical skills courses (health professions), Clinical teaching (health professions), Paediatrics, Medical students

Background

Videos demonstrating clinical skills are considered "incredibly powerful" tools to augment bedside teaching (p. 777) [1]. Bedside teaching is regarded as the optimal method for teaching clinical skills [2] but it may be affected by contextual conditions, including variation across different instructors [3, 4] and the time instructors have available for bedside teaching beyond their clinical duties [5]. Bedside teaching may also be affected by teaching conditions, such as physical space constraints [6]. For example, recent evidence suggested that bedside teaching works better in groups with lower student-to-teacher ratios [7].

We previously compared video teaching of undergraduate paediatric students to bedside teaching in an under-resourced setting with a high student-to-teacher ratio [8]. We found that video teaching was not inferior, within a 10% margin, in terms of performance in an observed structured clinical examination (OSCE) [8]. These findings contributed to the development of an online multimedia paediatric physical examination skills (PPES) course. The overall purpose of the PPES course was to introduce undergraduate medical students to the paediatric examination. By the definition of Song et al. [9], the course qualifies as a 'complex multimedia environment'–a learning environment providing multiple representations of information, such as text, sound, images, animation, and video. The PPES course included several types of clinical recordings: skills demonstration videos (narrated and unnarrated), audio recordings, clinical photographs, and infographics. Some of the clinical photographs were derived from the videos, and others were independent of the videos. The distinctive features of the course included the use of local South African patients and instructors in clinical settings in the videos, the use of unnarrated videos for student self-assessment of knowledge, coupled with narrated and annotated feedback videos, and specially developed interactive infographics highlighting clinical examination skills that are difficult to learn.

According to Sweller, teaching with technology must fit into the human cognitive architecture and capacity for learning [10]. Cognitive load theory proposes empirically derived instructional design principles to ameliorate the limited processing capacity of working memory [10, 11]. The practical implications of cognitive load theory are to minimise extraneous processing and preserve working memory capacity for learning through deliberate instructional design. For example, extraneous processing can be minimised by avoiding non-essential and redundant information and using temporal and spatial contiguity [12]. Cognitive load theory is most beneficial for high-complexity material where many elements must be learned simultaneously [10]. The cognitive load principles applied to the PPES course are described in the Methods section.

Our evaluation of this course was motivated by three primary considerations: (1) the critical importance of evaluating online teaching within its specific implementation context [13], (2) the limited number of evaluations of online paediatric clinical skills courses that include different types of clinical recordings, and (3) methodological limitations in previous comparative studies of online learning components and video-based studies [13]. In addition to providing information to improve our course, this study adds to the evidence base for online paediatric clinical skills teaching. Our findings are relevant to course designers, paediatricians, and other clinical teachers and researchers.

Methods

Study setting

The study was conducted at the University of the Witwatersrand (Wits University) in Johannesburg, South Africa, in 2022. The University offers a six-year undergraduate-entry medical degree that merges with the third year of study for students pursuing a four-year postgraduate degree, known as the Graduate Entry Medical Program (GEMP). GEMP I and II are pre-clinical years. The GEMP II curriculum includes an introductory clinical skills programme, which is assessed at the end of the year by a paper-based OSCE consisting of multiple discipline-specific, five-minute stations. The knowledge-based assessment, focusing on cognitive skills, is less resource-intensive and is regarded as an appropriate introduction to the OSCE format for pre-clinical students.

The paediatric physical examination skills online course

The Paediatric Physical Examination Skills (PPES) course was developed by paediatricians, a health professions educationalist, curriculum designers, and instructional designers. The aim of the course at the GEMP II level was to introduce the basic paediatric examination and augment the lecture-based and bedside teaching. The course was hosted on the University's learning management system, Ulwazi, the University's instance of Canvas. Ulwazi was introduced in January 2021.

The course consists of five modules: General Examination, Vital Signs Measurement, Nutritional Assessment, Respiratory Distress and Noisy Breathing

Table 1 Design principles used in the course*

	Principle	Application in the PPES course
Reduce extraneous processing	Omit extraneous material in slides or scripts	Captions consisted of phrases and not whole sentences.
	Use cues to highlight information essential for learning	Symbols and labels were used to highlight specific clinical features and examination techniques.
	Do not use captions that repeat the narration	The videos do not have subtitles.
	Present text near the corresponding part of the graphics	Text is placed close to the corresponding parts of the graphics.
	Present corresponding words and pictures simultaneously rather than successively.	The narration is synchronised with the images.
Manage essential processing	Present material in manageable user-controlled segments rather than a continuous unit.	The course consisted of structured modules.
	Present words as spoken text	Narration was used where applicable.
Foster generative processing	Voice principle	South African paediatricians narrated the videos.
	Insert generative learning activities.	At pauses in the video, the learner is asked to type in a brief explanation of the preceding segment.
	Worked-out examples principle	Clinical examination techniques were demonstrated in a sequential manner.
	Self-explanation principle	Students were asked to record their unique observations of clinical scenarios.

*Based on Mayer [14] and van Merriënboer and Sweller [15]

Assessment, and Assessment of Hydration and Perfusion. The sequence of the clinical recordings approximates the first two steps of Peyton's four-step teaching approach (demonstration, deconstruction, comprehension, and performance) [7]. The students initially watch a non-narrated video. They are then asked to record their observations (testing video). The testing videos are followed by feedback videos that provide explanations (either in narrative or text form). The video length ranges from 30 s to 5 min. The videos are followed by clinical photographs and specially designed infographics explaining complex concepts. The course also includes three types of additional materials: instructions for bedside clinical skills drills, multiple-choice quizzes, and downloadable PDFs with further information and exercises. Table 1 outlines the design principles applied in the course based on recommendations by Mayer [14] and van Merriënboer and Sweller [15] to reduce the cognitive load of multimedia educational materials.

Table 2 Structure of the paediatric OSCE station

Group	Testing Date	Time	Overview of the OSCE question
1 and 2	13 October	Morning/Afternoon	Anthropometry interpretation (height/BMI Z-scores), BP cuff selection, manual BP procedure
3	25 October	Morning	Growth measurement techniques: height measurement (stadiometer), head circumference, growth chart plotting and interpretation
4	26 October	Morning	Growth parameter interpretation, MUAC measurement, nutritional classification, oedema assessment
5	27 October	Morning	Growth interpretation, respiratory rate measurement, respiratory distress signs, and abnormal breath sounds

The GEMP II OSCE

The 2022 GEMP II cohort was divided into five groups for the three-and-a-half-hour paper-based OSCE[1]. Table 2 shows the timing and structure of the OSCE, which was conducted over five sessions on four days. Two groups of students took the OSCE in morning and afternoon sessions on 13 October. The paediatric station, comprising four questions, was one of 12 five-minute stations. The paediatric questions asked students to either describe an examination technique, interpret anthropometric parameters, or identify and interpret physical signs after viewing photographs or video recordings of ill children. No communication was possible between the two groups of students participating in the OSCE sessions on 13 October, as they were asked the same question (Table 2).

Study design

This observational study consisted of a quasi-experimental, non-equivalent group evaluation design and a descriptive, cross-sectional, voluntary online survey. All 351 GEMP II students at Wits University in 2022 were eligible to participate in the study.

[1] The OSCE questions were answered on paper.

Impact evaluation

The following datasets were used for the evaluation. The GEMP II class data ($n = 351$) provided demographic and background information. Course activity reports from the University's Business Intelligence Services and course-related analytic data from Ulwazi ($n = 347$) were used to determine the total number of views of core module content pages. We also had access to the cohort's promotion marks in the previous year (2021) ($n = 328$) and their final OSCE marks ($n = 347$).

The outcome variable for the impact evaluation was students' marks for the paediatrics station in their final OSCE in 2022. The predictive variable was activity on the course, measured as the total number of views of core module content ("page views"). We used the median number of page views to divide the 351 students into low and high course users.

We used a propensity score matching analysis to adjust for potential selection biases of course use [16]. Propensity score matching and regression adjustment provide a more accurate estimate of causal effects than using an unmatched sample [17]. Propensity scores calculate the probability of a participant's treatment group as a function of the measured confounders for that participant [18]. Low users and high users were matched on gender, age, type of course entry (non-graduate or graduate), previous academic performance (2021-year marks), and OSCE group before calculating the average treatment effect (ATE) using the kmatch ado program in Stata (version 18). In this case, the ATE is the adjusted difference in the mean paediatric OSCE marks between the low and high users [16, 19]. Propensity score matching analysis was also used to calculate the ATE for different subgroups in the cohort. Additional details of the propensity score analysis are included in Supplementary Material 1.

Survey

The online questionnaire developed for the study (Supplementary Material 2) was administered in REDCap. The questionnaire was piloted with 55 GEMP III students from the two final paediatric rotations in 2021 and modified based on their feedback. The final questionnaire consisted of six sections with closed- and open-ended questions about students' use of the modules, the general usefulness and usability of the course, access and navigation difficulties, socio-demographic information, and the hospital for their paediatric rotation. As recommended by our Human Research Ethics Committee, we offered reimbursement for the time spent completing the survey in the form of a meal voucher redeemable at the Medical School cafeteria.

An individualised survey link was emailed to students after they completed the OSCE in November 2022. The first page of the survey included participant information and requested informed consent before participants proceeded with the survey (Supplementary Material 2). All survey items had to be answered. The survey remained open for two months, with reminder emails at approximately two-week intervals.

The closed-ended responses were analysed using descriptive statistics in Stata. The textual open-ended responses were analysed using content analysis to systematically categorise the student responses and identify patterns in their usage and experience of the online course [20]. Content analysis was deemed appropriate given the discrete nature and large number of open-ended survey responses.

The following questions (Supplementary Material 2) were analysed using content analysis: the three open-ended responses in Sect. 2 (general usefulness of the online course) and the three in Sect. 3 (accessibility of the course). The responses to these questions were transferred to Microsoft Excel and coded individually as follows. Each response was examined for meaning units (words or phrases relevant to the research). One response could have more than one meaning unit. Each meaning unit was assigned a code and a frequency of 1. The codes were renamed and refined iteratively by two experienced coders, who held discussions to resolve any contrasting opinions until a comprehensive coding system was developed. Once all the responses for a question had been coded, the codes were grouped into subcategories, then categories, and the frequencies were counted. The results for the three open-ended responses in Sect. 2 (Supplementary Material 2) are displayed graphically to facilitate pattern identification [21]. Selected participant quotations have been used as examples of categories or subcategories [22].

The Wits University Human Research Ethics Committee (Medical) approved the study (Ethics Certificate M210663). We also obtained permission for the study from the Deputy Registrar, the Head of the School of Clinical Medicine, and the Unit for Undergraduate Medical Education.

Results

Figure 1 shows the sequence of the study. The impact evaluation was conducted for the entire 2022 cohort, excluding four students who had not taken the OSCE. The response rate for the post-OSCE survey was 74.9% (260 completed responses out of 347 total responses).

Demographics and background information

Table 3 summarises the demographic and background information for the entire 2022 cohort as well as the sub-group of survey respondents. The majority of the cohort were female (65.2%) and non-graduate admissions (58.1%), with a mean age of 23. White students were the

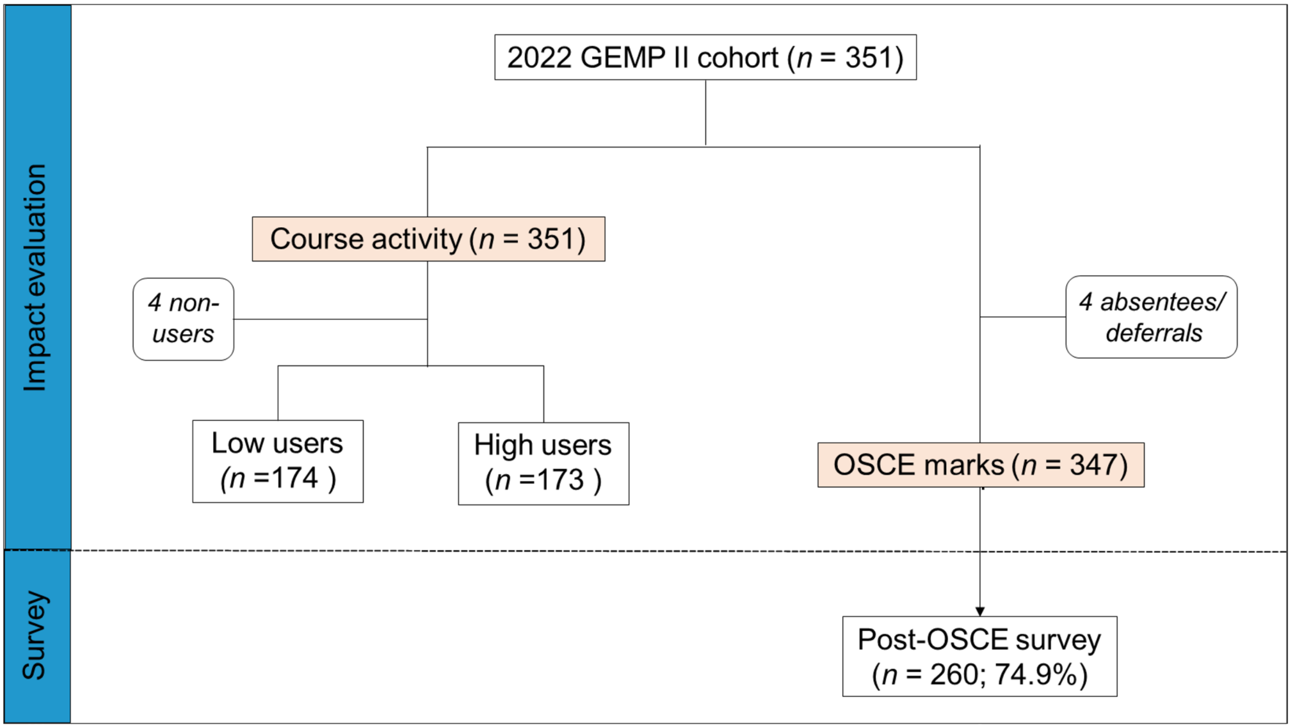

Fig. 1 Study flow diagram

largest racial group (36.3%). Most students were of urban origin (84.5%) and had attended private schools or the more affluent quintiles 4 and 5 public schools (88.6%). South African public schools are classified using a quintile system based on socioeconomic indicators, including the average income and unemployment rates in a school's geographical location [23, 24].

Comparing the respondents to the non-respondents, the former consisted of significantly more females than males ($p = 0.004$) (Table 3). There were no significant differences between the respondents and non-respondents in terms of age, background factors, and OSCE group.

Impact evaluation

Table 4 presents the number of PPES course page views for the high- and low-user groups. The total number of views ranged from 0 to 1219, with a mean of 270.9 ± 216.7 and a median of 195.

The overall mean OSCE score for the cohort was 80.0% [95%CI: 79.4–80.7]. The mean score for the pediatric station was 79.8% [95% CI: 78.3–81.2]. Figure 2 compares the median scores for low (77.8%) and high (84.6%) users of the PPES course. The rightward shift of the high-user curve shows the effect of course usage on the OSCE marks. Most students in both groups scored above the OSCE pass mark of 60%, though the failure rate was significantly lower for high users than for low users of the PPES course in the unmatched analysis ($X^2 = 3.9$, $p = 0.048$).

Selection bias occurs when participants are not randomly assigned to treatment groups, which can lead to biased estimates of treatment effects [25]. In this study, attributes such as diligence, self-regulation, and motivation may be related to both higher use of the course and OSCE performance, overestimating the impact of course activity on performance. Logistic regression analysis of the predictors of course use showed that females were 1.7 times more likely (p = 0.029) to be in the high user group, as were higher academic achievers in the previous year–each 10% increase in the cohorts' 2021 year-mark increased (their odds of being in the top half of the user group by 1.54 (p = 0.030). Age (p = 0.593) and type of course entry (GEMP or non-GEMP) (p = 0.119) did not significantly affect the likelihood of being in the high user group.

The analysis of impact after propensity score matching (Table 5) confirmed that increased PPES course usage significantly affected performance in the paediatrics component of the OSCE: high course users scored an average of 7.0 [95%CI: 3.8–10.2] percentage points higher in the OSCE than low users. This equates to a Cohen's d of 0.249, indicating a small effect size.

The sub-group analysis (Table 5) suggests that the course had the greatest impact on the middle 60% of the class in terms of previous academic performance, with a matched ATE of 8.2% [95% CI: 4.3–12.2]. The impact did not appear statistically significant for students from schools in quintiles 1 and 5, rural areas, or those from

Table 3 Demographic and background information

	Entire cohort ($N=351$)	Survey respondents ($n=260$)	*p*-value (Comparing respondents to non-espondents)
Gender: Female [*n* (%)]	229 (65.2%)	181 (69.6%)	$X^2=8.458$, $p=0.004$
Age [mean ± SD]	22.9 ± 2.5	22.9 ± 2.5	$t=0.187$, $p=0.852$
Race*: [*n* (%)]			
Black	119 (33.9%)	91 (35.0%)	Fisher's exact, $p=0.095$
Coloured	19 (5.4%)	11 (4.2%)	
Indian/Asian	86 (24.5%)	58 (22.3%)	
White	127 (36.2%)	100 (38.5%)	
Nationality: South African [*n* (%)]	340 (96.9%)	252 (96.9%)	Fisher's exact, $p=1.000$
Graduate entry [*n* (%)	147 (41.9%)	113 (43.5%)	$X^2=1.030$, $p=0.310$
Area of origin: Rural	53 (15.5%)	36 (14.2%)	$X^2=1.320$, $p=0.250$
School socioeconomic status** [*n* (%)]			
Public quintile 1–3	39 (11.4%)	28 (11.0%)	$X^2=0.191$, $p=0.909$
Public quintile 4–5	158 (46.2%)	117 (46.1%)	
Private	145 (42.4%)	109 (42.9%)	
OSCE group: [*n* (%)]			
Group A	73 (20.8%)	52 (20.0%)	$X^2=0.2.85$, $p=0.579$
Group B	70 (19.9%)	54 (20.8%)	
Group C	71 (20.2%)	57 (21.9%)	
Group D	66 (18.8%)	47 (18.1%)	
Group E	71 (20.2%)	50 (19.2%)	

*The racial classifications used were those introduced during the apartheid era (1947–1994) according to the South African Population Registration Act (No. 30 of 1950). These classifications are still used to assist the South African government in redressing injustices against previously disadvantaged population groups

**Quintile 1 schools are in the poorest areas, and Quintile 5 in the wealthiest [24]

poorer feeder schools. However, the small sample sizes limit the power of these sub-group analyses.

Survey results

Most respondents viewed all five modules, with high viewership (> 90%) across all five modules; all respondents viewed the General Examination module (Table 6). The modular content was generally perceived as clearly presented (> 85% of respondents), with the Nutritional Assessment module reported as having the least clear content presentation. Most respondents agreed that the learning activities in the course helped them engage with the learning material (> 80% across the five modules) (Table 6). However, fewer respondents agreed that the learning activities helped prepare them for the OSCE (69%–78% across the five modules) (Table 6).

Most respondents indicated that the course had stimulated their interest in paediatrics (79.6%), regarded the video demonstrations as more helpful than the textbook (86.5%), and felt that the videos provided more structure than bedside teaching (74.2%). A slightly lower proportion agreed that it made them feel confident about the OSCE (62.3%).

Figure 3 displays the results of the question asking respondents whether the course was worth their while. Positive and negative perceptions are represented as positive and negative numbers in Fig. 3. The overwhelming perception was the overall value of the course ($n=321$). There were 12 subcategories of reasons for this perception–the primary reason being that 'the course provided an introduction to paediatrics' ($n=58$). One respondent stated:

> *"The course gave me the opportunity to learn a lot regarding paediatrics in a short period of time since Paed[iatric]s was not part of our clinical skills content".*

Another reason contributing to the overall value of the course ($n=321$) was the feeling that 'the course prepared them for their hospital rotations' ($n=51$), as explained by two respondents:

> *"Overall, the course was really useful in helping me to prepare for clinical visits, for the OSCE, and just gaining a better knowledge and appreciation for paediatrics overall. I was able to engage with doctors at the hospital more confidently".*

and

> *"It provided a structured and useful base for our clinical methods session, where we could build on that knowledge via bedside teaching".*

Table 4 Number of page views for high and low users ($n=347$*)

User group	N	Mean	SD	Min	Max	p50	IQR**
Low users	174	110.2	40.4	0	195	109	34
High users	173	432.5	201.1	198	1219	360	256
Total	347	270.9	216.7	0	1219	195	251

*Four students did not use the course (see Fig. 1)

**Interquartile range

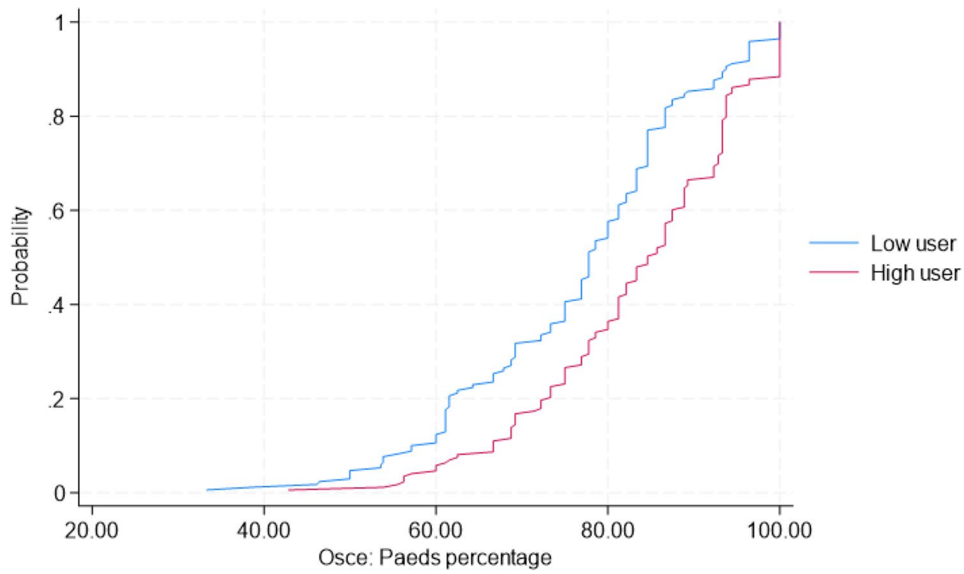

Fig. 2 Cumulative probability distribution of the OSCE scores for low and high users of the PPES online course

Table 5 Propensity score matched average treatment effect (ATE) for the total group and different sub-groups

	n	ATE*	95%CI	*p*-value
Total	**343**	**7.0**	**3.8–10.2**	**< 0.001**
*Previous academic performance***:				
Lowest 20% of the class	62	3.7	−2.3–9.7	0.217
Middle 60% of the class	195	8.2	4.3–12.2	< 0.001
Top 20% of the class	65	2.0	−3.9–7.9	0.501
Area of origin				
Rural	46	2.8	−4.1–9.7	0.414
Urban	267	5.0	1.4–8.6	0.007
Feeder school socioeconomic status				
Public Quintiles 1–4 (Poorest)	58	4.6	−2.2–11.3	0.181
Public Quintile 5 (Richest)	117	4.3	−0.8–9.4	0.096
Private	138	6.27	1.0–11.6	0.020

*Average treatment effect

**Based on 2021 final year marks

Table 6 Use and perceptions of the five modules throughout 2022 (*n* = 260)

Module	Viewed the module (%)	Content clearly presented (%)	Activities* promoted engagement (%)	Activities* prepared students for the OSCE (%)
General Examination	100.0	93.5	89.6	78.0
Vital Signs	98.1	93.0	88.3	77.7
Respiratory Disease and Noisy Breathing	97.7	90.6	87.8	77.1
Nutritional Assessment	96.2	86.4	81.2	73.6
Hydration and Perfusion	94.6	89.8	84.9	69.5

* Bedside clinical skills/drills and multiple-choice questions

The respondents also valued the course because it 'provided an approach to the paediatric examination' (*n* = 39), as expressed by the following respondents:

> *"I feel the large majority of the course was helpful in teaching an approach to a paediatric patient. I feel it is a good resource to refer back to moving forward".*

and

> *"I have a better understanding of how to approach a younger patient, what important things to look out for, and how certain things are different in children/infants than in adults, e.g., breath sounds".*

Respondents also felt that 'the course fostered an appreciation for paediatrics' (*n* = 29). One student said:

> *I learnt a lot from it. It was very informative, and it sparked a passion for paediatrics in me. I intend on going through it again for my electives.*

The negative perceptions about the overall value of the course (*n* = 8) (Fig. 3) included the notions that it 'did not adequately prepare students for the OSCE' (*n* = 3) or their hospital rotations (*n* = 2), as summed up in the following comment from one student :

> *I enjoyed the course and feel that I have more knowledge on how to examine children. However, I did struggle to apply this knowledge to the OSCE as the course taught a broad view of treating children, and the OSCE was very focused on one specific aspect only.*

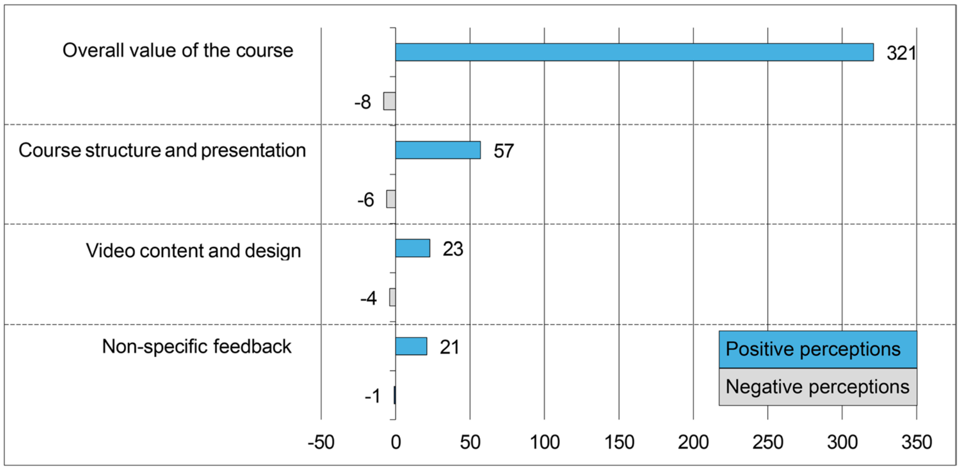

Fig. 3 Respondents' perceptions of whether the course was worth their while ($n = 441$). The numbers represent counts of codes

Figure 4 presents respondents' combined perceptions of the most helpful and least helpful aspects of the course for learning clinical skills in preparation for the final OSCE. To facilitate comparison, the most helpful aspects are represented as 'positive' and the least helpful aspects as 'negative'. Two broad categories related to the course design and video design were identified through the content analysis.

The most helpful aspect of the course design was that students felt it 'prepared them for the OSCE' ($n = 65$). In the words of one student:

> *The course allowed me to know what was expected of me in the OSCE. It allowed me to say, "Okay, I need to know XYZ for [the] respiratory examination.*

The next most helpful aspect of the course design was that it 'included both videos and pictures' ($n = 65$), as exemplified by the following quote:

> *" ...the videos and descriptive pictures were super, super useful. Being shown exactly how some clinical signs look that I've never seen before will make it so much easier for the next time I see it in person at a hospital".*

The overall course structure and presentation ($n = 24$) were also perceived as helpful:

> *It had a systematic approach and was straight to the point. It gave an explanation but also provided pictures and videos to help consolidate clinical examples and understanding.*

and

> *The course was well-structured with integrated activities to test the material. This helped me to learn the material efficiently and gain some confidence to apply it.*

The least helpful aspect of the course design (Fig. 4) was that the 'content was not comprehensive' ($n = 46$). Some respondents wanted more content about vital signs, history-taking, and anthropometry in the Vital Signs module ($n = 33$):

> *"I find that it helped me learn a lot about anthropometry and the different pathologies seen in children, specifically. However, I do wish that we were given the opportunity in the course to plot our own graph and interpret it for practice before the OSCE".*

Requests for more content were often aligned with what had been asked in the OSCE questions.

Another less helpful aspect was the perception that the 'course was poorly designed' ($n = 19$) (Fig. 4). Two respondents' comments illustrate this perception:

> *"So, the course itself was stuff we sort of knew already, just a bit of extra information. I would have rather had these small bits of info given over a lecture rather than a full online course".*

and

> *"Some stuff was repeated several times throughout the course without adding any value; it just made it longer to go through".*

The most helpful aspect of the video design in preparing respondents for the OSCE was that the 'video design was

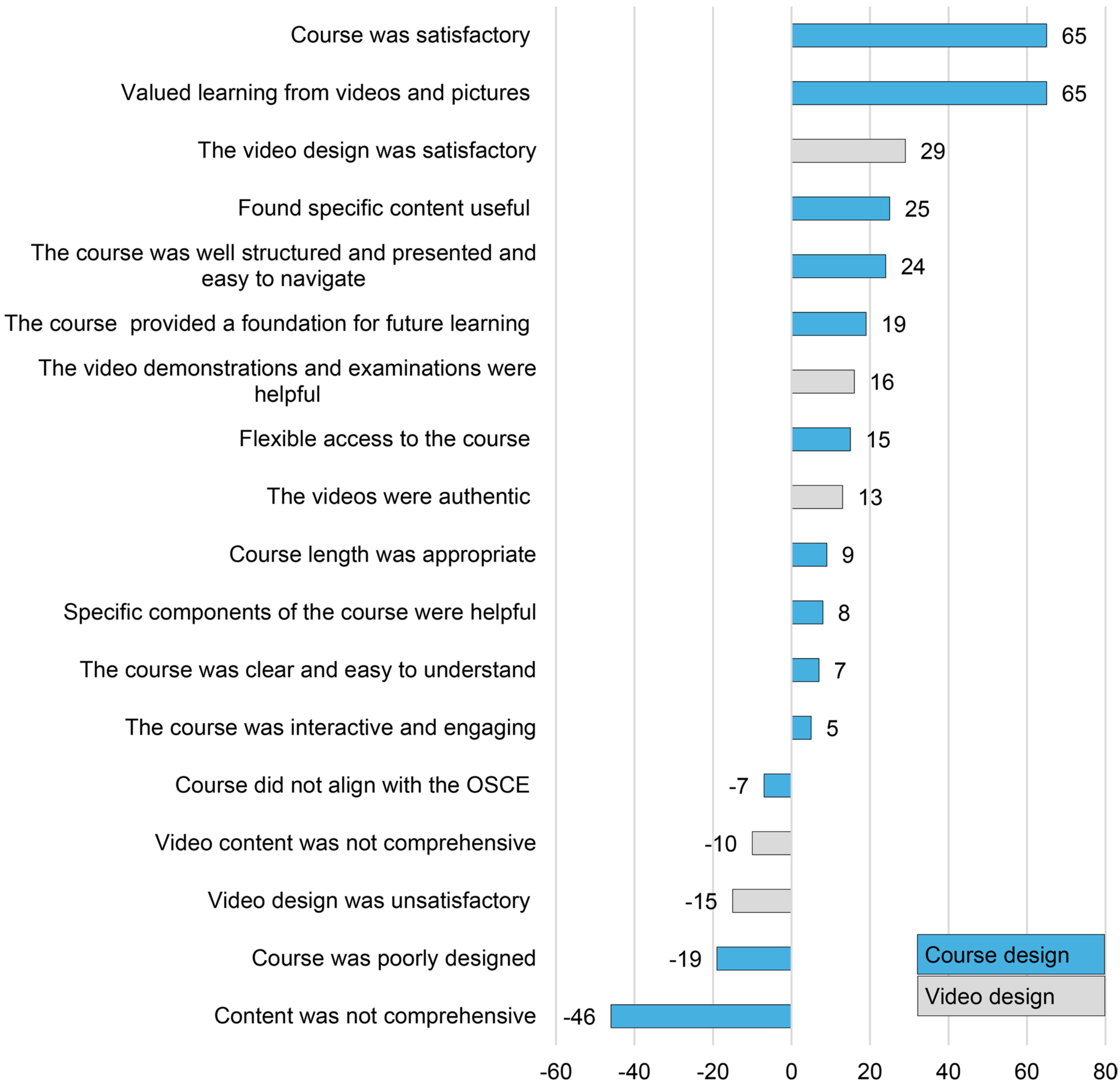

Fig. 4 Perceptions of the course and video design ($n = 425$). The legend represents the categories. The labels on the y-axis represent the subcategories, and the numbers represent counts of codes. Counts < 4 are not displayed

satisfactory' ($n = 29$) (Fig. 4). The video demonstrations and physical examinations, in particular, were regarded as helpful ($n = 16$). Another notable helpful aspect was that 'the videos were authentic' ($n = 13$), as explained by two respondents:

> *"The combination of a simple layout, videos of real patients with the pathologies, and assessments incorporated made for a great learning experience".*

and

> *"The doctors in the videos were examining actual babies instead of dummies, so it helped to see how they interact with the kids and hold their attention".*

The least helpful aspect of the videos was the perception that the 'video design was unsatisfactory' ($n = 15$). The reasons for this perception included that the videos were of 'poor quality' ($n = 3$), 'outdated ($n = 3$), and 'too long' ($n = 3$), as described by the following respondents:

"The videos were not always very clear".

and

"The videos were too long, plus none of the skills provided in the videos were assessed".

Another less helpful feature of the videos was that the 'video content was not comprehensive' ($n = 10$). The major reasons for this perception were that some respondents felt the videos seemed incomplete ($n = 3$) and failed to provide sufficient explanations ($n = 3$):

"Some videos seemed half-done, as in there should be more explanation of how the overall procedure was done".

The survey also asked students how frequently they experienced technical difficulties accessing the online course. The majority of respondents (93.4%) indicated that they never (76.9%) or rarely (16.5%) experienced technical difficulties, with 5.8% responding that they had 'sometimes' experienced problems. The primary technical barrier was internet connectivity issues (n = 29), including connection problems caused by scheduled power outages ($n = 8$). Some students explained how the scheduled power outages, called 'loadshedding' in South Africa, prevented access to the online course:

"...just when load shedding was in place and I couldn't access it at all".

and

Load shedding–lack of WiFi.

The other technical barrier was institutional technical issues ($n = 8$), which referred to difficulties accessing Ulwazi due to server downtime, account-related problems, ICT errors, and maintenance issues, particularly during critical examination periods. The following quotations exemplify the impact of account-related problems and maintenance-related issues:

"There was an ICT error recently where students could not access their accounts as they were deactivated and so could not access the course".

and

"The night before my OSCE, the Wits server was down, and I was unable to access the course via Ulwazi, which was very stressful".

Beyond these reported technical difficulties, almost all respondents (99.23%) indicated that it was easy to find the course on Ulwazi and navigate through it (97.69%).

Discussion

This study consisted of an impact evaluation to determine whether the use of an online paediatric clinical examination skills course influenced performance in an OSCE and an online survey exploring students' experiences and perceptions of the course.

The impact evaluation was based on the supposition that time spent on the course is a measure of course usage. The main finding was that students who spent more time using the course performed better in the OSCE. This finding remained robust even after adjusting for potential selection biases using propensity score matching, which controlled for student characteristics such as gender, age, type of course entry, previous academic performance, and OSCE group. While propensity score matching helped address selection bias, we acknowledge that unobserved confounding variables may still influence the relationship between course usage and OSCE performance.

Possible unobserved confounders in our study may include differences in students' ability to self-regulate their learning and manage their time effectively. According to Bandura's social cognitive theory of self-regulation, "behavior is extensively motivated and regulated by the ongoing exercise of self-influence" (p. 248) [26]. Self-influence becomes essential in online learning, where students must self-regulate their learning [27]. Self-regulated learning encompasses students' ability to set learning goals, monitor their progress, and adjust their learning strategies as needed [28, 29]. Students with stronger self-regulation skills may engage more actively and effectively with online resources [27, 30] and may also perform better in examinations due to their overall learning effectiveness, as demonstrated in one medical study [31]. Similarly, students' abilities to manage their time could influence both the extent of their engagement with supplementary materials and their examination preparation and performance. In our study, it is unclear to what extent self-regulation impacted performance, rather than specific course content or the extent of course usage.

Technical and socioeconomic factors could independently influence both students' ability to engage effectively with video-based learning materials and their overall academic performance. Course access and navigation were not major barriers in our study. It is possible, however, that the lowest performers in our study may not have been sufficiently technologically proficient or comfortable with digital learning to maximise the benefits of the online course. Similar challenges related to students' technological proficiency have been noted in other studies; for example, Huang et al. [31] indicated that participants who lacked computer skills may have felt frustrated when using online learning resources. The study by Hatlevik et al. [32] showed that digital self-efficacy impacts learning. Students' preference for traditional teaching methods has also been shown to affect learning [31]. Our results suggest that lower-performing students need additional learning support to utilise online learning resources effectively.

In our study, most students passed the OSCE, regardless of whether they used the course. This suggests that students demonstrated a high level of competence in the basic skills assessed in the OSCE. Nevertheless, increased course usage had a significant impact on students' OSCE scores in the paediatric component, with a small effect size (Cohen's d = 0.249). The pass rate was also significantly higher in the high-use group.

The scarcity of studies evaluating complex multimedia learning environments for undergraduate medical students makes it challenging to assess the findings from our impact evaluation of the stand-alone PPES course against those of studies comparing videos to other learning modalities. However, given the emphasis on videos in the PPES course, comparisons to the findings from recent meta-analyses of video-based learning are relevant. The Morgado et al. [33] meta-analysis reported moderate effects of video-based learning for knowledge acquisition (Cohen's d = 0.67) and skills development (Cohen's d = 0.76) in 25 studies across medical disciplines. However, the authors highlighted several potential limitations related to the risk of bias across the 40 total included studies, notably selection bias due to shortcomings in the randomisation process (57.5% of 40 studies) and performance bias due to deviations from the planned intervention protocols (47.5% of 40 studies) [33]. These authors also cautioned against assessing the true effectiveness of video-based learning across different educational contexts based on the small number of studies in some disciplines. Another meta-analysis of 50 empirical studies comparing video and traditional learning conditions across multiple disciplines, countries, and educational levels [34] also reported an overall moderate effect size for instructional videos across various educational contexts (Cohen's d = 0.492). Similar effect sizes were found for the specific outcome measures of knowledge acquisition (Cohen's d = 0.477) and skills development (Cohen's d = 0.493). The meta-analysis included only 10 studies focused on medical education.

The significant effectiveness of online teaching has been demonstrated in medical studies. However, heterogeneous online learning approaches were investigated. Where studies compared blended learning with lectures, the significantly better learning outcomes may be due to the experimental groups receiving quantitatively more teaching (an online component in addition to traditional teaching) than the control groups (only traditional teaching). For example, Durán-Guerrero et al. [35] compared radiology lectures to online modules plus lectures. Gong et al. [36] compared lectures with blended learning, which consisted of online resources, active learning before class, face-to-face learning, and simulated clinical training. In another blended learning study, all students attended four hours of lectures on ECG analysis and interpretation, with the experimental group subsequently having additional access to a web application with 24 ECGs [37].

Studies where the comparisons did not involve additional teaching may be less likely to show significant gains. For example, the Berland et al. [38] comparative analysis of online versus in-person opioid overdose prevention training found no meaningful differences between in-person and online training. However, in another study comparing video versus face-to-face teaching, the video group performed significantly better than the control group on the OSCE. These examples highlight the heterogeneity and lack of generalisability of studies involving online components, thus underscoring the need to evaluate such teaching within the context of its use, as has been done in this study.

The differential impact of our online course across the student subgroups warrants discussion. Based on previous academic performance, the middle-performing students derived the largest, statistically significant benefit from the course. In contrast, the lowest- and highest-performing students showed smaller, non-significant effects. Similarly, urban students benefited more than their rural counterparts, and students in private schools benefited more than those in public schools. The analyses are constrained by the small numbers in different sub-groups, so these effects require confirmation in more extensive studies.

Viljoen et al. [37] raised concerns about the financial and temporal cost of their ECG ONLINE software programme. These authors also highlighted the different types of expertise (medical and technical) required to develop the programme. However, small effect sizes, as found in this and other studies, do not necessarily indicate limited value, especially if the resources can be reused across multiple cohorts with minimal ongoing costs, potentially making it cost-effective despite the modest per-student impact.

The student experience of the PPES course was largely positive, as has been found in other studies [37, 39]. Applying cognitive load theory principles in the design of our course may have contributed to its effectiveness. One specific design element valued by respondents was the authenticity of the videos, which referred to the clinical demonstrations involving local patients and expert clinical teachers in real settings. Authentic learning in real-world contexts has long been considered to enhance procedural learning [40]. The students also valued learning clinical signs from videos and pictures. Two considerations for targeting lower-performing students are to include more quizzes and provide immediate feedback on why an answer was incorrect. Mirriahi et al. [41] suggest that students with low prior knowledge would benefit from quizzes with immediate feedback – prompt feedback may help students monitor their understanding and identify areas for improvement.

The major limitation of this study is the lack of a randomised evaluation design. As Heinrich et al. point out, in the absence of an experimental design, "assignment to treatment is frequently nonrandom" (p. 4) [16]. Randomisation was impossible in our study because the 2022 cohort had equal access to the PPES online course. We addressed potential selection biases using propensity score matching to adjust for selection bias [42, 43]. We matched low- and high-course users on the basis of gender, age, type of course entry (non-graduate or graduate), and previous academic performance (based on 2021 marks). Comparisons using five or six subclasses will typically remove 90% or more of the bias present in raw comparisons [44]. We previously acknowledged that unobserved confounding variables may have influenced the relationship between course usage and OSCE performance. Another limitation of the study is the limited scope of the OSCE, which focused on knowledge and not clinical skills. Also, the non-significant findings in some subgroups should not be interpreted as evidence of no effect but rather as inconclusive due to the limited sample sizes. In addition, we did not consider the long-term effects of the course in this study. Viljoen et al. [37] found that the blended learning group in their study retained significantly more knowledge than the conventional teaching group when assessed six months later, making the long-term retention of knowledge and skills from video-based online courses a relevant topic for future research.

Future research involving the PPES course should utilize larger sample sizes to achieve more conclusive results, particularly to confirm the subgroup findings observed in this study. Longitudinal studies examining the retention and transfer of skills to clinical practice would provide valuable insights into the long-term impact of such interventions, particularly during clinical rotations in GEMP 3 and 4. Research should also explore how to design, utilize, or incorporate online courses, such as the PPES course, into teaching to target lower-performing students. Additionally, cost-effectiveness analyses comparing online clinical skills courses to traditional teaching methods would help institutions make informed decisions about resource allocation for similar interventions, taking into account the need to support lower-performing students and non-cost-related pragmatic concerns, such as restrictions on physical space in under-resourced settings.

Conclusion

This study demonstrates that an online paediatric physical examination skills course, incorporating different types of clinical videos, can significantly improve students' clinical examination performance, particularly for those in the middle academic performance range. However, strategies are needed to utilise online courses more effectively to provide additional support for lower-performing students. Applying cognitive load theory principles in course design and using authentic clinical recordings may have contributed to the course's effectiveness. The findings have important implications for supplementing traditional bedside teaching, especially in resource-constrained settings. However, more insight into the specific design components that support learning from multimodal courses and longitudinal studies examining the transfer of clinical skills could inform cost-effective and sustainable institutional strategies to exploit the potential of video-based learning. When considering resource allocation decisions, institutions must weigh the potential impact against the investment required to develop online resources, especially for lower-performing students. Online resources, when viewed against the limitations of traditional bedside teaching (inconsistent instruction, space constraints, and limited patient availability), offer the valuable advantages of repeated access by large numbers of students and subsequent cohorts.

Abbreviations

ATE	Average treatment effect
GEMP	Graduate Entry Medical Program
OSCE	Observed structured clinical examination
PPES	Paediatric Physical Examination Skills
Wits University	University of the Witwatersrand

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-025-08177-5.

Supplementary Material 1: Propensity score matching results.

Supplementary Material 2: Online questionnaire.

Acknowledgements

The authors acknowledge Dr Shannon Leahy's involvement in developing the OSCE questions for the paediatric stations, Dr Stuart Pattinson's invaluable assistance with administering the survey, and the students who participated in this study.

Authors' contributions

AG, SGL, ZD, and DB conceptualised and designed the study. AG and DB collected and analysed the data. AG wrote the first draft of the manuscript. All authors contributed to the interpretation of the data and the revision of the manuscript. All authors read and approved the final version of the manuscript.

Funding
This work is based on the research supported in part by the National Research Foundation of South Africa for the grant, Unique Grant No. SRUG2203291054. The grantholder acknowledges that opinions, findings and conclusions or recommendations reported in this paper are those of the author(s), and that the NRF accepts no liability whatsoever in this regard.

Data availability
The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (Ethics Certificate Number M210663). Informed consent was obtained from all the participants involved in the study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Centre for Health Science Education, Faculty of Health Sciences, University of the Witwatersrand, 29 Princess of Wales Terrace, Parktown, Johannesburg, Gauteng 2193, South Africa
[2]Centre for Health Policy, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, 27 St Andrews Road, Parktown, Johannesburg, Gauteng 2193, South Africa
[3]Vaccines and Infectious Diseases Analytics Research Unit, University of the Witwatersrand, Johannesburg, Gauteng, South Africa
[4]Department of Paediatrics and Child Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, Gauteng, South Africa
[5]Faculty of Health Sciences, Office for Teaching and Learning, University of the Witwatersrand, 29 Princess of Wales Terrace, Parktown, Johannesburg, Gauteng 2193, South Africa

Received: 19 June 2025 / Accepted: 17 October 2025
Published online: 19 November 2025

References

1. Cooper D, Higgins S. The effectiveness of online instructional videos in the acquisition and demonstration of cognitive, affective and psychomotor rehabilitation skills. Br J Educ Technol. 2015;46:768–79. https://doi.org/10.1111/bjet.12166.
2. Brewer PE, Racy M, Hampton M, Mushtaq F, Tomlinson JE, Ali FM. A three-arm single blind randomised control trial of naïve medical students performing a shoulder joint clinical examination. BMC Med Educ. 2021;21:1–7. https://doi.org/10.1186/s12909-021-02822-5.
3. Hogg HDJ, Pereira M, Purdy J, Frearson RJR, Lau GB. A non-randomised trial of video and written educational adjuncts in undergraduate ophthalmology. BMC Med Educ. 2020;20:1–8. https://doi.org/10.1186/s12909-019-1923-1.
4. Mir MA, Marshall RJ, Herbert R, Duthie L, Evans W, Hall R. Comparison between videotape and personal teaching as methods of communicating clinical skills to medical students. Br Med J. 1984;289:31–4. https://doi.org/10.1136/bmj.289.6436.31.
5. McGee RG, Wark S, Mwangi F, Drovandi A, Alele F, Malau-Aduli BS. Digital learning of clinical skills and its impact on medical students' academic performance: a systematic review. BMC Med Educ. 2024;24:1477. https://doi.org/10.1186/s12909-024-06471-2.
6. Schwerdtfeger K, Wand S, Schmid O, Roessler M, Quintel M, Leissner KB, et al. A prospective, blinded evaluation of a video-assisted 4-stage approach during undergraduate student practical skills training. BMC Med Educ. 2014. https://doi.org/10.1186/1472-6920-14-104.
7. Giacomino K, Caliesch R, Sattelmayer KM. The effectiveness of the peyton's 4-step teaching approach on skill acquisition of procedures in health professions education: A systematic review and meta-analysis with integrated meta-regression. PeerJ. 2020;8:1–26. https://doi.org/10.7717/peerj.10129.
8. George A, Blaauw D, Green-Thompson L, Hajinicolaou C, Lala N, Parbhoo K, 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;16:1–16. https://doi.org/10.1186/s41239-019-0164-z.
9. Song HS, Kalet AL, Plass JL. Interplay of prior knowledge, self-regulation and motivation in complex multimedia learning environments. J Comput Assist Learn. 2016;32:31–50. https://doi.org/10.1111/jcal.12117.
10. Sweller J. Cognitive load theory, learning difficulty, and instructional design. Learn Instr. 1994;4:295–312. https://doi.org/10.1016/0959-4752(94)90003-5.
11. Chandler P, Sweller J. Cognitive load theory and the format of instruction. Cogn Instr. 1991;8:293–332. https://doi.org/10.1207/s1532690xci0804_2.
12. Mayer RE. The past, present, and future of the cognitive theory of multimedia learning. Educ Psychol Rev. 2024;36:1–25. https://doi.org/10.1007/s10648-023-09842-1.
13. Ellaway RH, Pusic M, Yavner S, Kalet AL. Context matters: emergent variability in an effectiveness trial of online teaching modules. Med Educ. 2014;48:386–96. https://doi.org/10.1111/medu.12389.
14. Mayer RE. Evidence-based principles for how to design effective instructional videos. J Appl Res Mem Cogn. 2021;10:229–40. https://doi.org/10.1016/j.jarmac.2021.03.007.
15. van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. Med Educ. 2010;44:85–93. https://doi.org/10.1111/j.1365-2923.2009.03498.x.
16. Heinrich C, Maffioli A, Vázquez G. A Primer for Applying Propensity-Score Matching Impact-Evaluation Guidelines. Washington, DC: Inter-American Development Bank; 2010. Available from: https://ideas.repec.org/p/idb/spdwps/1005.html.
17. Stuart E, Green K. Using full matching to estimate causal effects in nonexperimental studies: examining the relationship between adolescent marijuana use and adult outcomes. Dev Psychol. 2008;44:395–406. https://doi.org/10.1037/0012-1649.44.2.395.
18. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70:41–55. https://doi.org/10.1093/biomet/70.1.41.
19. Herbert RD. How to estimate treatment effects from reports of clinical trials. I: continuous outcomes. Aust J Physiother. 2000;46:229–35. https://doi.org/10.1016/s0004-9514(14)60334-2.
20. Krippendorff K. Content analysis: An introduction to its methodology. 4th ed. SAGE; 2019.
21. Anderson G, Arsenault N. Fundamentals of educational research. 2nd ed. London: RoutledgeFalmer; 1998.
22. Eldh AC, Årestedt L, Berterö C. Quotations in qualitative studies: reflections on constituents, custom, and purpose. Int J Qual Methods. 2020. https://doi.org/10.1177/1609406920969268.
23. Ogbonnaya UI, Awuah FK. Quintile ranking of schools in South Africa and learners' achievement in probability. Stat Educ Res J. 2019;18:106–19. https://doi.org/10.52041/serj.v18i1.153.
24. Republic of South Africa Department of Education. Amended National Norms and Standards for School Funding, South African Schools Act, 1996 (Act No 84 of 1996). Republic of South Africa: 2006. Available from: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.gov.za/sites/default/files/gcis_document/201409/33971gon47.pdf.
25. Starks H, Diehr P, Curtis JR. The challenge of selection bias and confounding in palliative care research. J Palliat Med. 2009;12:181–7. https://doi.org/10.1089/jpm.2009.9672.
26. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process. 1991;50:248–87.
27. Edisherashvili N, Saks K, Pedaste M, Leijen Ä. Supporting self-regulated learning in distance learning contexts at higher education level: systematic literature review. Front Psychol. 2022;12:1–21. https://doi.org/10.3389/fpsyg.2021.792422.
28. Cho KK, Marjadi B, Langendyk V, Hu W. The self-regulated learning of medical students in the clinical environment - a scoping review. BMC Med Educ. 2017;17:1–13. https://doi.org/10.1186/s12909-017-0956-6.

29. van Houten-Schat MA, Berkhout JJ, van Dijk N, Endedijk MD, Jaarsma ADC, Diemers AD. Self-regulated learning in the clinical context: a systematic review. Med Educ. 2018;52:1008–15. https://doi.org/10.1111/medu.13615.
30. Ballouk R, Mansour V, Dalziel B, McDonald J, Hegazi I. Medical students' self-regulation of learning in a blended learning environment: a systematic scoping review. Med Educ Online. 2022;27:1–7. https://doi.org/10.1080/10872981.2022.2029336.
31. Huang Z, Yang J, Wang H, Chen B, Zheng D, Chen H. Integration of massive open online course (MOOC) in ophthalmic skills training for medical students: outcomes and perspectives. Asia-Pac J Ophthalmol. 2022;11:543–8. https://doi.org/10.1097/APO.0000000000000548.
32. Hatlevik OE, Throndsen I, Loi M, Gudmundsdottir GB. Students' ICT self-efficacy and computer and information literacy: determinants and relationships. Comput Educ. 2018;118:107–19. https://doi.org/10.1016/j.compedu.2017.11.011.
33. Morgado M, Botelho J, Machado V, Mendes JJ, Adesope O, Proença L. Video-based approaches in health education: a systematic review and meta-analysis. Sci Rep. 2024;14:23651. https://doi.org/10.1038/s41598-024-73671-7.
34. Lin Y, Yu Z. A meta-analysis evaluating the effectiveness of instructional video technologies. Technol Knowl Learn. 2024;29:2081–115. https://doi.org/10.1007/s10758-023-09669-3.
35. Durán-Guerrero JA, Ulloa-Guerrero LH, Salazar-Díaz LC. Blended learning: an effective methodology for teaching radiology to medical students. Rev Fac Med. 2019;67:273–7. https://doi.org/10.15446/revfacmed.v67n2.69862.
36. Gong J, Ruan M, Yang W, Peng M, Wang Z, Ouyang L, et al. Application of blended learning approach in clinical skills to stimulate active learning attitudes and improve clinical practice among medical students. PeerJ. 2021. https://doi.org/10.7717/peerj.11690.
37. Viljoen CA, Millar RS, Manning K, Burch VC. Effectiveness of blended learning versus lectures alone on ECG analysis and interpretation by medical students. BMC Med Educ. 2020;20:1–16. https://doi.org/10.1186/s12909-020-02403-y.
38. Berland N, Lugassy D, Fox A, Goldfeld K, Oh SY, Tofighi B, et al. Use of online opioid overdose prevention training for first-year medical students: a comparative analysis of online versus in-person training. Subst Abus. 2019;40:240–6. https://doi.org/10.1080/08897077.2019.1572048.
39. Huang AY, Ryan A, Bearman M, Molloy E. The influence of clinical practice video on learning conversations in postgraduate medical education: a scoping review. Focus on Health Professional Education: A Multi-Professional Journal. 2024;25:26–43. https://doi.org/10.11157/fohpe.v25i3.695.
40. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65:S63-7. https://doi.org/10.1097/00001888-199009000-00045.
41. Mirriahi N, Jovanović J, Lim LA, Lodge JM. Two sides of the same coin: video annotations and in-video questions for active learning. Educ Technol Res Dev. 2021;69:2571–88. https://doi.org/10.1007/s11423-021-10041-4.
42. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med. 1997;127:757–63. https://doi.org/10.7326/0003-4819-127-8_part_2-199710151-00064.
43. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. Stat Med. 2007;26:20–36. https://doi.org/10.1002/sim.2739.
44. Cochran W. The effectiveness of adjustment by subclassification in removing bias in observational studies. Biom. 1968;24:295–313.

Publisher's Note