

Implementation of the objective structured clinical examination in medical education: a systematic review of its applicability and effectiveness

Implementación del Examen Clínico Objetivo Estructurado en educación médica: revisión sistemática sobre su aplicabilidad y efectividad

Sergio G. Roca  • Nataly Duran*   • Jaime Mercado 

Universidad Privada Domingo Savio, Santa Cruz, Bolivia.

*Corresponding author

Recepción: 20-04-2026

Aceptación: 03-07-2026

Publicación: 31-07-2026

ABSTRACT

The Objective Structured Clinical Examination (OSCE) constitutes a standardized tool for assessing clinical competencies in medical students. This systematic review aimed to analyze its implementation and effectiveness in face-to-face and virtual modalities. A search for studies published between 2019 and 2025 was conducted following the PRISMA guidelines, including research related to student performance and the assessment of clinical competencies. The results showed that OSCE enables the objective evaluation of clinical skills, communication, clinical reasoning, and decision-making; however, virtual formats present limitations in assessing procedures and physical examination skills. In addition, an increase in academic stress associated with this assessment was identified. It is concluded that OSCE is an effective and adaptable method, although it requires greater standardization, examiner training, and the development of hybrid models.

Keywords: Objective Structured Clinical Examination, medical education, medical students, clinical competence, educational assessment, computer simulation.

RESUMEN

El Examen Clínico Objetivo Estructurado (ECO) constituye una herramienta estandarizada para evaluar competencias clínicas en estudiantes de medicina. Esta revisión sistemática tuvo como objetivo analizar su implementación y efectividad en modalidades presencial y virtual. Se realizó una búsqueda de estudios publicados entre 2019 y 2025 siguiendo las directrices PRISMA, incluyendo investigaciones relacionadas con el desempeño estudiantil y la evaluación de competencias clínicas. Los resultados evidenciaron que el ECO permite valorar de manera objetiva habilidades clínicas, comunicación, razonamiento clínico y toma de decisiones; sin embargo, los formatos virtuales presentan limitaciones para evaluar procedimientos y la exploración física. Además, se identificó un incremento del estrés académico asociado a esta evaluación. Se concluye que el ECO es un método efectivo y adaptable, aunque requiere mayor estandarización, capacitación de evaluadores y desarrollo de modelos híbridos.

Palabras clave: Examen Clínico Objetivo Estructurado, educación médica, estudiantes de medicina, competencia clínica, evaluación educacional, simulación por computador.

Cite as: Roca, S. G., Duran, N., & Mercado, J. (2026). Implementation of the objective structured clinical examination in medical education: a systematic review of its applicability and effectiveness. *Revista Gregoriana de Ciencias de la Salud*, 3(2), 163-184. <https://doi.org/10.36097/rgcs.v3i2.3227>

© Author(s) 2026

INTRODUCTION

Current medical education aims to train professionals capable of managing complex clinical situations, based on students' theoretical knowledge, practical skills, and attitudes; therefore, the Objective Structured Clinical Examination (OSCE) allows these areas to be objectively measured (Cueller et al., 2023). Since its introduction by Harden in 1975, the OSCE has evolved as a method that overcomes the subjectivity of traditional clinical examinations by assessing core competencies such as history-taking, physical examination, clinical reasoning, and effective patient communication.

The OSCE is a key tool in medical education for objectively and standardly assessing students' clinical competencies. It uses a series of stations where students face simulated clinical scenarios, allowing for a comprehensive evaluation (Ramos et al., 2019). In this regard, evaluating the effectiveness of the OSCE ensures that clinical competencies are measured accurately and reliably. Recent studies confirm that it is a reliable method for determining clinical skills in specific scenarios (Martínez González et al., 2018).

Although this assessment has been widely used to measure clinical competencies in medical students, doubts persist regarding its direct impact on learning and clinical performance. Furthermore, the current literature shows inconsistencies in its implementation and effectiveness, making it difficult to standardize this evaluative method. These uncertainties highlight the need for additional research to optimize its application in medical education (Gómez Peña et al., 2019).

A 2022 systematic review analyzed studies in Latin America and found significant differences in the number of stations, duration, and areas assessed in the OSCEs, hindering the comparison of results and the standardization of the methodology. This underscores the need for research establishing clear guidelines for its effective application (Cardona-Arias, 2022).

Added to this challenge is the recent forced migration toward virtual environments due to the COVID-19 pandemic, which has raised new questions about the validity and fidelity of the OSCE in remote formats, as well as the equivalence of results between both modalities. Although efforts have been made to adapt clinical stations to digital platforms, there are still no unified criteria to guarantee their effectiveness in this new context.

As a result of the review conducted, this study aimed to evaluate the implementation and effectiveness of the OSCE in the training of clinical competencies in medical students, identifying strengths, weaknesses, and areas for improvement in its application, both in face-to-face and virtual

modalities, to provide evidence that guides its standardization in both settings.

METHODOLOGY

The present study was designed as a systematic literature review, following the guidelines of the PRISMA 2020 statement (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). The review was not prospectively registered in a protocol database (such as PROSPERO), given its exploratory nature and its focus on the qualitative synthesis of the available evidence.

To guide the search and selection of studies, the research question was defined using the PICO framework: Population (undergraduate medical students), Intervention (application of the OSCE as a tool for assessing clinical competencies), Comparison (not applicable, as this is a descriptive review), and Outcomes (measurement of clinical competencies, including practical skills, communication, decision-making, and other relevant competencies assessed through the OSCE).

Inclusion and exclusion criteria were established to select the most relevant articles. Empirical investigations (clinical trials, quasi-experimental studies, and observational studies) were included, as well as systematic reviews, bibliographic reviews, meta-analyses, and scoping reviews, focusing on undergraduate medical students who had been assessed using the OSCE for their clinical competencies. Only articles published in English or Spanish between January 2019 and February 2025 were considered. Excluded were letters to the editor, editorials, conference abstracts, and theses; studies focusing exclusively on other health professionals (nursing, physiotherapy, or postgraduate); evaluations that did not use the OSCE methodology or that combined multiple methods without a separate analysis of this tool; studies that did not provide clear data on the impact on clinical competencies or that were limited to administrative or logistical aspects; and publications in languages other than English or Spanish, or published before 2019.

Systematic searches were conducted in two recognized databases in the health sciences field: PubMed and Mendeley. The selection of these databases was based on their broad coverage of biomedical and educational literature, as well as their institutional accessibility. The search strategy was built using a combination of key terms and Boolean operators, employing the following search string: "Objective structured clinical review" AND "medical student". Filters were applied in both databases to retrieve only studies published within the established date range

(2019-2025) and in the languages of interest (English and Spanish). Additionally, the reference lists of selected articles were reviewed to identify potential additional studies not captured by the electronic search (manual searching or "snowballing").

The article selection was carried out in six stages, following the workflow recommended by PRISMA 2020. In the first stage, studies older than five years relative to the search date were excluded. In the second stage, only open-access articles were selected to ensure full-text availability. In the third stage, two reviewers independently screened the titles of the retrieved records to assess their preliminary relevance; disagreements were resolved by consensus or through the intervention of a third reviewer. In the fourth stage, the abstracts of preselected studies were read to verify their alignment with the review objective and the inclusion criteria. In the fifth stage, the full texts of articles that passed the previous phase were analyzed, exhaustively confirming compliance with all eligibility criteria. Finally, in the sixth stage, duplicate articles identified across the different databases were removed. The flow diagram of the selection process, with the number of records identified, screened, assessed, and included at each phase, is presented in Figure 1.

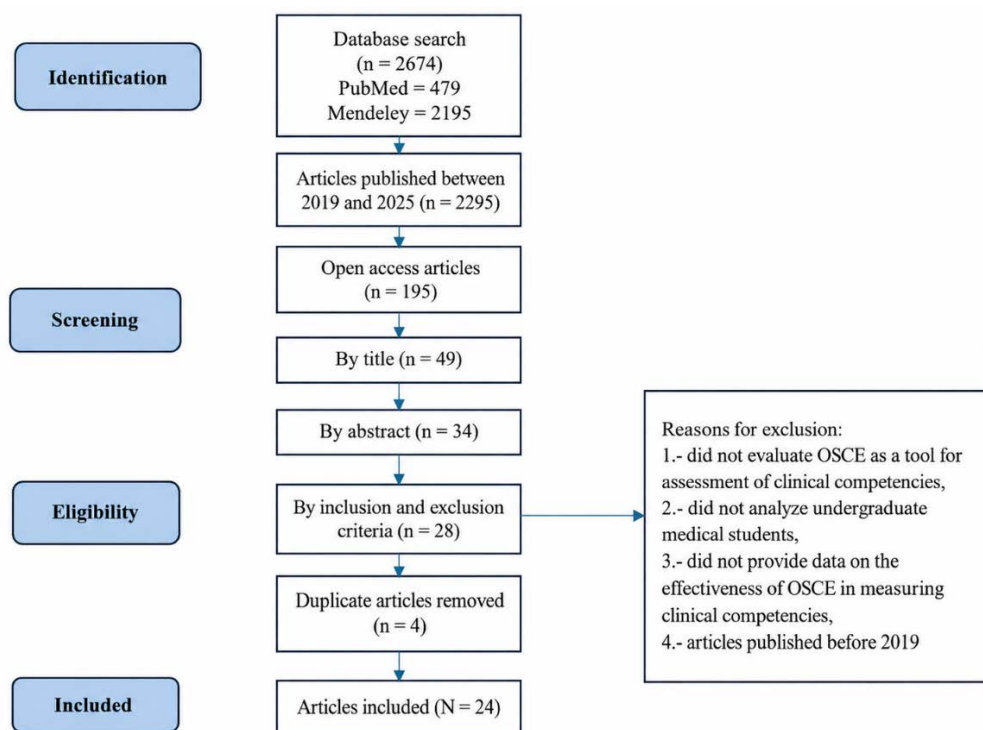


Figure 1. PRISMA 2020 flow diagram of the study identification, screening, eligibility, and inclusion process.

Data extraction was performed by a single reviewer and verified by a second reviewer to minimize errors. The following information was collected from each selected study: complete bibliographic reference (title, authors, and publication year), study design, sample size, study population (including demographic and academic characteristics), intervention applied (description of the OSCE: number of stations, duration, areas assessed, and modality —face-to-face or virtual—), and key results on the effectiveness of the OSCE in measuring clinical competencies. The data were organized into a standardized evidence table to facilitate comparison across studies.

The methodological quality of the included studies was independently assessed by two reviewers, using tools adapted to each study design. For observational and quasi-experimental studies, the modified Newcastle-Ottawa scale was used, while for systematic reviews, the AMSTAR 2 tool was applied. Disagreements in scoring were resolved through discussion until consensus was reached. No study was excluded based on its methodological quality, but this information was considered in the qualitative analysis of the findings.

Given the expected heterogeneity in study designs, populations, interventions, and outcome measures, a meta-analysis was not performed. Instead, the data were analyzed qualitatively using a narrative synthesis approach. The findings were grouped thematically to identify trends in the implementation and effectiveness of the OSCE in the training of clinical competencies in medical students, as well as the strengths, weaknesses, and areas for improvement reported in the literature. Likewise, the methodologies used across the different studies, their findings, and their implications for medical education were compared, with special attention to the differences between the face-to-face and virtual modalities.

Since this systematic review was based entirely on previously published and openly accessible scientific literature, it did not require approval from an ethics committee or the participation of human subjects. All data were extracted from secondary sources, respecting copyright and the usage policies of each publisher.

RESULTS AND DISCUSSION

Table 1 provides a comprehensive summary of the 24 studies included in this systematic review, detailing the primary author, year of publication, study design, sample characteristics, and country of origin for each investigation.

Table 1. Characteristics of the articles included in the review

No.	Author(s)	Type of study	Sample	Country
1	Chan et al. (2023)	Revision systematic	17 studies included	International
2	Felthun et al. (2021)	Retrospective observational	285 senior students	Australia
3	Felthun et al. (2021)	Revision systematic	Eleven items were included	International
4	Krishnamoorthy et al. (2022)	Instrument validation	30 students from the third semester of the medical degree.	India
5	Komasawa et al. (2020)	Study correlational	94 medical students	Japan
6	Schwill et al. (2020)	Mixed study (quantitative and qualitative)	74 student evaluators and 307 students evaluated	Germany
7	Saad et al. (2022)	Descriptive phenomenological qualitative.	23 stakeholders, including examiners, academics, simulated patients, professional staff, students, and quality assurance examiners.	Australia
8	Sharma et al. (2022)	Cross-sectional study	100 medical students	India
9	Shimizu et al. (2024)	Validation design based on Messick's validity framework	22 panelists	Japan
10	Chang et al. (2023)	Review study	21 comparative studies	International
11	Arekat et al. (2022)	Mixed cross-sectional study	124 final-year medical students and 60 teachers	Kingdom of Bahrain
12	Bajra et al. (2023)	Quasi-experimental pre-and post -intervention design.	133 third and fourth-year medical students	USA
13	Bernard et al. (2020)	Rehearsal-controlled randomized	Pre-clerkship students	USA
14	Ruwizhu et al. (2022)	Study pilot	17 students	Zimbabwe
15	Wilkes et al. (2024)	Study observational	-	Australia
16	Alsahafi et al. (2023)	Revision systematic	14 studies reviewed	International
17	Martínez et al. (2023)	Randomized, double-blind controlled trial	50 fourth-year medical students	Chili
18	Hawezy et al. (2024)	Investigation of empirical qualitative	132 students from the 4th and 6th year of medicine	Iraq
19	Al-Hashimi et al. (2023)	Literature review	Not applicable (literature review)	Kingdom United
20	Saeed et al. (2024)	Post-event observational study with survey	142 participants (105 completed the survey)	Saudi Arabia

No.	Author(s)	Type of study	Sample	Country
21	Aqib et al. (2024)	Development and implementation of a virtual tool for medical education (preprint, under review).	Six medical students and residents from Western University and McMaster University participated in the testing phase.	Canada
22	Iida et al. (2021)	Study experimental pilot.	16 third-year medical students (women) in Japan.	Japan
23	Shen et al. (2020)	OSCE validation study using the Electronic Health Record (EHR).	119 third-year medical students.	USA
24	Choi et al. (2023)	Study comparative quasi-experimental	83 medical students participated in the virtual OSCE.	USA

The included literature demonstrates substantial geographical diversity, encompassing empirical research from North America (the United States and Canada), Asia (India, Japan, Iraq, Saudi Arabia, Bahrain, and China), Oceania (Australia), Europe (Germany and the United Kingdom), South America (Chile), and Africa (Zimbabwe), alongside several multi-national systematic reviews. The sample sizes exhibit considerable variation, ranging from small pilot cohorts of 16 to 17 participants (e.g., Iida et al., 2021; Ruwizhu et al., 2022) to large-scale evaluations involving over 300 students and faculty members (Schwill et al., 2020), reflecting the diverse scope, objectives, and resource availability of the primary studies. This tabulated synthesis serves as the foundational evidence base from which all subsequent qualitative analyses and thematic interpretations were derived, ensuring full transparency and replicability in accordance with PRISMA 2020 reporting standards (Item 19 – Study characteristics).

In alignment with PRISMA 2020 guidelines, the systematic mapping of study characteristics presented in Table 1 is essential for assessing the applicability, methodological robustness, and potential biases of the retrieved evidence. The marked heterogeneity in study designs is immediately apparent: the table includes randomized controlled trials (Bernard et al., 2020; Martínez et al., 2023), quasi-experimental pre-post intervention designs (Bajra et al., 2023; Choi et al., 2023), cross-sectional surveys (Sharma et al., 2022), mixed-method approaches (Schwill et al., 2020; Arekat et al., 2022), qualitative phenomenological studies (Saad et al., 2022; Hawezy et al., 2024), instrument validation frameworks (Krishnamoorthy et al., 2022; Shimizu et al., 2024; Shen et al., 2020), systematic and literature reviews (Chan et al., 2023; Felthun et al., 2021; Chang et al., 2023; Alsahafi et al., 2023; Al-Hashimi et al., 2023), and even a preprint under

peer review (Aqib et al., 2024). This broad methodological spectrum, while enriching the review with multiple perspectives, inherently precluded a pooled quantitative synthesis (meta-analysis) due to the lack of uniform outcome measures, variable intervention protocols, and differing comparator groups. Consequently, and as reflected in the PRISMA 2020 flow diagram (Figure 1), a narrative synthesis approach was adopted to meaningfully integrate findings across these disparate designs, paying careful attention to contextual factors such as the timing of the study relative to the COVID-19 pandemic, the modality of the OSCE (face-to-face vs. virtual), and the specific competencies targeted by each evaluation.

Table 1 shows three clusters that inform the review objectives. The first cluster comprises rigorously designed comparative trials and quasi-experimental studies (e.g., Bernard et al., 2020; Martínez et al., 2023; Bajra et al., 2023; Choi et al., 2023), which constitute the core evidence supporting the equivalence and effectiveness of virtual OSCE formats relative to traditional face-to-face assessments, particularly in terms of overall pass rates and global competency scores. The second cluster encompasses qualitative, mixed-methods, and cross-sectional investigations (e.g., Saad et al., 2022; Schwill et al., 2020; Arekat et al., 2022; Hawezy et al., 2024; Sharma et al., 2022), which contribute nuanced insights into stakeholder perceptions, the psychological impact of high-stakes assessments (Iida et al., 2021), and the practical barriers to implementing standardized examinations across resource-limited settings. The third cluster groups validation, pilot, and instrument-development studies (e.g., Krishnamoorthy et al., 2022; Shimizu et al., 2024; Shen et al., 2020; Ruwizhu et al., 2022; Komasaawa et al., 2020), which highlight ongoing efforts to refine the psychometric properties of the OSCE, integrate digital tools such as the Electronic Health Record, and adapt the methodology to emerging educational technologies. This compilation enables critical evidence appraisal and provides a framework for thematic synthesis.

Table 2 offers a comprehensive and multidimensional synthesis of the 24 studies included in this systematic review. This structured tabulation serves as an analytical scaffold that enables a direct cross-comparison of heterogeneous interventions, ranging from traditional in-person Objective Structured Practical Examinations (OSPE) in physiology (Sharma et al., 2022) and large-scale Grand OSCEs integrating multiple specialties (Hawezy et al., 2024), to fully remote teleOSCE formats leveraging videoconferencing platforms such as Zoom (Saeed et al., 2024; Arekat et al., 2022), and quasi-experimental comparisons between pre-pandemic and pandemic-era cohorts (Felthun et al., 2021b).

Table 2. Characteristics, results, and areas for improvement in the implementation of the virtual and face-to-face OSCE

Reference	OSCE implementation method	Skills clinics evaluated	Key results	Strengths identified	Weaknesses or limitations	Areas for improvement
Sharma et al. (2022)	OSPE with structured stations for clinical physiology.	system, respiratory system, abdomen, nerves, cranial, and the central nervous system.	94.4% considered OSPE a good experience, and 93.5% were satisfied with the content.	Objective, structured evaluation and high acceptance.	Physical fatigue, need for teacher training, and limited interaction with evaluators.	Incorporate OSPE into formative assessments and explore its use in summative assessments.
Hawezy et al. (2024)	GOSCE integrates multiple specialties into a single exam.	Clinical skills, communication, decision making, and patient safety.	80% perceived the exam as fair, and 85.6% indicated broad coverage of competencies. The feedback strengthened learning and confidence for summative assessments.	Comprehensive and interdisciplinary evaluation.	Insufficient preparation time and variable perception of fairness.	Optimize preparation and logistics times.
Al-Hashimi et al. (2023)	OSCE training with multiple structured stations.	Practical skills, communication, reasoning, and clinical decision-making.	The simulator obtained slightly higher scores (+2.8), with no significant differences.	Identify learning needs and improve clinical skills.	Variability due to emotional factors and evaluator subjectivity.	Standardize criteria, increase stations, and decrease student stress.
Martinez et al. (2023)	Randomized trial comparing virtual clinical simulator (Body Interact®) and small group discussion.	Adherence to clinical guidelines, diagnosis, treatment, and decision-making.	The simulator obtained slightly higher scores (+2.8), with no significant differences.	Randomized design and OSCE developed by experts with national standards.	Brief intervention, Hawthorne effect, and active comparator.	Conduct prolonged interventions, follow-up, and comparison with traditional teaching.
Saeed et al. (2024)	Virtual OSCE via Zoom with workrooms and station rotation.	Medical history, physical examination, data interpretation, communication, and therapeutic planning.	High satisfaction (4.7/5); most preferred the virtual modality, although there were doubts about its equivalence with the face-to-face OSCE.	Accessibility, cost savings, and good organization.	Little time for feedback, technical problems, and less practical interaction.	Optimize digital tools and technological readiness.

Reference	OSCE implementation method	Skills clinics evaluated	Key results	Strengths identified	Weaknesses or limitations	Areas for improvement
Arekat et al. (2022)	OSCE online via Zoom with simulated stations.	Communication, reasoning, and clinical decision-making.	The virtual OSCE showed high validity and reliability.	Lower costs, accessibility, and maintenance of standards.	It does not allow for a proper evaluation of the physical examination.	Incorporate haptic technologies and augmented reality.
Bernard et al. (2020)	OSCE formative and summative assessments are integrated with clinical notes, presentations, and clinical reasoning.	Medical history, communication, physical examination, clinical reasoning, and medical documentation.	High satisfaction; 100% indicated feeling prepared for clinical practice.	Curriculum integration and detailed feedback.	High cost and high demand for trained teachers.	Optimize resources and strengthen standardization through technology.
Chang et al. (2023)	Comparison between OSCE online and written exams.	Expert role doctor and communicator.	The OSCE online exam was comparable to the written exams.	Suitable for assessing specific skills.	Coverage is limited by some skills clinics.	Expand the competencies assessed in the OSCE.
Shimizu et al. (2024)	Messick framework.	Patient safety: error identification, communication, and double-checking.	Good reliability and low correlation with written exams.	Solid methodological validation and focus on essential competencies.	Validation is limited to an institutional context.	Incorporate more scenarios and validate in other environments.
Ruwizhu et al. (2022)	Pilot OSCE in psychiatry with five stations and simulated patients.	Clinical interview, mental status examination, risk assessment, and communication.	Good organization; deficiencies were identified in continuity of questions and interpersonal skills.	Use of trained simulated patients and training of evaluators.	High cost, need for training, and low discrimination at some stations.	Standardize workstations, optimize time, and strengthen interpersonal skills.
Wilkes et al. (2024)	Comparison between face-to-face OSCE and virtual via videoconference.	Mental status examination, communication, differential diagnosis, and therapeutic planning.	Scores were higher online; no relevant differences in some domains.	Viability and good reliability of the virtual format.	Limited assessment of practical skills and dependence on connectivity.	Develop the best strategies for evaluating procedural skills and integrating digital platforms.

Reference	OSCE implementation method	Skills clinics evaluated	Key results	Strengths identified	Weaknesses or limitations	Areas for improvement
Aqib et al. (2024)	OSCE online and in-person with real-time simulations and standardized checklists.	Communication, clinical reasoning, differential diagnosis, physical examination, and CanMEDS competencies.	Better performance and confidence in virtual mode; no significant differences between modalities.	Accessibility, immediate feedback, and standardization.	Technical problems, less practical interaction, and variable perception of effectiveness.	Optimize platforms, increase simulations, and develop international standards.
Bajra et al. (2023)	TeleOSCE via videoconference and remote practice with MonkeyJacket.	Communication in telemedicine, virtual physical examination, clinical interview, reasoning, and management plans.	Self-efficacy and communication skills improved; deficiencies in confidentiality and limitations of teleconsultation persisted.	Interactivity, free access, and inter-institutional practice.	Limited evaluation of physical examination and specific skills by specialty.	teleconsultation skills and expand the platform to other specialties.
Alsahafi et al. (2023)	TeleOSCE with digital tools, pre-workshops, and standardized stations.	Communication, clinical interview, virtual physical examination, empathy, and decision-making.	Self-efficacy and performance improved with structured feedback.	Integration of telemedicine, active learning, and objective assessment.	Limitations in evaluation skills for distance learning physics.	Standardize feedback and incorporate further scenarios clinicians.
Iida et al. (2021)	OSCE with groups of five students, simulated patients, and video recording.	Communication, diagnosis, treatment, procedural skills, and stress response.	Psychological stress and fatigue were observed before the exam, with no significant changes in oxidative stress.	Objective assessment of multiple competencies and review through recordings.	Stress can affect performance; limited practical assessment and variability between evaluators.	Implement strategies to reduce stress and optimize feedback.
Shen et al. (2020)	OSCE station focused on the use of electronic health records (EHR).	Patient communication, EHR integration, and clinical documentation.	Communication improved after training; performance in clinical documentation decreased.	Structured teaching and a feasible large-scale assessment.	Difficulty in integrating the EHR without affecting the doctor-patient interaction, and differences in familiarity with the platform.	Strengthen training in documentation and use of the EHR during the consultation.
Krishnamoorthy	Development	Communication,	Valid and reliable	Standardized	Insufficient time	Optimize time,

Reference	OSCE implementation method	Skills clinics evaluated	Key results	Strengths identified	Weaknesses or limitations	Areas for improvement
et al. (2022)	and validation of a scale to evaluate OSCE attributes with eight structured stations.	clinical reasoning, procedural skills, and formative and summative assessment.	scale ($\alpha = 0.81$); 80% considered that the OSCE covered the learning objectives.	evaluation, immediate feedback, and high acceptance.	at stations, stress, and difficulties in uniform implementation.	expand its application, and validate it in larger samples.
Felthun et al. (2021a)	TeleOSCE via videoconference with remote observation and comparison with face-to-face OSCE.	Communication, clinical reasoning, and procedures adapted to the virtual environment.	Good acceptance and validity comparable to the face-to-face OSCE.	Flexibility, integration of telemedicine, and cost reduction.	Technological dependence and limited assessment of manual skills.	Improve remote sensing and develop hybrid methods.
Felthun et al. (2021b)	Retrospective comparison between teleOSCE and face-to-face OSCE.	Communication, physical examination, clinical reasoning, and management in various specialties.	There were no significant differences in most domains; better performance in some areas of the teleOSCE.	A viable, accessible, and lower-cost method.	Limitations in evaluating procedures and technological dependence.	Develop hybrid models and expand research in different specialties.
Komasawa et al. (2020)	OSCE with seven stations and two trained evaluators; comparison with CBT and clinical performance.	Medical interview, physical examination, neurology, emergencies, and clinical techniques.	The OSCE correlated with clinical performance, unlike the CBT.	An objective tool for predicting clinical performance.	Poor correlation between theoretical knowledge and practical skills; possible bias among evaluators.	Integrate advanced simulations and assess the long-term impact.
Schwill et al. (2020)	OSCE training with peer assessment, teleOSCE, and clinical simulation in primary care.	Skills clinics basics, medical interview, physical examination, telemedicine, and response to emergencies.	High satisfaction with peer evaluation, greater self-efficacy, and correlation between OSCE performance and clinical practice.	Lower cost, development of teaching skills, and strengthening of clinical communication.	Possible peer bias, limitations in confidentiality, and poor correlation with computerized tests.	Improve objectivity, strengthen confidentiality training, and integrate OSCE with computerized assessments.

Reference	OSCE implementation method	Skills clinics evaluated	Key results	Strengths identified	Weaknesses or limitations	Areas for improvement
Chan et al. (2023)	vOSCE via videoconference with real-time evaluation and comparison with face-to-face OSCE.	Communication, clinical consultation, data interpretation, diagnosis, and telemedicine.	The vOSCE was a viable alternative; students reported less anxiety than in the face-to-face OSCE.	Accessibility, lower cost, and standardized assessment of communication skills.	Limited assessment of physical skills, connectivity problems, and less appreciation of nonverbal communication.	Develop hybrid models, optimize platforms, and standardize evaluator training.
Choi et al. (2023)	Comparison between virtual and face-to-face OSCE through five training stations with self-assessment and peer assessment. OSCE virtual via videoconference with students, simulated patients, and examiners in remote mode.	Advanced clinical communication, medical interviews, and patient interaction.	The virtual OSCE scored lower in interview content, with no differences in structure; students felt emotionally supported.	Educational value comparable to face-to-face learning, immediate feedback, and accessibility.	Less effective in assessing non-verbal communication and natural interaction with patients.	Improve the teaching of interview content, strengthen virtual interaction, and standardize the evaluation.
Saad et al. (2022)		Medical history, communication, clinical reasoning, decision making, and telemedicine.	The virtual OSCE was adequate for assessing medical history and communication; it had limitations in assessing physical skills.	Accessibility, continuity of assessment during the pandemic, and cost reduction.	Difficulty in evaluating clinical procedures, technological dependence, and less interaction with the patient.	Implement hybrid models, optimize digital platforms, and strengthen examiner training.

Table 2 also captures methodological innovations, including the integration of the Electronic Health Record (EHR) into OSCE stations (Shen et al., 2020), the use of validated psychometric frameworks such as Messick's validity framework (Shimizu et al., 2024), and the adoption of peer-assessment strategies (Schwill et al., 2020). By distilling the essential characteristics and outcomes of each study into a standardized format, Table 2 facilitates a transparent, replicable, and critically informed narrative synthesis, fully aligned with PRISMA 2020 reporting standards for the presentation of study characteristics and results.

A dominant and recurrent theme emerging from Table 2 concerns the comparative effectiveness, feasibility, and inherent trade-offs between virtual and traditional face-to-face OSCE modalities. The accumulated evidence robustly indicates that virtual OSCEs are broadly viable, logistically advantageous, and well-received by students and examiners alike, offering substantial benefits in terms of cost reduction, geographical accessibility, and continuity of assessment during public health restrictions (Saeed et al., 2024; Alsahafi et al., 2023; Aqib et al., 2024). Notably, Felthun et al. (2021b) demonstrated that teleOSCE yielded comparable or even superior scores in certain domains—particularly physical examination ($P < 0.001$)—relative to pre-pandemic in-person assessments, a finding attributed to enhanced observational clarity through optimized camera angles and examiner focus. However, a consistent and critical limitation permeating nearly all virtual implementations is the compromised evaluation of procedural and hands-on clinical skills, including palpation, auscultation, detailed neurological assessments, and other manual examination techniques that require direct physical patient contact (Saad et al., 2022; Wilkes et al., 2024; Choi et al., 2023). This persistent deficiency has catalyzed a consensus across multiple studies on the imperative for hybrid methodologies that strategically reserve in-person stations for psychomotor competencies while leveraging digital platforms for communication, history-taking, clinical reasoning, and data interpretation stations (Bajra et al., 2023; Arekat et al., 2022; Chang et al., 2023). Furthermore, the table illuminates important psychosocial dimensions: while Iida et al. (2021) documented significant elevations in stress and fatigue biomarkers (s-HHV-6 and u-5-HT) prior to OSCE administration, Chan et al. (2023) paradoxically reported lower self-reported anxiety levels among candidates in virtual formats, suggesting that modality-specific emotional responses are complex, context-dependent, and warrant dedicated investigation to optimize student well-being without compromising assessment rigor.

Beyond the virtual–presential dichotomy, Table 2 reveals several cross-cutting

methodological and implementation challenges that persist across the entire body of evidence, irrespective of modality or setting. First, marked heterogeneity in examiner calibration, checklist standardization, station logistics, and feedback protocols constitutes a recurrent weakness, as explicitly documented in the "areas for improvement" column across multiple entries (Krishnamoorthy et al., 2022; Schwill et al., 2020; Ruwizhu et al., 2022). This variability underscores the pressing need for internationally harmonized guidelines, structured examiner training programs, and the adoption of validated, specialty-specific rubrics to enhance inter-rater reliability and ensure equitable assessment across different institutions and geographical contexts. Second, the integration of ancillary digital tools—exemplified by the use of the Electronic Health Record (EHR) in Shen et al. (2020)—presents a double-edged opportunity: while it enhances the authenticity of medical documentation and aligns assessment with contemporary clinical practice, it simultaneously risks disrupting the physician–patient therapeutic alliance if students are not adequately trained to balance screen-based tasks with active, empathetic communication. Third, the substantial diversity in sample sizes (ranging from small pilot cohorts of 16 participants to large-scale evaluations of over 300 subjects), study designs (from double-blinded randomized controlled trials to qualitative phenomenological inquiries and preprint protocols under peer review), and outcome measures prevents direct meta-analytic pooling, thereby limiting the statistical generalizability of individual findings. Nevertheless, the collective evidence synthesized in Table 2 robustly affirms the OSCE as a psychometrically sound and pedagogically invaluable instrument for clinical competency assessment, while simultaneously delineating a clear and actionable roadmap for future refinement. Priority research directions include the development and validation of augmented reality and haptic technologies to bridge the gap in remote physical examination evaluation, the implementation of longitudinal studies to assess the sustained impact of OSCE-based training on postgraduate clinical performance, and the co-creation of consensus-based, culturally adaptable frameworks that standardize OSCE delivery, feedback quality, and examiner accreditation across diverse educational and resource-limited settings.

The findings of this systematic review substantiate the relevance of the OSCE as a central instrument for the standardized assessment of clinical competencies in medical students. Multiple studies demonstrate that the OSCE facilitates an objective measurement of practical, communicative, and clinical reasoning skills, ensuring reproducibility and fairness in evaluative processes (Chan et al., 2023). In this regard, comparative evidence between the face-to-face and

virtual modalities suggests that both can achieve similar levels of effectiveness; however, notable differences are documented in students' perceptions of the digital environment and in the feasibility of assessing specific competencies that require direct physical interaction (Felthun et al., 2021b).

The adaptability of the OSCE to virtual settings, accelerated by the COVID-19 pandemic, constitutes one of the most prominent findings in recent literature. The implementation of teleOSCE has proven to be feasible and well-accepted by students, ensuring the continuity of clinical assessment in remote contexts without substantially compromising the overall validity of the examination (Saeed et al., 2024). Nevertheless, significant limitations persist in the evaluation of procedural skills (such as physical examination maneuvers, palpation techniques, and auscultation) and in the interpretation of physical signs that require direct patient contact. This shortcoming underscores the need to design hybrid methodologies that complement virtual assessment with specific in-person sessions for these competencies, thus combining technological flexibility with the rigor of traditional clinical interaction (Saad et al., 2022).

Another emerging aspect in the reviewed evidence is the incorporation of digital technologies into OSCE stations, particularly the use of the Electronic Health Record (EHR). The integration of the EHR can enhance the accuracy of medical documentation and promote more structured clinical communication, aligning assessment with real-world practice; however, its improper use tends to divert the examiner's or student's attention toward the screen, interfering with rapport and active listening with the simulated patient. This finding highlights the necessity of prior training for students in the simultaneous management of digital records and human interaction, so that technology acts as a facilitator rather than a barrier in the physician–patient relationship (Shen et al., 2020).

The emotional and psychological component of the OSCE has also been subject to analysis in the literature. It has been documented that the examination situation produces a considerable increase in stress and fatigue biomarkers, which potentially impacts cognitive performance, decision-making during the test, and students' mental health in the medium term (Iida et al., 2021). To alleviate these effects, the implementation of psychological support programs and the design of evaluative scenarios that reduce anxiety without compromising academic rigor are suggested—for example, through early familiarization with the station format or the inclusion of strategic breaks. Likewise, immediate and structured feedback emerges as a crucial component that transforms the evaluative experience into a formative opportunity, favoring the consolidation of

competencies, real-time error correction, and enhanced self-confidence (Arekat et al., 2022). Nevertheless, variability in grading criteria and heterogeneity in examiner training remain important barriers to achieving fully equitable and reproducible assessment across different sites or panels (Krishnamoorthy et al., 2022).

In summary, this systematic review substantiates that the OSCE endures as a reliable and effective methodological cornerstone when assessing clinical competencies in undergraduate medical education. While the virtual format represents a valuable and necessary alternative in certain contexts—especially under in-person restrictions—its full effectiveness depends on complementarity with face-to-face sessions for specific skills, as well as on rigorous standardization of grading processes and continuous training of assessors. Furthermore, the importance of addressing students' emotional well-being and consciously leveraging digital tools to avoid compromising the quality of simulated care is highlighted. Future research should focus on evaluating the long-term impact of the OSCE on sustained learning, refining hybrid models that optimize the fidelity and validity of the evaluative process, and developing consensual guidelines that harmonize its implementation across diverse educational settings, both face-to-face and virtual.

CONCLUSIONS

The findings of this systematic review confirm that OSCE is an effective and reliable assessment strategy for measuring clinical competencies among medical students. Its application in virtual environments has demonstrated feasibility and educational value, particularly for communication skills, clinical reasoning, and decision-making; however, limitations remain in the assessment of physical examination and procedural abilities. Future implementations should prioritize standardized evaluation criteria, examiner training, technological improvements, and integrated psychological support strategies to reduce student stress and enhance the validity and fairness of clinical assessment.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

AUTHOR CONTRIBUTIONS

Conceptualization: Sergio G. Roca and Nataly Duran. **Data curation:** Jaime Mercado

and Nataly Duran. **Formal analysis:** Sergio G. Roca and Nataly Duran. **Funding acquisition:** Jaime Mercado and Nataly Duran. **Investigation:** Sergio G. Roca and Nataly Duran. **Methodology:** Jaime Mercado and Nataly Duran. **Project administration:** Sergio G. Roca and Nataly Duran. **Resources:** Jaime Mercado and Nataly Duran. **Software:** Sergio G. Roca and Nataly Duran. **Supervision:** Jaime Mercado and Nataly Duran. **Validation:** Sergio G. Roca and Nataly Duran. **Visualization:** Jaime Mercado and Nataly Duran. **Writing – original draft:** Sergio G. Roca and Nataly Duran. **Writing – review and editing:** Jaime Mercado and Nataly Duran.

REFERENCES

- Al-Hashimi, K., Said, U. N., & Khan, T. N. (2023). Formative Objective Structured Clinical Examinations (OSCEs) as an Assessment Tool in UK Undergraduate Medical Education: A Review of Its Utility. *Cureus*. <https://doi.org/10.7759/CUREUS.38519>
- Alsahafi, A., Ling, D. L. X., Newell, M., & Kropmans, T. (2023). A systematic review of effective quality feedback measurement tools used in clinical skills assessment. *MedEdPublish* (2016), 12, 11. <https://doi.org/10.12688/MEP.18940.2>
- Aqib, A., Fareez, F., Assadpour, E., Babar, T., Kokavec, A., Wang, E., Lo, T., Lam, J.-P., & Smith, C. (2024). Development of a Novel Web-Based Tool to Enhance Clinical Skills in Medical Education. *JMIR Medical Education*, 10, e47438-e47438. <https://doi.org/10.2196/47438>
- Arekat, M., Shehata, M. H., Deifalla, A., Al-Ansari, A., Kumar, A., Alsenbesy, M., Alshenawi, H., El-Agroudy, A., Husni, M., Rizk, D., Elamin, A., Salah, A. Ben, & Atwa, H. (2022). Evaluation of the Utility of Online Objective Structured Clinical Examination Conducted During the COVID-19 Pandemic. *Advances in Medical Education and Practice*, 13, 407-418. <https://doi.org/10.2147/AMEP.S357229>
- Bajra, R., Srinivasan, M., Torres, E. C., Rydel, T., & Schillinger, E. (2023). Training future clinicians in telehealth competencies: outcomes of a telehealth curriculum and teleOSCEs at an academic medical center. *Frontiers in Medicine*, 10. <https://doi.org/10.3389/FMED.2023.1222181>
- Bernard, A. W., Thomas, L., Rockfeld, J., & Cassese, T. (2020). Expanding OSCE-related Learning Opportunities For Pre-Clerkship Students: Insights From an Assessment for Learning Curriculum. *Journal of Medical Education and Curricular Development*, 7, undefined-undefined. <https://doi.org/10.1177/2382120520940663>

- Cardona-Arias, J. A. (2022). Revisión sistemática de las aplicaciones de la evaluación clínica objetiva estructurada (ECO-E) en la formación médica latinoamericana. *Boletín Semillero de Investigación En Familia*, 4(1), 2022. <https://doi.org/10.22579/27448592.827>
- Chan, S. C. C., Choa, G., Kelly, J., Maru, D., & Rashid, M. A. (2023). Implementation of virtual OSCE in health professions education: A systematic review. *Medical Education*, 57(9), 833-843. <https://doi.org/10.1111/MEDU.15089>
- Chang, O., Holbrook, A. M., Lohit, S., Deng, J., Xu, J., Lee, M., & Cheng, A. (2023). Comparability of Objective Structured Clinical Examinations (OSCEs) and Written Tests for Assessing Medical School Students' Competencies: A Scoping Review. *Evaluation & the Health Professions*, 46(3), 213-224. <https://doi.org/10.1177/01632787231165797>
- Choi, A., Morrison, L. J., Murtha, T., & Talwalkar, J. (2023). A Comparison Between In-Person and Virtual OSCE on Advanced Communication Skills for Senior Medical Students (Sci200). *Journal of Pain and Symptom Management*, 65(5), undefined-undefined. <https://doi.org/10.1016/J.JPAINSYMMAN.2023.02.250>
- Cueller, A. G., Dueñas, R. S., Espinoza, E. M., Fuertes, S. P., Gómez, J. E., Haro, L. M., Medina, N. P., & Calderón, L. R. (2023). *Evaluación de competencias y resultados de aprendizaje en el curriculum médico 2022* [Preprint]. ResearchGate. <https://doi.org/10.13140/RG.2.2.33620.68486>
- Felthun, J. Z., Taylor, S., Shulruf, B., & Allen, D. W. (2021a). Assessment methods and the validity and reliability of measurement tools in online objective structured clinical examinations: a systematic scoping review. *Journal of Educational Evaluation for Health Professions*, 18. <https://doi.org/10.3352/JEEHP.2021.18.11>
- Felthun, J. Z., Taylor, S., Shulruf, B., & Allen, D. W. (2021b). Empirical analysis comparing the tele-objective structured clinical examination (teleOSCE) and the in-person assessment in Australia. *Journal of Educational Evaluation for Health Professions*, 18, 23. <https://doi.org/10.3352/JEEHP.2021.18.23>
- Gómez Peña, L., Dávalos Espinosa, L. G., Rodríguez Prieto, P. F., Blanco Zabala, E., Viera Hernández, R. V., Rocha Cisneros, I. de la C., Gómez Peña, L., Dávalos Espinosa, L. G., Rodríguez Prieto, P. F., Blanco Zabala, E., Viera Hernández, R. V., & Rocha Cisneros, I. de la C. (2019). La evaluación clínica objetiva estructurada desde el área de formación médica general. *Investigación En Educación Médica*, 8(32), 40-50.

<https://doi.org/10.22201/FACMED.20075057E.2019.32.18133>

- Hawezy, D. J., Barzinjy, S. T., Saadadin, G., Hawezy, G. A., & Baqer, W. (2024). Perception of undergraduate medical students and examiners towards grand objective structured clinical examination. *BMC Medical Education*, 24(1). <https://doi.org/10.1186/S12909-024-06388-W>
- Iida, T., Ito, Y., Kanazashi, M., Murayama, S., Miyake, T., Yoshimaru, Y., Tatsumi, A., & Ezoe, S. (2021). Effects of Psychological and Physical Stress on Oxidative Stress, Serotonin, and Fatigue in Young Females Induced by Objective Structured Clinical Examination: Pilot Study of u-8-OHdG, u-5HT, and s-HHV-6. *International Journal of Tryptophan Research : IJTR*, 14. <https://doi.org/10.1177/11786469211048443>
- Komasawa, N., Terasaki, F., Nakano, T., & Kawata, R. (2020). Relationships between objective structured clinical examination, computer-based testing, and clinical clerkship performance in Japanese medical students. *PloS One*, 15(3). <https://doi.org/10.1371/JOURNAL.PONE.0230792>
- Krishnamoorthy, Y., Kannusamy, S., Ganesh, K., Thulasingham, M., Lakshminarayanan, S., & Kar, S. (2022). Development and Validation of Scale Assessing the Preparedness of Objective Structured Clinical Examination in Undergraduate Competency based Medical Education. *Indian Journal of Community Medicine : Official Publication of Indian Association of Preventive & Social Medicine*, 47(4), 522-526. https://doi.org/10.4103/IJCM.IJCM_1392_21
- Martínez González, A., Trejo Mejía, J. A., Martínez González, A., & Trejo Mejía, J. A. (2018). ¿Cómo realizar un ECOE? *Investigación En Educación Médica*, 7(28), 98-107. <https://doi.org/10.22201/FACMED.20075057E.2018.28.18123>
- Martinez, F. T., Soto, J. P., Valenzuela, D., González, N., Corsi, J., & Sepúlveda, P. (2023). Virtual Clinical Simulation for Training Amongst Undergraduate Medical Students: A Pilot Randomised Trial (VIRTUE-Pilot). *Cureus*, undefined-undefined. <https://doi.org/10.7759/CUREUS.47527>
- Ramos, J. M., Martínez-Mayoral, M. A., Sánchez-Ferrer, F., Morales, J., Sempere, T., Belinchón, I., & Compañ, A. F. (2019). Análisis de la prueba de evaluación clínica objetiva estructurada (ECOE) de sexto curso en la Facultad de Medicina de la Universidad Miguel Hernández de Elche. *Educación Médica*, 20, 29-36.

<https://doi.org/10.1016/J.EDUMED.2017.07.020>

- Ruwizhu, T., Nyamukapa, R., Mazhandu, F., Mutambara, J., Mangezi, W., & Whitwell, S. (2022). Piloting the use of objective structured clinical examinations (OSCEs) to assess undergraduate medical students' clinical competence in psychiatry in Zimbabwe. *BJPsych International*, 19(3), 75-77. <https://doi.org/10.1192/BJI.2021.49>
- Saad, S. L., Richmond, C., Jones, K., Schlipalius, M., Rienits, H., & Malau-Aduli, B. S. (2022). Virtual OSCE Delivery and Quality Assurance During a Pandemic: Implications for the Future. *Frontiers in Medicine*, 9. <https://doi.org/10.3389/FMED.2022.844884>
- Saeed, E., Hamad, M. H., Alhuzaimi, A. N., Aljamaan, F., Elsenterisi, H., Assiri, H., Alhasan, K., Bashiri, F. A., Kambal, M., Khalil, M. S., Abdulghani, H. M., Al-Tawfiq, J. A., Al-Eyadhy, A., & Tamsah, M.-H. (2024). Virtual Objective Structured Clinical Examination (OSCE) Training in the Pandemic Era: Feasibility, Satisfaction, and the Road Ahead. *Cureus*, 16(6). <https://doi.org/10.7759/CUREUS.61564>
- Schwill, S., Fahrbach-Veeser, J., Moeltner, A., Eicher, C., Kurczyk, S., Pfisterer, D., Szecsenyi, J., & Loukanova, S. (2020). Peers as OSCE assessors for junior medical students - a review of routine use: a mixed methods study. *BMC Medical Education*, 20(1). <https://doi.org/10.1186/S12909-019-1898-Y>
- Sharma, A., Tahilramani, H., Misra, M., Somani, S., & Tak, A. (2022). Objective Structured Practical Examination as a Tool of Assessing Practical Skills in Clinical Physiology: Perceptions of Medical Students and Faculty. *Scripta Medica (Banja Luka)*, 53(1), 13-20. <https://doi.org/10.5937/SCRIPTAMED53-35975>
- Shen, E., Cristiano, J. A., & Ellis, L. R. (2020). The Electronic Health Record Objective Structured Clinical Examination Station: Assessing Student Competency in Patient Notes and Patient Interaction. *MedEdPORTAL: The Journal of Teaching and Learning Resources*, 16, 10998. https://doi.org/10.15766/MEP_2374-8265.10998
- Shimizu, I., Tanaka, K., Mori, J.-I., Yamauchi, A., Kato, S., Masuda, Y., Nakazawa, Y., & Kanno, H. (2024). Objective Structured Clinical Examination to Assess Patient Safety Competencies of Japanese Medical Students: Development and Validation Argument. *Cureus*, 16(11). <https://doi.org/10.7759/CUREUS.73969>
- Wilkes, F. A., Looi, J. C. L., Maguire, P. A., Bonner, D., Reay, R. E., Brazel, M., Keightley, P., Tedeschi, M., Wardle, C., & Kramer, D. (2024). Online medical student OSCE

examinations during the first three years of the COVID-19 pandemic compared to three years pre-pandemic: An Australian experience in psychiatry and addiction medicine. *Medical Teacher*, 46(6), 776-781. <https://doi.org/10.1080/0142159X.2023.2279918>

Statement on the use of artificial intelligence: The authors acknowledge the use of generative artificial intelligence (AI) and AI-assisted technologies solely to improve the readability, language quality, and clarity of this manuscript. AI tools may also have been used to create or edit non-scientific graphical content, such as graphical abstracts, icons, illustrations, diagrams, or educational figures, provided that such use is transparently disclosed and does not misrepresent or fabricate research findings. AI tools were not used to generate, modify, enhance, interpret, or analyze scientific data, medical images (e.g., radiographs, computed tomography scans, magnetic resonance imaging, ultrasound images, histopathological images, microscopy images, or other diagnostic images), or any figures that constitute primary research data. The authors reviewed and edited all AI-assisted content as necessary and assume full responsibility for the accuracy, integrity, and originality of the final published work.

Disclaimer / Editor's Note: All publications' statements, opinions, and data are solely those of the individual authors and contributors, not *Revista Gregoriana de Ciencias de la Salud* or the editors. *Revista Gregoriana de Ciencias de la Salud* and/or the editors disclaim all responsibility for any injury to persons or property resulting from any ideas, methods, instructions, or products referred to in the content.