An Explorative Review of the Constructs, Metrics, Models, and Methods for Evaluating e-Learning Performance in Medical Education

Deborah Oluwadele[1,2], Yashik Singh[1] and Timothy Adeliyi[2]
[1]Department of Telemedicine, School of Nursing and Public Health, University of KwaZulu-Natal, Durban, South Africa
[2]Department of Informatics; University of Pretoria, South Africa

deborah.oluwadele@up.ac.za (corresponding author)
singhy@ukzn.ac.za
timothy.adeliyi@up.ac.za

Abstract: The performance evaluation of e-learning in medical education has been the subject of much research lately. Researchers are yet to achieve a consensus on the definition of performance or the suitable constructs, metrics, models, and methods to help understand student performance. Through a systematic review, this study put forward a working definition of what constitutes performance evaluation to reduce the ambiguity, arbitrariness, and multiplicity surrounding performance evaluation of e-learning in medical education. A systematic review of published articles on performance evaluation of e-learning in medical education was performed on the SCOPUS, Web of Science, PubMed, and EBSCOHost databases using search terms deduced from the PICOS model. Following the PRISMA guidelines relevant published papers were searched and exported to Endnote. Screening and quality appraisal were done on Rayyan. Three thousand four hundred and thirty-nine published studies were retrieved and screened using predetermined inclusion and exclusion criteria. One hundred and three studies passed all the criteria and were reviewed. The reviewed literature used 30 constructs to operationalize performance. The leading constructs are knowledge and effectiveness. Both constructs were used by 60% of the authors of the reviewed literature to define student performance. Knowledge gain, satisfaction, and learning outcome are the most common metrics used by 81%, 26%, and 15% of the reviewed literature to measure student performance. The study discovered that most researchers forget to evaluate the “e” or electronic component of e-learning when evaluating performance. The constructs operationalized and metrics measured were primarily focused on learning outcomes with minimal focus on technology-related metrics or the influence of the electronic mode of delivery on the learning process or evaluation outcome. Only 6% of the reviewed literature applied evaluation models to guide their evaluation process - mostly the Kirkpatrick evaluation model. Also, most of the included studies used randomization as an experimental control method, mainly using pre-and post-test surveys. Modern evaluation methods were rarely used. Only 1% of the reviewed literature used Google Analytics, and 2% used data from a learning management system. This study increments the existing body of knowledge in performance evaluation of e-learning in medical education by providing a convergence of constructs, metrics, models, and methods and proposing a roadmap to guide students’ performance evaluation process from the synthesis of findings and the gaps identified through the systematic review of existing literature in the domain. This roadmap will assist in informing researchers of grey areas to consider when evaluating performance to ensure more quality research outputs in the domain.

Keywords: e-Learning, e-Learning evaluation, Factors, e-Learning performance, Medical education, Roadmap, Systematic literature review

1. Background

The rapid development of digital technologies, such as the Internet of Things (IoT) artificial intelligence (AI), and robotics, as well as globalization, fundamentally alter civilization as Information Technology (IT) is harnessed as a tool for social transformation. This is apparent in Japan’s ‘Society 5.0’ concept. This concept seeks to revolutionize industries, living spaces, and public habits through the integration of IT (Fukuyama, 2018). Society 5.0 is based on four key constructs - a society that is human-centered, integrates cyberspace and physical space, is knowledge-intensive, and is data-driven. (Deguchi et al., 2020). Aside from using technology to optimize processes like admission and registration, e-learning is one of the critical ways that higher education institutions continue to leverage technology. Fukuyama (2018) designed a chart summarizing how Keidanren, Japan’s business federation, re-aligned its corporate behavior charter to realize a sustainable society by proactively delivering on Sustainable Development Goals (SDGs) through Society 5.0. This chart identified seventeen SDGs that will be achieved using information technologies, such as big data, the Internet of Things, and Virtual Reality, as enablers. This includes Educational Technology (e-Learning system).

Reference this paper: Oluwadele, D, Singh, Y and Adeliyi,T., 2023. An Explorative Review of the Constructs, Metrics, Models, and Methods for Evaluating e-Learning Performance in Medical Education, *The Electronic Journal of e-Learning, 21(5)*, pp 394-412, available online at www.ejel.org

According to Clark and Mayer (2016), e-learning is the delivery of instructional content via digital devices like desktops, laptops, tablets, or mobile phones. In the current teaching and learning process, e-learning is used to support education, advance knowledge, enhance performance, and enhance students' learning outcomes. Evaluation is crucial to implementing e-learning because it provides a pathway to monitor project milestones and deliverables (Galas et al., 2018) and retrospection of what works and needs to be improved. As the world recovers from the threats of COVID-19 and students in higher education institutions return to face-to-face learning, it is tempting to consider e-learning an outdated contingency approach to teaching in turbulent times. However, considering the burden of diseases versus the capacity of healthcare professionals to manage this burden, especially in Low-to-Middle Income Countries (LMICs), e-learning is appreciated for its potential to strengthen the human resource capacity in health.

The evaluation of e-Learning in Medical Education(e-LMED) raises questions such as "What should be evaluated, and how?" Should the unit of analysis be focused on the methodology, users, learning content, or technology used, quantity of knowledge and skills acquired, benefits derived from applying these, or the context that influences the suitability of all the factors named earlier? (Cairó, Barreiro and Solsona, 2000). To date, the pedagogical and empirical answers to these questions create a divergence rather than a convergence of the body of knowledge in the domain. This is because researchers have yet to achieve a consensus on the definition of performance evaluation or the suitable constructs, metrics, models, and methods to help evaluate student performance in the domain. Hence the need to operationalize performance evaluation of e-learning in the domain.

Operationalization is the process of turning an ambiguous definition into a specific one that can be measured and tested. Constructs, metrics, models, and methods can be used to operationalize performance evaluation. The concepts, variables, or phenomena a researcher plans to explore are called constructs in research. Constructs are typically abstract and vague; hence, operationalizations or indicators that can be observed or assessed are necessary to measure them (Rubin and Babbie, 2016). We refer to these indicators as metrics. Metrics are precise measurements used to evaluate performance or monitor advancement toward a given purpose or goal. A metric is a measure that can be verified, expressed in quantitative or qualitative terms, and described as a standard (Melnyk, Stewart and Swink, 2004). Although factors and metrics are different ideas, they are frequently combined when assessing performance. In research, the term "factors" refers to various components or characteristics that may impact a study's findings or the reliability of its conclusions. To create accurate and trustworthy research results, it is essential to recognize and account for these elements (Salkind, 2010). Factors can aid in determining the root causes of performance, whereas metrics can be considered as varying operationalization of constructs to offer a practical means of measuring performance.

Even though theories, models, and frameworks are often used interchangeably, there exist apparent differences between these phenomena. Theories describe, explain, and predict factors that influence an outcome. They explain the meaningful relationship between constructs and how a construct may change the behavior of another construct (Foy et al., 2011). Models describe without necessarily explaining the process of translating research into practice. They may be used to make assumptions of cause and effect between constructs (Bauer et al., 2015). Frameworks provide a broad set of constructs that describe concepts and data without explaining cause and effect between the constructs. They may also offer prescriptive steps that summarize how implementation should ideally be conducted (Meyers, Durlak and Wandersman, 2012).

Through a systematic literature review, this study put forward a working definition of what constitutes performance evaluation to reduce the ambiguity, arbitrariness, and multiplicity surrounding performance evaluation of e-learning in medical education by exploring the constructs, metrics, models, and methods used in the domain. This paper presents a brief contextual background to our study. The next part of the paper presents the method, study design, and analysis results. The study concludes with the discussion and conclusion sessions where results are interpreted, the study's limitations are highlighted, and recommendations are made for future research.

2. Materials and Methods

2.1 Systematic Review

Literature reviews play a crucial role in academic research to gather and examine the body of knowledge of a domain. However, narrative literature reviews are rid of systematic investigation owing to the lack of methodological steps that help to conduct the review in a scientifically replicable manner (Linnenluecke, Marrone and Singh, 2020). Hence, this study used a Systematic Literature Review (SLR) method. SLR is regarded as a valuable method of research because it conforms to scientific method principles by being "designed to locate, appraise, and synthesize the best available evidence" relative to the study's aim to deliver "informative and evidence-based" research (Boland, Dickson and Cherry, 2017). SLR Planning, Execution, and Reporting are the three key phases of the SLR execution process. During these phases, the steps involved include developing research questions, conducting a thorough search for primary studies, assessing the quality of the studies that were included, identifying and extracting the data needed to answer the research questions, summarizing and synthesizing the study results, interpreting the findings to determine their applicability, and finally, writing reports. This section describes the systematic approach adopted to ensure the study is replicable and the results are transparent for other researchers. All the literature published so far on the evaluation of student performance in an e-learning context in medical education was explored, and our roadmap and empirical findings in response to the research questions raised are presented.

2.2 Review Protocol

Before conducting the systematic review, the review was planned by defining a review protocol using the guidelines by Kitchenham and Charters (Kitchenham et al., 2009). During this stage, research questions were chosen, a protocol was developed, and all the authors validated the procedure to determine whether the approach was practical. Along with the study topics, beginning search strings and publication selection criteria are included. After all of this information was defined, the protocol was amended again. The internal processes of the Plan Review stage are depicted in Figure 1.

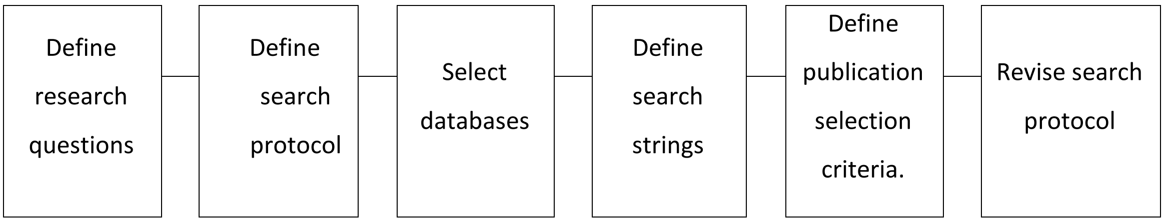

Figure 1: The Internal Process of the "Plan Review" Phase

2.3 Research Questions

This SLR seeks to answer the main research question: "How is performance operationalized and measured in an e-learning context in medical education? The Research Questions (RQs) presented in Table 1 were raised to answer this question.

Table 1: Study Research Questions and Their Purposes

	Research Questions	Purpose
RQ1	How is performance operationalized in literature in e-LMED?	This research question was poised to explore how performance is defined by pinpointing the constructs evaluated in the reviewed literature.
RQ2	What are the metrics and factors measured when evaluating performance in e-LMED?	This question aims to identify the metrics or factors measured when student performance is evaluated.
RQ3	What are the methods used to evaluate performance in e-LMED?	This identifies the research design, techniques, and approaches used to evaluate student performance in the reviewed literature.
RQ4	What models, theories, or frameworks are used to evaluate performance in e-LMED?	This highlights the conceptual models, theories, or frameworks used as guiding principles to define constructs and inform metrics evaluated to understand student performance in the reviewed literature.
RQ5	What gap exists in the literature on performance evaluation in e-LMED?	This question identifies possible research gaps in performance evaluation in the domain by synthesizing all the findings from RQ1-RQ4.

2.4 Data Sources

Scopus, Web of Science, PubMed, and EBSCOHost databases were searched to retrieve the literature related to the performance evaluation of students in e-LMED. These databases were identified as those with the highest number of literature on the research subject.

2.5 Search Strategy

This Systematic Review was carried out between July 2022 and April 2023. These databases were searched with terms developed using the PICOS (Population, Intervention, Comparisons, Outcomes, and Setting) model. The

PICOS model ensures scientific diligence and objectivity of reviews by prescribing methodological standards that enhance the value of the scientifically published literature reviews and guarantee their robust reproducibility (Saaiq and Ashraf, 2017). Using the PICOS model, the Population was defined as e-learning publications—any publication on e-learning. The Intervention construct of the PICOS model was used to delimit the Population of literature focusing on evaluating performance in an e-learning context. The Comparison and outcome constructs were not applicable because of the aim of the study. Within the defined intervention, however, the Setting construct explained the context within which this study will consider studies on performance evaluation in e-learning. Hence, the setting was defined as medical education. Thus, this study searched the relevant databases for publications on e-learning (Population) evaluating performance (intervention) in medical education (setting). Table 2 shows how the PICOS model was used to define the search criteria.

Table 2: Search criteria using the PICOS model

Search Criteria for Studies in the Review	
Population	*E-learning publications; any intervention, course, program, or module run online. The Population will examine papers from all over the world.*
Interventions	*Performance Evaluation: Any construct, factors, methods, or models used to assess e-learning interventions*
Comparisons	*Not applicable*
Outcomes of interest	*Not applicable*
Setting	*Medical Education; any program, module, training, or intervention in medicine.*

The databases were searched using keywords derived from the search criteria and associated synonyms used by researchers in published literature. To improve the accuracy of the search strategy, sharpen the result, and ensure a limited number of false-positive and false-negative results, the search words were updated as many times as feasible. The final search terms and the outcomes are displayed in Table 3.

Table 3: Search string construction

Databases	*Search Terms*	Results
PubMed	*(e-learning OR "online learning" OR "distance learning" OR "virtual learning" OR "digital learning" OR "web-based learning") AND ("performance evaluation" OR "performance assessment" OR "performance appraisal" OR "Course evaluation" OR Evaluation OR "performance-based assessment") AND ("medical education" OR "medical training" OR "Tele education" OR Telemedicine OR "medical school" OR "medical students" OR "medical curriculum" OR "medical student education" OR "clinical education"))*	1375
PUBMED MeSH	*(("education, distance"[MeSH Terms] OR online education[Text Word] NOT "blended learning") AND (Performance[All Fields] AND evaluation[All Fields]) AND ("education, medical"[MeSH Terms] OR Medical Education[Text Word]))*	112
Scopus	*TITLE-ABS-KEY ((e-learning OR "online learning" OR "distance learning" OR "virtual learning" OR "digital learning" OR "web-based learning") AND ("performance evaluation" OR "performance assessment" OR "performance appraisal" OR "Course evaluation" OR Evaluation) AND ("medical education" OR "medical training" OR "tele education" OR telemedicine) AND (universities OR colleges OR "medical school"))*	450
Web of Science	*(e-learning OR "online learning" OR "distance learning" OR "virtual learning" OR "digital learning" OR "web-based learning") AND ("performance evaluation" OR "performance assessment" OR "performance appraisal" OR "Course evaluation" OR Evaluation) AND ("medical education" OR "medical training" OR "tele education" OR telemedicine) AND (universities OR colleges OR "medical school")*	190
EBSCOHost	*(online learning OR e-learning OR distance learning) AND (performance evaluation OR performance assessment OR performance-based assessment OR student engagement OR learner engagement OR student participation) AND (medical education OR medical school OR medical students OR medical curriculum OR medical student education OR clinical education)*	1312
	Total	3439

2.6 Inclusion and Exclusion Criteria

After identifying the relevant studies, the inclusion and exclusion criteria were defined and applied to arrive at the final studies included in the synthesis. Table 4 shows the exclusion criteria for the study.

Table 4: Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
• Primary Studies • Studies that evaluate students' performance in clinical subjects. • Studies that evaluate students' performance in e-LMED within a postgraduate or graduate university setting.	• Studies not written in English language. • Studies evaluating student performance in a blended learning context • Studies with no full-text access • Studies on challenges and solutions of facilitators in e-LMED • Studies that assess e-learning programs but not students' performance in e-LMED • Studies not done in a medical school setting • Studies whose main objective is to compare performance in e-learning to face-to-face learning • Studies evaluating students' satisfaction or perception only • Studies that do not assess the clinical aspects of performance • Studies evaluating e-learning platforms in medical education • Studies whose main objective is to describe the adoption or utilization of technology

2.7 Data Extraction and Screening

The retrieved studies were extracted from different databases in Excel format and exported directly to Endnote to keep track of the references. 3439 studies were exported to Endnote from the databases queried. 353 duplicates were identified by Endnote and deleted, after which 3086 studies were exported to Rayyan for quality assessment and screening. Rayyan further identified 167 duplicates during the first screening session. These duplicates were carefully examined and deleted after confirmation. Hence, 2919 studies were analyzed using the inclusion and exclusion criteria on Rayyan. After the first round of screening to eliminate duplicates, abstract screening was conducted. All 2919 studies were screened according to the inclusion and exclusion criteria.

Studies excluded for being the wrong Population were mainly those that evaluated blended learning instead of e-learning and those that evaluated e-learning outside of the medical education context. Studies excluded for having the wrong outcome included those that evaluated e-learning in medical education but focused on other outcomes rather than student performance. Some studies focused on outcomes such as stress and anxiety level, teacher efficacy, perception, and attitude only, while others conducted program evaluations. Studies excluded for being the wrong study design include those that compared e-learning to traditional or blended learning, systematic literature reviews, and those evaluating students' and facilitators' performance. The wrong publication type describes studies such as letters to the editor, while foreign language describes studies excluded for not being published in English. 2476 studies were excluded in this round of screening. The third and final round of screening was conducted on the full text of the 443 reviewed literature. The screening discovered 173 studies had the wrong objectives, 108 were not conducted in a University setting, 26 had no full-text access, and 12 had wrong outcomes. Hence, the final studies included and analyzed were 103. Figure 2 shows the PRISMA diagram used to depict the screening process of this study.

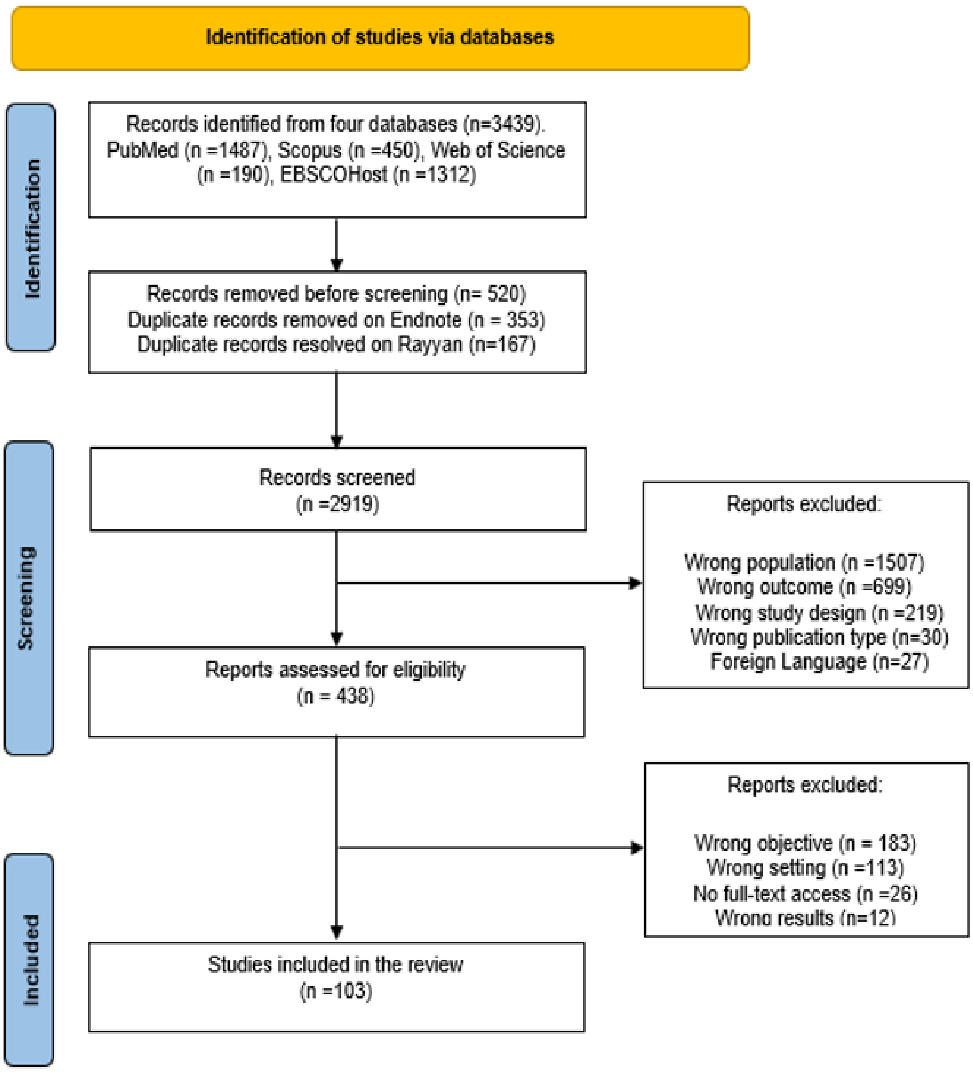

Figure 2: PRISMA Diagram

2.8 Quality Assurance

Quality assurance was conducted to ensure that only relevant studies were included. The second author and third authors triangulated the search terms developed by the first author. All authors agreed on the search terms used and on the quality of data to be extracted beforehand.

Where there were conflicts in the included or excluded articles, another collaborator was added to Rayyan to peruse the literature's full text together over a conference call with all the authors and resolved the conflicts using the inclusion and exclusion criteria stipulated for the study. Figure 3 shows the collaboration of all authors and the conflicts identified and resolved on Rayyan.

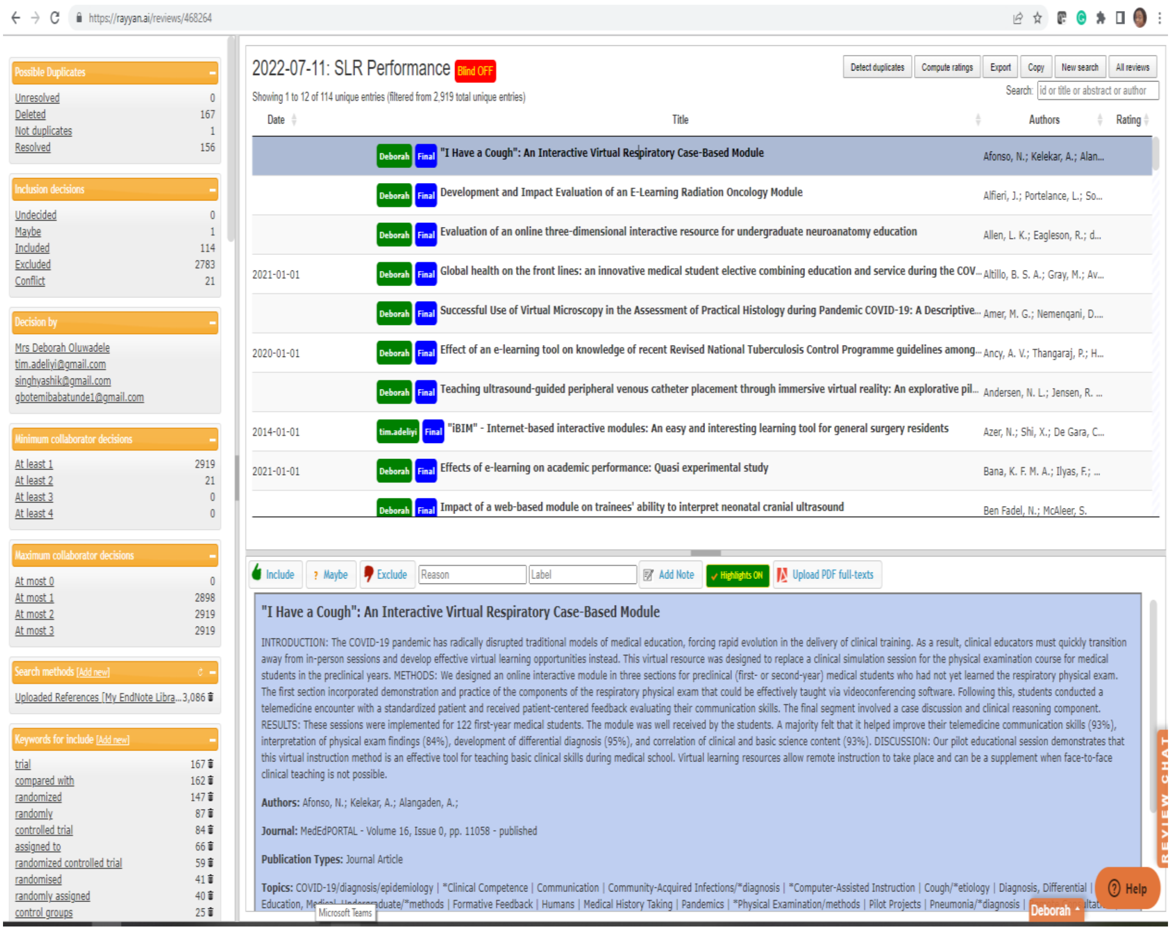

Figure 3: Quality assurance process on Rayyan

3. Results and Analysis

3.1 Overview of the Publication

103 (3%) studies met the inclusion criteria and were reviewed. Figure 4 shows the historical distribution of reviewed literature. The reviewed literature was published in the last two decades (2002-2022); the highest number of publications was in 2021 when 16 studies were published. The domain experienced sporadic growth between the years 2020 and 2022. This feat could be attributed to the emergence of COVID-19 and the adoption of e-learning to mitigate the lockdown restriction. Figure 5 shows the distribution of the types of publications of the reviewed literature.

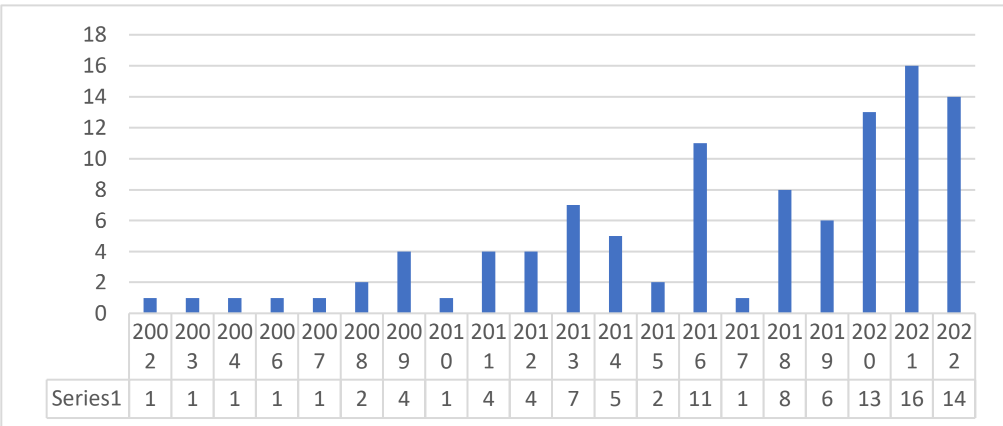

Figure 4: Historical distribution of reviewed literature

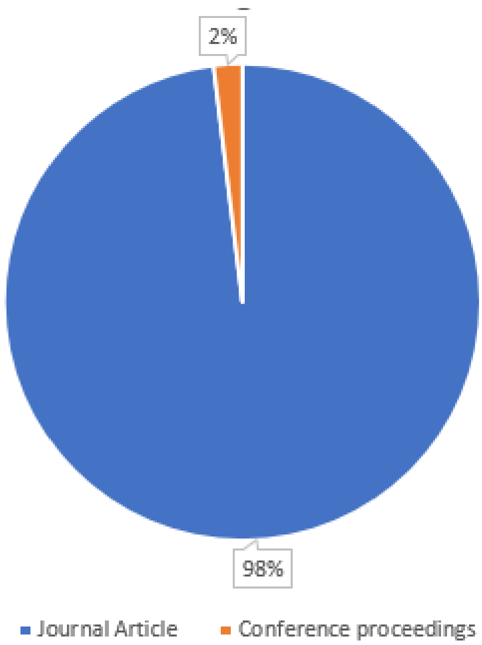

Figure 5: Publication types of the reviewed literature

The extracted data was tabulated on an Excel spreadsheet, with sheets created to analyze each research question. This section presents the analysis of all the research questions posed for this study.

RQ1 - How is performance operationalized in literature in e-LMED?

This RQ seeks to define student performance in e-LMED. The justification for RQ1 is to align the perspective of researchers to the terminologies or constructs used by authors in the domain to describe student performance. These constructs are reflected in the aim of the reviewed studies. The 103 reviewed literature used 30 constructs to operationalize performance (Figure 6). Thirty-three studies operationalize performance as knowledge. These authors evaluated students' performance by analyzing knowledge gain ((Allen, Eagleson and de Ribaupierre, 2016, Amer and Nemenqani, 2020, Andersen et al., 2021, Bergman et al., 2016, Bernardo et al., 2004, Bhat et al., 2022, Bracken et al., 2021, Brewer et al., 2016, Ens, Janzen and Palmert, 2016, Ganji et al., 2022, Gaupp, Körner and Fabry, 2016, Gillan et al., 2018, McCoy et al., 2020, Yilmaz et al., 2021) knowledge Change (Elbeddini and Tayefehchamani, 2021), knowledge acquisition (Fransen et al., 2018), knowledge retention (Lorenzo-Alvarez et al., 2019), knowledge transfer (Krahe et al., 2020) knowledge improvement (Camargo et al., 2014), learning outcomes (Choi-Lundberg, Cuellar and Williams, 2016, Krahe et al., 2020, Kulier et al., 2008, Liu et al., 2021), and academic performance (Gupta et al., 2020, McLaughlin et al., 2013, van Bonn et al., 2022) in students who underwent a specific e-learning intervention by evaluating the test or exam scores.

The next popular term used to operationalize performance is effectiveness. Twenty-seven authors operationalized performance by evaluating the effectiveness of e-learning in improving students' knowledge and application of key concepts or skills (Afonso, Kelekar and Alangaden, 2020, Alfieri et al., 2023, Burnette et al., 2009, Chang Chan et al., 2019, Chua et al., 2022, Corrigan et al., 2012, Darici et al., 2021, Falusi et al., 2022, Felder, Fauler and Geiler, 2013, Fuji and Galt, 2015, Helms et al., 2009, Kukolja-Taradi et al., 2008, Nomura et al., 2021, Onyeka et al., 2020, Puljak and Sapunar, 2011, Saiboon et al., 2021, Schilling et al., 2006, Schneider, Albers and Muller-Mattheis, 2015, Sichani, Mobarakeh and Omid, 2018, Sikkens et al., 2018, Smolle, Prause and Smolle-Jüttner, 2007, Viteri Jusué et al., 2020, Waugh et al., 2022, Webb and Choi, 2014, Yilmaz et al., 2021). Terms such as module effectiveness, learning effect, and simply effectiveness were clustered together regardless of their terminology. The effectiveness of the e-learning interventions was evaluated by using students' scores to determine their performance. Other popular constructs used to operationalize performance include impact, perception, and attitude. Notably, constructs that hinge on the electronic components of e-learning were not popularly considered in the operationalization of performance; these include accessibility, confidence, user experience, and user interface. These constructs have a frequency of less than 10% of the cumulative frequency of the 30 constructs.

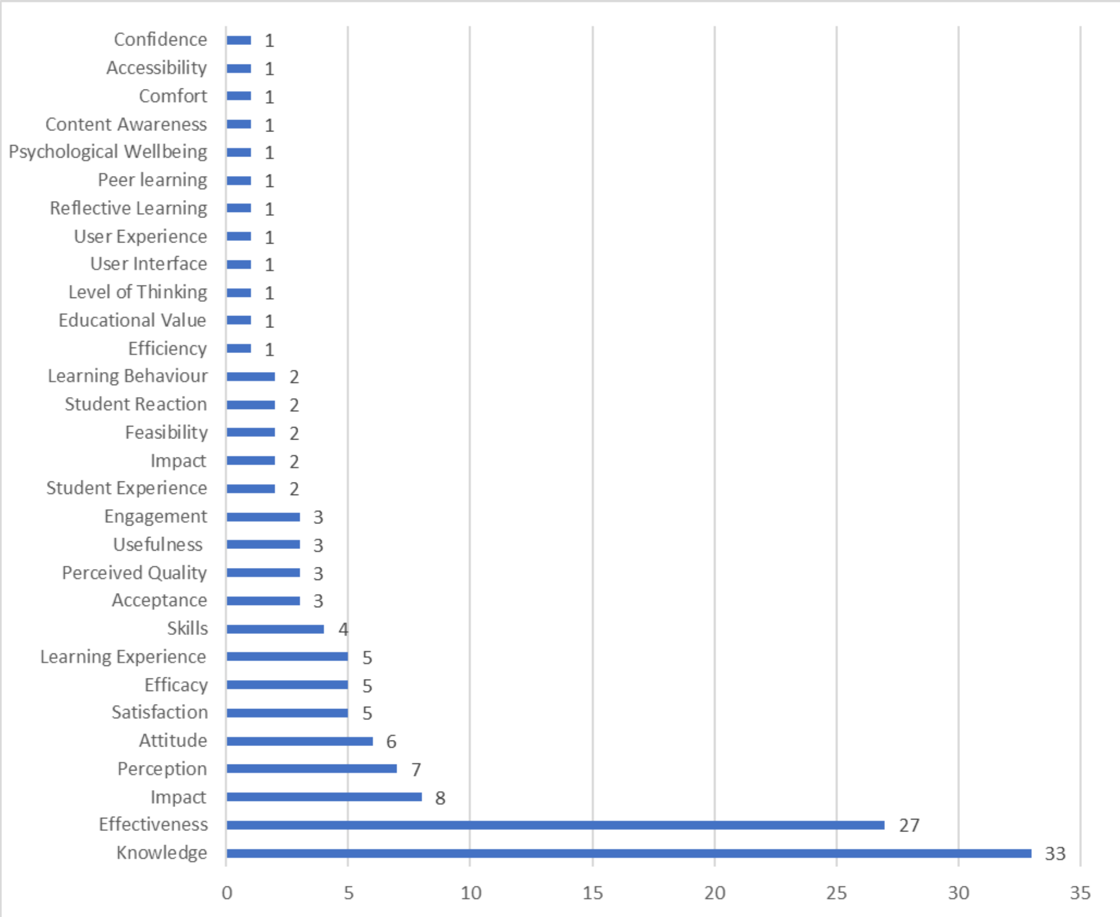

Figure 6: Constructs used to operationalize performance

RQ2. What are the metrics or factors measured when evaluating performance in e-LMED?

This question helps us identify the factors or metrics measured when evaluating student performance. Performance metrics were used to track and measure the achievement of predefined learning outcomes or specified as aims of the studies. Performance factors provide essential insights into what works and what needs to be improved, modified, or changed to ensure the achievement of objectives. An analysis of these factors could impact the perception of quality and the necessary gaps to be filled. Authors of the reviewed literature neither referred to the variables measured as factors or metrics nor made reference to the distinction between the two concepts. Different metrics were also used to evaluate performance, and most of these studies did not focus on how these factors influence the evaluation outcomes of these metrics.

The reviewed literature used 77 factors and metrics. These factors had a cumulative frequency of 311. Knowledge gain is the most common metric used by 81% of the reviewed literature and had a frequency of 83 (27%). The next factor with a relatively high frequency was satisfaction. 26% of the reviewed literature measured student satisfaction to understand performance. This metric had a frequency of 27 (9%). Other factors with higher frequencies include learning outcome (n=15), clinical skills (n=13), attitude (n=12), usefulness (n=11), and confidence (n=10). Out of the 77 metrics used to evaluate performance, only 14 (18%) of the metrics/factors were related to the electronic dimension of e-learning. This includes metric/factors such as adoption (Yilmaz et al., 2021), implementation (Yilmaz et al., 2021), maintenance (Yilmaz et al., 2021, Sekine et al., 2022), internet access (Sekine et al., 2022), access to materials (Choi-Lundberg, Cuellar and Williams, 2016, Gupta et al., 2020, Stevens et al., 2019), systems thinking (Gaupp, Körner and Fabry, 2016), familiarity with computers (Diessl et al., 2010), availability (Bridge, Jackson and Robinson, 2009), course functionality (Bernardo et al., 2004), ease of navigation (McLaughlin et al., 2013, Tan, Ross and Duerksen, 2013), ease of use (Elbeddini and Tayefehchamani, 2021, Webb and Choi, 2014), user interface (Corrigan et al., 2012, Gillan et al., 2018), usage (Choi-Lundberg, Cuellar and Williams, 2016, Corrigan et al., 2012, Gupta et al., 2020), and usability (Cipriano et al., 2013, Felder, Fauler and Geiler, 2013, Rusingiza et al., 2022, Webb and Choi, 2014).

Of all the studies evaluated, only Zare-Bidaki et al. (2022) followed a systematic approach to evaluation by defining the study context, applying an evaluation framework (RE-AIM) to evaluate the effectiveness of the course, defining the data sources and the metrics used, and most importantly, using technical data analytics techniques (Google Analytics and Moodle Learning Management System). Choi-Lundberg, Cuellar and Williams (2016) developed Dissection Audio-visual Resources (DAVR) to improve medical students' preparation for learning from dissection sections. The study analyzed data from the e-learning management system and survey, using factors related to the electronic dimension of e-learning, such as access and usage, to identify the underlying cause of performance in the metric evaluated (student scores). Corrigan et al. (2012) also paid attention to technological factors such as usability and user interface and their impact on performance metrics such as satisfaction and knowledge acquisition in dental education. Notably, Gupta et al. (2020) also seek to explain performance evaluation metrics (course examination score and satisfaction) via factors related to the e-electronic component of e-learning, such as usage rate and access to the adaptive e-learning platform-firecracker. The study affirmed a positive correlation between usage and improved cardiovascular system performance during a Bachelor of Medicine, Bachelor of Surgery (MBBS) program. Webb and Choi (2014) used "tracking web usage data" – a faculty's standard audit for online learning resources, combined with pre-and post-tests, summative course assessment, and questionnaires to investigate the impact of technological factors such as usability and ease of use on performance metrics such as knowledge gain and knowledge transfer to clinical practice.

Figure 7 provides the factors/metrics used to evaluate performance in the reviewed literature.

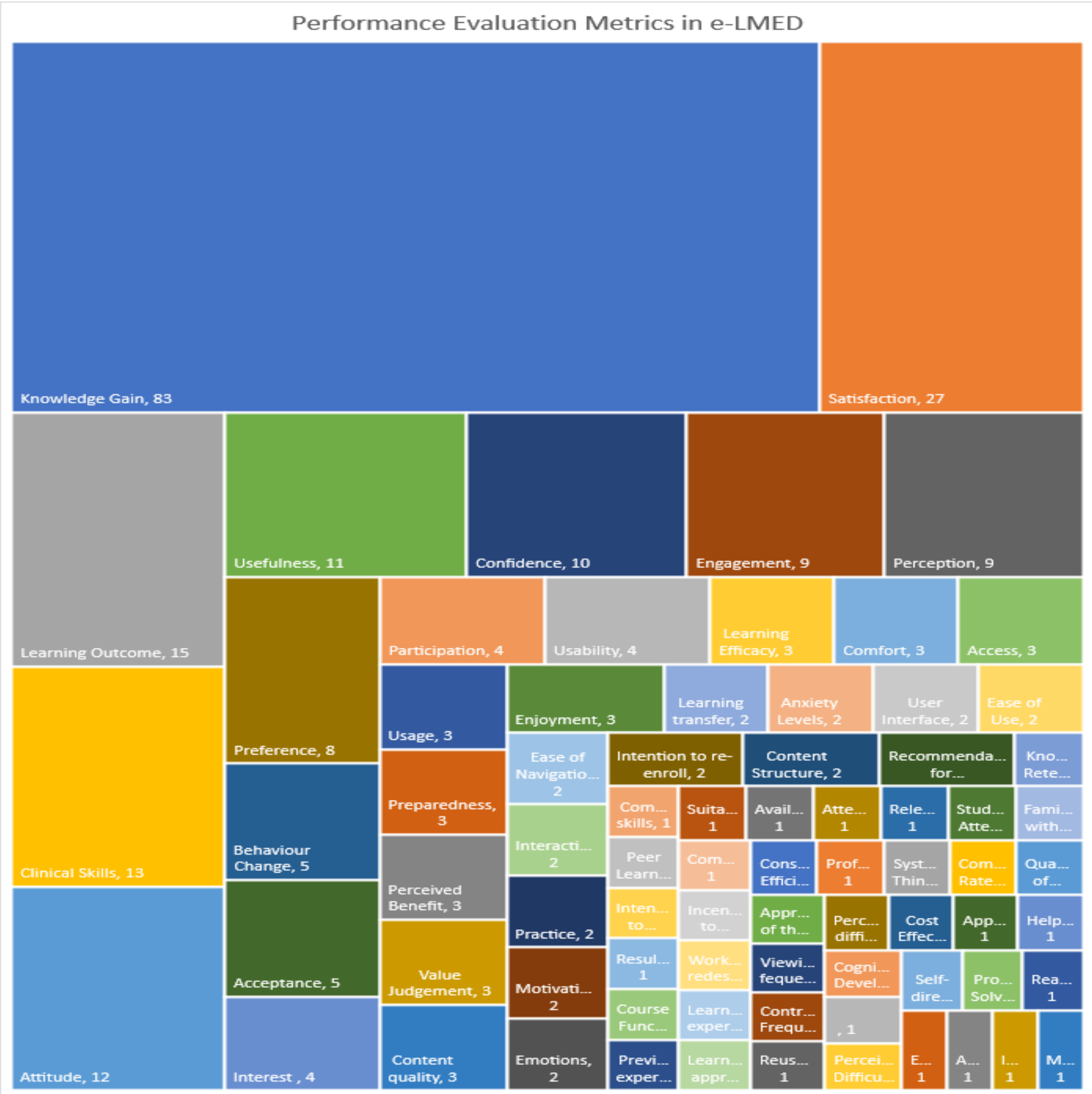

Figure 7: Metrics and factors used to measure performance

RQ3. What are the methods used to evaluate performance in e-LMED?

This RQ identifies the research design, tool, and techniques used to evaluate performance in e-LMED. Most studies did not specify whether they used quantitative, qualitative, or mixed methods. However, inference can be made from the instruments used for data collection that most studies used mixed methods, and the next most common method is quantitative methods, while an insignificant percentage used qualitative methods. Much of the reviewed literature used randomized control trials to examine the performance outcome of the intervention group versus the control group. The most common way to measure performance is through surveys (n=67). Next is the post-test (n=57), pretest (n=47), and exam (n=13). Other contemporary methods such as Google Analytics (Yilmaz et al., 2021), tracking data (Webb and Choi, 2014), digital dashboard data (Gupta et al., 2020), and online evaluation system (Liu et al., 2021) were identified, even though they were uncommon. Figure 8 presents the methods used to evaluate performance in e-LMED.

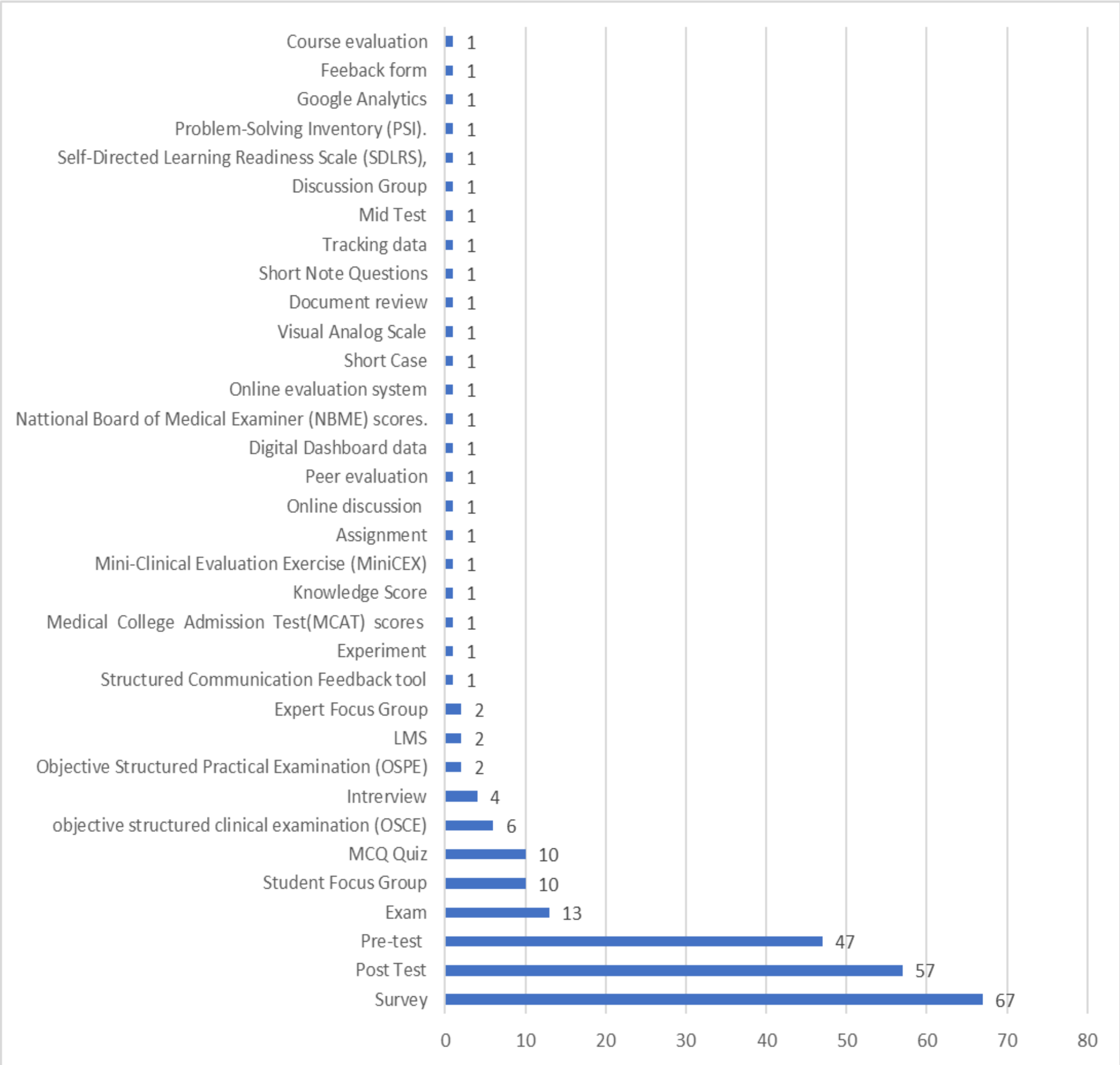

Figure 8: Research Methods used to evaluate performance in e-LMED

RQ4 - What are the models, theories, or frameworks used to evaluate performance in e-LMED? This section highlights the framework, models, or theories used as guiding principles to define constructs and inform metrics for evaluating student performance in e-LMED.

Theoretical approaches are used more frequently in implementation science to understand better and explain how and why implementations succeed or fail. Nilsen (2020) proposed a taxonomy that distinguishes between various theories, models, and frameworks in implementation science to make it easier to choose and use the most appropriate methods in implementation research and practice and encourage interdisciplinary communication among implementation researchers. Five different types of theoretical approaches are employed in implementation science. These are Process models, Determinant frameworks, Classic theories, Implementation theories, and Evaluation frameworks (Nilsen, 2020). The use of theories, models, and frameworks in implementation science could be categorized into three main goals: (1) describing and directing the process of putting research into practice; (2) comprehending and illuminating what factors affect implementation outcomes; and (3) evaluating implementation.

E-learning evaluation frameworks provide researchers with a structure to evaluate the success of e-learning implementation by explicitly measuring the concepts and constructs. They also help identify potential inhibitors and facilitators that must be addressed while guiding the implementation process.

Of the 103 studies, only six authors used a performance evaluation framework to guide their evaluation process. Almost all authors used the Kirkpatrick Evaluation Model. Ben Fadel and McAleer (2020), Naidoo et al. (2021), Zare-Bidaki et al. (2022) used a performance evaluation framework to guide their evaluation process. They also used the instructional design model; ADDIE (Analyze, Design, Develop, Implement, and Evaluate) Instructional design model to guide the design of their e-learning interventions. Zare-Bidaki et al. (2022) further used Kern's Six Step Curriculum development approach model. Table 5 details studies that used the performance evaluation model.

Table 5: Models used for performance evaluation of e-LMED

Author's Code	Authors	Title	Models
A2	Alfieri et al., 2012	Development and Impact Evaluation of an E-Learning Radiation Oncology Module	Kirkpatrick Evaluation Framework
A10	Ben Fadel & McAleer,2020	Impact of a web-based module on trainees' ability to interpret neonatal crania ultrasound	ADDIE Instructional design model, Kirkpatrick Evaluation Model
A43	Halawa et al., 2017	Distance Learning in Clinical Transplantation; A Successful Model in Post-graduate Education	Kirkpatrick Evaluation Model
A67	Moreira et al. 2019	Learner's perception, knowledge and behaviour assessment within a breast imaging E-Learning course for radiographers	Kirkpatrick's evaluation Model.
A69	Naidoo et al., 2021	Design, Implementation, and Evaluation of a Distance Learning Framework to Adapt to the Changing Landscape of Anatomy Instruction in Medical Education During COVID-19 Pandemic: A Proof-of-Concept Study	Analyse, Design, Develop, Implement, and Evaluate (ADDIE - a distance learning theory) model. Kirkpatrick Evaluation Model. Also Garrison's community inquiry, Siemens' connectivism and Harasim's online-collaborative-learning; and improved using Anderson's DL-model.
A103	Yilmaz et al., 2021	RE-AIMing COVID-19 online learning for medical students: a massive open online course evaluation	RE-AIM (reach, effectiveness, adoption, implementation, and maintenance) framework "analysis, design, development, implementa-tion, and evaluation" (ADDIE) instructional design model with Kern's six-step curriculum development ap-proach,

RQ5 - What gaps exist in performance evaluation in e-LMED?

This question analyzes the areas for improvement in performance evaluation in e-LMED through the synthesis of all the findings from RQ1-RQ4.

The top constructs used to define student performance in e-LMED include knowledge, effectiveness, impact, and perception. The analysis of these constructs shows that very little attention is paid to evaluating student performance using technology constructs. While learning is the subject of attention, the electronic components of e-learning seem to have been forgotten. It could be implied that authors are not paying enough attention to how facilitating learning via electronic applications could impact student performance. For example, of the thirty constructs used to operationalize performance, only about six constructs relate to the technology aspect of e-learning. These constructs have a frequency of less than 10% of the total cumulative frequency of 134. Furthermore, e-learning in medical education seems to not thrive past the use of e-learning courses. The domain appears not to explore more systems, applications, and technologies that expand technology use cases in medical education. Another critical gap that exists and needs to be filled is that the reviewed literature rarely used a performance evaluation framework to guide their evaluation process. The few (6 studies) that used a performance evaluation framework mainly used the Kirkpatrick evaluation model. While this model is quite popular and widely accepted, it has been heavily critiqued for its inflexibility in evaluating technology-enabled learning (Oluwadele, Singh and Adeliyi, 2022). This is because the model presents metrics that can be measured to guide student performance evaluation but does not provide factors that could impact these metrics or the metrics that can evaluate the technology component of e-learning.

Additionally, performance evaluation is conducted haphazardly. Authors do not focus on metrics that evaluate the technology components of e-learning – we coin this phenomenon "evaluating e-learning minus the e." Authors evaluate performance in e-learning the same way they would traditional learning. Noesgaard and Ørngreen (2015) had identified this pattern earlier and questioned whether e-learning and face-to-face learning should be defined and evaluated the same way. Their study recommended that future researchers and designers critically consider the identified definitions, measures, and factors when designing effective e-learning. Very few of the reviewed literature endeavored to map out how the use of technology impacted student performance based on the metrics used. Therefore, it isn't easy to appreciate the impact of technology factors on the performance metrics measured. Conclusively, several methods are being used to evaluate the performance of students in e-LMED. However, there is heavy reliance on surveys, pre-and post-test. A negligible percentage of authors of the reviewed literature used contemporary data collection and analysis methods. Although Oluwadele, Singh and Adeliyi (2023) revealed that authors in the e-LMED domain are actively collaborating, the rate of collaboration with authors from the computer science domain was affirmed to be very low. This might also explain why the technology component of e-learning is poorly explored in the reviewed literature. Cross-disciplinary collaboration between Medical Education and computer science domains will ensure that the trends in performance evaluation in e-learning in medical education are evaluated using more advanced data analysis and visualization tools to better understand patterns and trends in the data.

4. Discussion

Medical education and the health professions are confronted with a slew of global difficulties that are context-specific yet recurring. These difficulties have been defined as wicked problems that defy recognized solutions and are perceived differently by different people (Mennin, 2021). Mennin (2021) highlighted ten challenging issues in medical education. Four of these issues are relevant to the context of and resonate with the findings of this study. Firstly, the decomposition of the medical curriculum due to the pandemic and its subsequent impact of migration to online learning on students' opportunities for clinical encounters. Secondly, the quality of teaching as a result of the move to online learning and its rebounding impact on the validity and quality of performance-based assessments. Thirdly, the inflexibility of the domain to change is evident in the challenges experienced by clinicians for their continued professional development, despite all the advancements in the domain. And lastly, the difficulty in collaboration due to the setup of the domain. Collaboration is an alien term because "the currency of health professionals is an independent practice, and in research, training is about being an independent investigator. Assessment of learners is based on individual performance and the extent to which learners can function independently. Promotion is based on individual achievement". The issues highlighted above encapsulate and perfectly express the findings of this review and unveil the "elephant in the room."

Operationalization is the process of transforming an abstract concept into a precise definition that is measurable and testable. There is no standard definition for the performance evaluation of e-learning in medical education or a standardized guideline for the metrics, methods, or framework for guiding the implementation. Operationalization helps define the variables of interest. Without clear and precise operational definitions, researchers run the risk of measuring unrelated concepts or using different methods every time. Operationalization decreases subjectivity, reduces the chance of bias in the research, and improves the validity of studies. The reviewed literature used 30 constructs to operationalize performance; 32% (33 studies) operationalized performance with knowledge, 26% (27 studies) operationalized performance with effectiveness, and 8% (8 studies) operationalized performance with impact. 8% of the reviewed literature did not operationalize performance at all. These three constructs are the key constructs used by 66% of authors of the reviewed literature to evaluate performance. The 27 other constructs account for 34% of the total constructs identified.

Similarly, the reviewed literature seemed to randomly adopt metrics and factors to measure performance with little thought to the suitability of the metrics for the constructs being measured. Akin to this is the fact that a thorough attempt is not made to highlight a cause-effect relationship between factors and metrics. For instance, how does internet availability impact students' perception of or satisfaction with an e-learning intervention? Or how does the language of instruction impact the achievement of learning outcomes? Instead, the achievement of learning outcomes is evaluated and reported directly without seeking to understand what factors helped students to achieve these learning outcomes or otherwise. The trend of lack of traceability and synch between constructs, factors, and metrics raises a plethora of rhetorical questions, the chief of which is quality. How is the quality of publications evaluating e-LMED established? The next is how do we establish the critical success factors of e-learning in medical education from a consolidation of empirical evidence of publications in the domain? The answers to these questions will assist stakeholders and e-learning developers in understanding the factors to cater for during implementation and to measure during evaluation. Therefore, findings and feedback from every evaluation process and study will become a valuable feed-forward tool. This cycle will initiate increased growth in the domain.

E-learning evaluation frameworks provide researchers with specific concepts to measure, and structures to assess how well e-learning has been implemented. They assist in directing the implementation process and aid in the identification of potential facilitators and inhibitors that need to be addressed. However, only 6% (6 studies) of the reviewed literature used an evaluation framework to guide their performance evaluation process, while 5 out of these 6 studies used the Kirkpatrick Evaluation Model (Kirkpatrick, 1975). The Kirkpatrick evaluation model has four levels of evaluations of training outcomes: reaction, learning, behavior, and result. The model proposes that the effectiveness of training can be evaluated by measuring students' reactions to the training, the level of knowledge and skills acquired from the training, the transfer of knowledge and skills gained from the training to practices at the workplace, and the measurable impacts of the transferred knowledge and skills in participant's workplace (Oluwadele, Singh and Adeliyi, 2022).

Although Kirkpatrick presents clear concepts that can be used to evaluate training outcomes or student performance, these concepts have no bearing on the technology dimension of training and are, therefore, not the most suitable model for evaluating performance in an e-learning context. The model has been criticized by many scholars (Reio et al., 2017). One of the critiques of the Kirkpatrick evaluation model relevant to the context of this study is that the model is hierarchical, and the levels are placed in ascending order. The higher levels are therefore perceived as more useful and significant than the lower ones (Alliger and Janak, 1989). This reservation holds for almost all studies that used the Kirkpatrick evaluation model in the reviewed literature. For instance, Naidoo et al. (2021) only evaluated the first two levels of the Kirkpatrick evaluation model, Alfieri et al. (2023), and Moreira et al. (2018) evaluated the first three levels of the model but reported that evaluating the fourth level (impact of knowledge transfer on medical students' work organization) was not feasible, hence beyond their scope of the study. Halawa et al. (2017) used Phillips (1998) version of the Kirkpatrick evaluation model, which added a fifth concept of Return on Investment (ROI) to evaluate if the training provided a positive ROI; however, the study specified that trying to evaluate how learning changed behavior was difficult and impractical. (i.e., at Kirkpatrick level 3) because it would be challenging for the instructors to objectively assess the students at their places of employment. Only Ben Fadel and McAleer (2020) evaluated the four levels of the Kirkpatrick evaluation model.

The standard metrics used for performance evaluation are primarily focused on learning outcomes. A mental clustering of the top ten metrics for evaluating performance in e-LMED as identified in the reviewed literature; knowledge, satisfaction, learning outcome, clinical skills, attitude, usefulness, confidence, engagement, perception, and preference suggests that they are a perfect fit for measuring only the top three levels of Kirkpatrick evaluation model. These top ten metrics have 63% of the total frequency of all 77 constructs. Hence, we can infer that most of the reviewed literature only evaluated the first three levels of the Kirkpatrick evaluation model. Ruggeri, Farrington and Brayne (2013) established this fact by attesting that the narrow scope of current evaluations is a significant obstacle to comprehending the effects of e-learning interventions in medical education. The authors found it concerning that the scope of evaluation in present e-learning for health is frequently restricted to user satisfaction and enjoyment.

Given these points, it is recommended that authors evaluating performance in e-LMED use the steps below as a guide or roadmap for their evaluation process as presented in Figure 9:

1. Identify the constructs you intend to study (for instance -knowledge),
2. Choose the variables to delimit the constructs (for example - knowledge gain, knowledge retention, or Knowledge transfer)
3. Specify the indicators or metrics to measure the specified variables (learning outcome, completion rate, skills, engagement rate)
4. Explore mediating factors (including technology factors) that could influence the outcome of the measured metrics. (For instance – what is the impact of the "user interface" on engagement rate, completion rate, and learning outcome)
5. Choose the most suitable method to measure the metrics (For instance, a quantitative method using pre-and post-test, exams, learning analytics to measure knowledge gain, and qualitative focus group discussion or open-ended survey to measure knowledge retention or knowledge transfer)
6. Adopt and leverage performance evaluation models to guide constructs and metrics to evaluate using a top-down or bottom-up approach. Otherwise, critique the limitations of existing models and propose a new model.

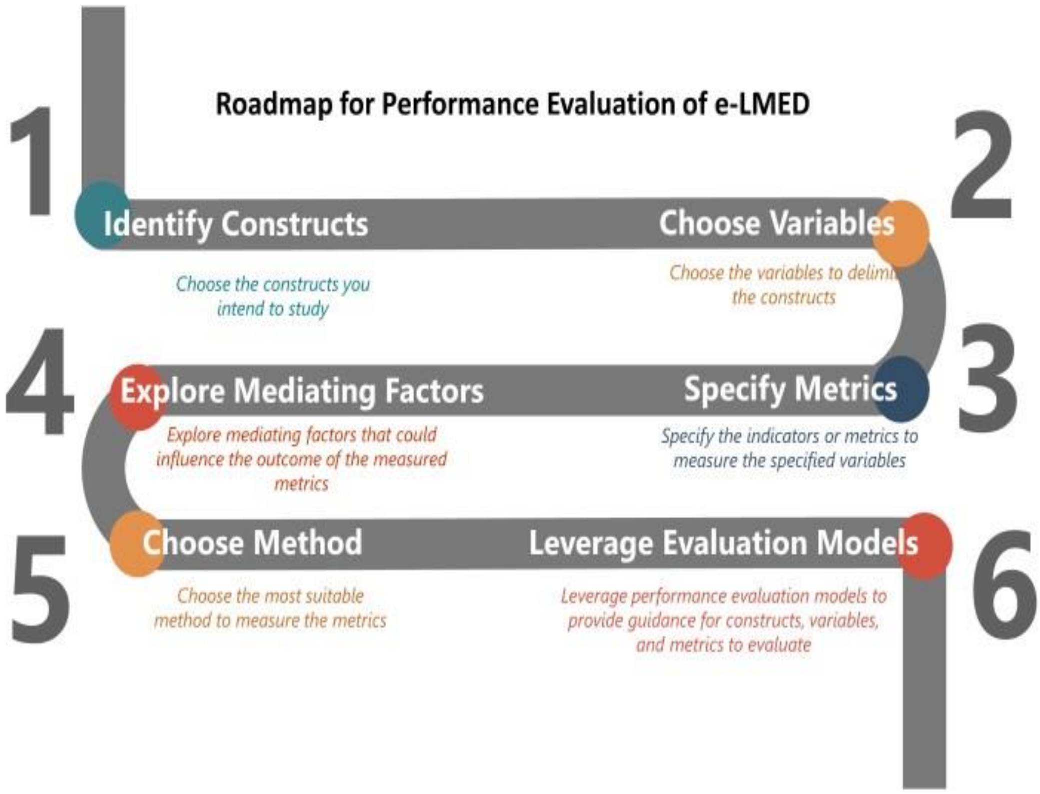

Figure 9: Proposed roadmap for performance evaluation of e-LMED

As e-learning adoption widens in higher education institutions, the issue of quality assurance continues to grow. Concerns are expressed concerning performance evaluation models that are solely based on pedagogical opinions rather than software perspectives (Farhan, Talib and Mohammed, 2019). To fill this gap, the construct-variable-metric-factors-method-model alignment matrix would be used to develop a Requirement Traceability Metrix (RTM) as a possible quality assurance technique for peer review of publications in e-LMED in the future.

5. Conclusion

Despite the ubiquity of research evaluating students' performance in e-LMED, the domain is yet to achieve a consensus on the definition of performance or the suitable constructs, metrics, models, and methods to help understand student performance. Through a systematic review, this study put forward a working definition of what constitutes performance evaluation to reduce the ambiguity, arbitrariness, and multiplicity surrounding performance evaluation of e-learning in medical education. This study answers these research questions: How is performance operationalized in literature in e-LMED? What are the metrics and factors measured when evaluating performance in e-LMED? What are the methods used to evaluate performance in e-LMED? What are the models, theories, or frameworks used to evaluate performance in e-LMED? What gaps exist in performance evaluation in e-LMED?

The study revealed that knowledge is the most prominent term for operationalizing performance. Most studies used knowledge gain and students' satisfaction as metrics to measure the knowledge acquired from e-learning courses using surveys and pre-and post-tests. Very few studies measured technology-related metrics or explored the impact of technology-related factors on the outcome of the metrics measured. Similarly, evaluation models were barely used to guide the evaluation process. The few studies that used evaluation models use the Kirkpatrick evaluation model, which has been famously critiqued as unsuitable for evaluating e-learning because the constructs presented by the model are unrelated to technology.

One of the many questions raised in the discussion was how the quality of publications evaluating e-LMED was assessed. Hence this study proposed a roadmap for performance evaluation to guide evaluation processes and suggests future research maps construct-variable-metric-factors-method-model as a Requirement Traceability Matrix for quality assurance of publications in e-learning in medical education.

Another important question is how to establish the critical success factors of e-learning in medical education from a consolidation of empirical evidence of publications in the domain. Hence, the authors suggest that future researchers and designers use modern data analytics techniques to identify the critical success factor of e-learning in medical education.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Afonso, N., Kelekar, A. & Alangaden, A. 2020. "I Have a Cough": An Interactive Virtual Respiratory Case-Based Module. MedEdPORTAL, 16, 11058.

Alfieri, J., Portelance, L., Souhami, L., Steinert, Y., Mcleod, P., Gallant, F. & Artho, G. 2023. Development and Impact Evaluation of an E-Learning Radiation Oncology Module. International Journal of Radiation Oncology Biology Physics, 82, E573-E580.

Allen, L. K., Eagleson, R. & De Ribaupierre, S. 2016. Evaluation of an online three-dimensional interactive resource for undergraduate neuroanatomy education. Anat Sci Educ, 9, 431-9.

Alliger, G. M. & Janak, E. A. 1989. Kirkpatrick's levels of training criteria: Thirty years later. Personnel psychology, 42, 331-342.

Amer, M. G. & Nemenqani, D. M. 2020. Successful Use of Virtual Microscopy in the Assessment of Practical Histology during Pandemic COVID-19: A Descriptive Study. J Microsc Ultrastruct, 8, 156-161.

Andersen, N. L., Jensen, R. O., Posth, S., Laursen, C. B., Jørgensen, R. & Graumann, O. 2021. Teaching ultrasound-guided peripheral venous catheter placement through immersive virtual reality: An explorative pilot study. Medicine (Baltimore), 100, e26394.

Bauer, M. S., Damschroder, L., Hagedorn, H., Smith, J. & Kilbourne, A. M. 2015. An introduction to implementation science for the non-specialist. BMC psychology, 3, 1-12.

Ben Fadel, N. & Mcaleer, S. 2020. Impact of a web-based module on trainees' ability to interpret neonatal cranial ultrasound. BMC Med Educ, 20, 489.

Bergman, J., Ballon-L, A, E., Lerman, S. E., Kwan, L., Bennett, C. J. & Litwin, M. S. 2016. Engaging Physician Learners Through a Web-Based Platform: Individualized End-of-Life Education. Am J Hosp Palliat Care, 33, 748-54.

Bernardo, V., Ramos, M. P., Plapler, H., De Figueiredo, L. F., Nader, H. B., Anção, M. S., Von Dietrich, C. P. & Sigulem, D. 2004. Web-based learning in undergraduate medical education: development and assessment of an online course on experimental surgery. Int J Med Inform, 73, 731-42.

Bhat, G. M., Bhat, I. H., Shahdad, S., Rashid, S., Khan, M. A. & Patloo, A. A. 2022. Analysis of Feasibility and Acceptability of an E-Learning Module in Anatomy. Anat Sci Educ, 15, 376-391.

Boland, A., Dickson, R. & Cherry, G. 2017. Doing a systematic review: A student's guide. Doing a Systematic Review, 1-304.

Bracken, K., Levinson, A. J., Mahmud, M., Allice, I., Vanstone, M. & Grierson, L. 2021. Spiralling Pre-clerkship Concepts into the Clinical Phase: Augmenting Knowledge Transfer Using Innovative Technology-Enhanced Curriculum Activities. Med Sci Educ, 31, 1-14.

Brewer, Z. E., Ogden, W. D., Fann, J. I., Burdon, T. A. & Sheikh, A. Y. 2016. Creation and Global Deployment of a Mobile, Application-Based Cognitive Simulator for Cardiac Surgical Procedures. Semin Thorac Cardiovasc Surg, 28, 1-9.

Bridge, P. D., Jackson, M. & Robinson, L. 2009. The effectiveness of streaming video on medical student learning: a case study. Medical education online, 14, 11.

Burnette, K., Ramundo, M., Stevenson, M. & Beeson, M. S. 2009. Evaluation of a web-based asynchronous pediatric emergency medicine learning tool for residents and medical students. Acad Emerg Med, 16, S46-50.

Cairó, O., Barreiro, J. & Solsona, F. The importance of metrics and evaluation processes in knowledge based systems. Proceedings of 2nd International Workshop on Computer Science and Information Technologies, 2000. Citeseer.

Camargo, L. B., Raggio, D. P., Bonacina, C. F., Wen, C. L., Mendes, F. M., Bönecker, M. J. & Haddad, A. E. 2014. Proposal of e-learning strategy to teach Atraumatic Restorative Treatment (ART) to undergraduate and graduate students. BMC Res Notes, 7, 456.

Chang Chan, A. Y., Custers, E., Van Leeuwen, M. S., Bleys, R. & Ten Cate, O. 2019. Does an Additional Online Anatomy Course Improve Performance of Medical Students on Gross Anatomy Examinations? Med Sci Educ, 29, 697-707.

Choi-Lundberg, D. L., Cuellar, W. A. & Williams, A. M. 2016. Online dissection audio-visual resources for human anatomy: Undergraduate medical students' usage and learning outcomes. Anat Sci Educ, 9, 545-554.

Chua, W. L., Ooi, S. L., Chan, G. W. H., Lau, T. C. & Liaw, S. Y. 2022. The Effect of a Sepsis Interprofessional Education Using Virtual Patient Telesimulation on Sepsis Team Care in Clinical Practice: Mixed Methods Study. J Med Internet Res, 24, e35058.

Cipriano, S. D., Dybbro, E., Boscardin, C. K., Shinkai, K. & Berger, T. G. 2013. Online learning in a dermatology clerkship: piloting the new American Academy of Dermatology Medical Student Core Curriculum. J Am Acad Dermatol, 69, 267-72.

Clark, R. C. & Mayer, R. E. 2016. E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning, john Wiley & sons.

Corrigan, M., Mchugh, S., Sheikh, A., Lehane, E., Shields, C., Redmond, P., Kerin, M. & Hill, A. 2012. Surgent University: The Establishment and evaluation of a national online clinical teaching repository for surgical trainees and students. Surgical Innovation, 19, 200-204.

Darici, D., Reissner, C., Brockhaus, J. & Missler, M. 2021. Implementation of a fully digital histology course in the anatomical teaching curriculum during COVID-19 pandemic. Annals of Anatomy, 236.

Deguchi, A., Hirai, C., Matsuoka, H., Nakano, T., Oshima, K., Tai, M. & Tani, S. 2020. What is society 5.0. Society, 5, 1-23.

Diessl, S., Verburg, F. A., Hoernlein, A., Schumann, M., Luster, M. & Reiners, C. 2010. Evaluation of an internet-based e-learning module to introduce nuclear medicine to medical students: a feasibility study. Nucl Med Commun, 31, 1063-7.

Elbeddini, A. & Tayefehchamani, Y. 2021. Development and evaluation of an online medication safety module for medical students at a rural teaching hospital: the Winchester District Memorial Hospital. BMJ Open Qual, 10.

Ens, A., Janzen, K. & Palmert, M. R. 2016. Development of an Online Learning Module to Improve Pediatric Residents' Confidence and Knowledge of the Pubertal Examination. J Adolesc Health, 60, 292-298.

Falusi, O. O., Weisz, J., Clarence, I., Lichtenstein, C., Coddington, D., Avent, G., Beers, L. & Ottolini, M. 2022. Leveraging E-Learning and Community Assets to "TEACH" Residents to Address Child Poverty. Acad Pediatr.

Farhan, M. K., Talib, H. A. & Mohammed, M. S. 2019. Key factors for defining the conceptual framework for quality assurance in e-learning. Journal of Information Technology Management, 11, 16-28.

Felder, E., Fauler, M. & Geiler, S. 2013. Introducing e-learning/teaching in a physiology course for medical students: acceptance by students and subjective effect on learning. Adv Physiol Educ, 37, 337-42.

Foy, R., Ovretveit, J., Shekelle, P. G., Pronovost, P. J., Taylor, S. L., Dy, S., Hempel, S., Mcdonald, K. M., Rubenstein, L. V. & Wachter, R. M. 2011. The role of theory in research to develop and evaluate the implementation of patient safety practices. BMJ quality & safety, 20, 453-459.

Fransen, F., Martens, H., Nagtzaam, I. & Heeneman, S. 2018. Use of e-learning in clinical clerkships: effects on acquisition of dermatological knowledge and learning processes. International journal of medical education, 9, 11-17.

Fuji, K. T. & Galt, K. A. 2015. An Online Health Informatics Elective Course for Doctor of Pharmacy Students. American Journal of Pharmaceutical Education, 79, 1-9.

Fukuyama, M. 2018. Society 5.0: Aiming for a new human-centered society. Japan Spotlight, 27, 47-50.

Galas, A., Pilat, A., Leonardi, M. & Tobiasz-Adamczyk, B. 2018. Research Project Evaluation—Learnings from the PATHWAYS Project Experience. International journal of environmental research and public health, 15, 1071.

Ganji, J., Shirvani, M. A., Motahari-Tabari, N. & Tayebi, T. 2022. Design, implementation and evaluation of a virtual clinical training protocol for midwifery internship in a gynecology course during COVID-19 pandemic: A semi-experimental study. Nurse Education Today, 111.

Gaupp, R., Körner, M. & Fabry, G. 2016. Effects of a case-based interactive e-learning course on knowledge and attitudes about patient safety: a quasi-experimental study with third-year medical students. BMC Med Educ, 16, 172.

Gillan, C., Papadakos, J., Brual, J., Harnett, N., Hogan, A., Milne, E. & Giuliani, M. E. 2018. Impact of high-fidelity e-learning on knowledge acquisition and satisfaction in radiation oncology trainees. Current Oncology, 25, e533-e538.

Gupta, S., Ojeh, N., Sa, B., Majumder, M. A. A., Singh, K. & Adams, O. P. 2020. Use of an Adaptive e-Learning Platform as a Formative Assessment Tool in the Cardiovascular System Course Component of an MBBS Programme. Adv Med Educ Pract, 11, 989-996.

Halawa, A., Sharma, A., Bridson, J. M., Lyon, S., Prescott, D., Guha, A. & Taylor, D. 2017. Distance Learning in Clinical Transplantation: A Successful Model in Postgraduate Education. World Journal of Education, 7, 74-78.

Helms, A., Denson, K., Brown, D. & Simpson, D. 2009. One Specialty at a Time: Achieving Competency in Geriatrics Through an e-Learning Neurology Clerkship Module. Academic Medicine, 84, S67-S69.

Kirkpatrick, D. L. 1975. Techniques for evaluating training programs. Evaluating training programs, 1-17.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J. & Linkman, S. 2009. Systematic literature reviews in software engineering–a systematic literature review. Information and software technology, 51, 7-15.

Krahe, A. M., Ketterer, M. C., Offergeld, C., Hildenbr & , T. 2020. [Evaluation of a structured e-learning-based approach to CT anatomy of the paranasal sinuses for medical students : A pilot study]. Hno, 70, 468-475.

Kukolja-Taradi, S., Dogas, Z., Dabić, M. & Drenjancevic Perić, I. 2008. Scaling-up undergraduate medical education: enabling virtual mobility by online elective courses. Croat Med J, 49, 344-51.

Kulier, R., Hadley, J., Weinbrenner, S., Meyerrose, B., Decsi, T., Horvath, A. R., Nagy, E., Emparanza, J. I., Coppus, S. F., Arvanitis, T. N., Burls, A., Cabello, J. B., Kaczor, M., Zanrei, G., Pierer, K., Stawiarz, K., Kunz, R., Mol, B. W. & Khan, K. S. 2008. Harmonising evidence-based medicine teaching: a study of the outcomes of e-learning in five European countries. BMC Med Educ, 8, 27.

Linnenluecke, M. K., Marrone, M. & Singh, A. K. 2020. Conducting systematic literature reviews and bibliometric analyses. Australian Journal of Management, 45, 175-194.

Liu, L. W., Yuan, S. Y., Zhang, W. B., Wang, Z. D., Zhao, C. Y., Pan, Y. C., Yan, B. & Wang, L. 2021. Development and assessment of an online virtual orthodontic curriculum. Journal of Dental Education, 86, 509-516.

Lorenzo-Alvarez, R., Rudolphi-Solero, T., Ruiz-Gomez, M. J. & Sendra-Portero, F. 2019. Game-Based Learning in Virtual Worlds: A Multiuser Online Game for Medical Undergraduate Radiology Education within Second Life. Anat Sci Educ, 13, 602-617.

Mccoy, L., Lewis, J. H., Simon, H., Sackett, D., Dajani, T., Morgan, C. & Hunt, A. 2020. Learning to Speak Up for Patient Safety: Interprofessional Scenarios for Training Future Healthcare Professionals. J Med Educ Curric Dev, 7, 2382120520935469.

Mclaughlin, J. E., Griffin, L. M., Esserman, D. A., Davidson, C. A., Glatt, D. M., Roth, M. T., Gharkholonarehe, N. & Mumper, R. J. 2013. Pharmacy Student Engagement, Performance, and Perception in a Flipped Satellite Classroom. American Journal of Pharmaceutical Education, 77, 1-8.

Melnyk, S. A., Stewart, D. M. & Swink, M. 2004. Metrics and performance measurement in operations management: dealing with the metrics maze. Journal of operations management, 22, 209-218.

Mennin, S. 2021. Ten global challenges in medical education: wicked issues and options for action. Medical Science Educator, 31, 17-20.

Meyers, D. C., Durlak, J. A. & Wandersman, A. 2012. The quality implementation framework: a synthesis of critical steps in the implementation process. American journal of community psychology, 50, 462-480.

Moreira, I. C., Ramos, I., Rua Ventura, S. & Pereira Rodrigues, P. 2018. Learner's perception, knowledge and behaviour assessment within a breast imaging E-Learning course for radiographers. Eur J Radiol, 111, 47-55.

Naidoo, N., Azar, A. J., Khamis, A. H., Gholami, M., Lindsbro, M., Alsheikh-Ali, A. & Banerjee, Y. 2021. Design, Implementation, and Evaluation of a Distance Learning Framework to Adapt to the Changing Landscape of Anatomy Instruction in Medical Education During COVID-19 Pandemic: A Proof-of-Concept Study. Frontiers in Public Health, 9.

Nilsen, P. 2020. Making sense of implementation theories, models, and frameworks. Implementation Science 3.0, 53-79.

Noesgaard, S. S. & Ørngreen, R. 2015. The effectiveness of e-learning: an explorative and integrative review of the definitions, methodologies and factors that promote e-learning effectiveness. Electronic Journal of E-learning, 13, 278-290.

Nomura, O., Irie, J., Park, Y., Nonogi, H. & Hanada, H. 2021. Evaluating Effectiveness of YouTube Videos for Teaching Medical Students CPR: Solution to Optimizing Clinician Educator Workload during the COVID-19 Pandemic. Int J Environ Res Public Health, 18.

Oluwadele, D., Singh, Y. & Adeliyi, T. T. A Hybridized Framework for Designing and Evaluating E-Learning Students' Performance in Medical Education. 2022 International Conference on Engineering and Emerging Technologies (ICEET), 2022. IEEE, 1-6.

Oluwadele, D., Singh, Y. & Adeliyi, T. T. E-Learning Performance Evaluation in Medical Education—A Bibliometric and Visualization Analysis. Healthcare, 2023. MDPI, 232.

Onyeka, T. C., Iloanusi, N., Namisango, E., Onu, J. U., Okunade, K. S., Mohammed, A. D., Gadanya, M. A., Nagoma, A. U., Ojiakor, S., Ilo, C., Okuzu, O., Oduche, C., Ugwu, N. & Allsop, M. J. 2020. Project OPUS: Development and evaluation of an electronic platform for pain management education of medical undergraduates in resource-limited settings. PLoS ONE, 15.

Phillips, J. J. 1998. Level four and beyond: An ROI model. Evaluating corporate training: Models and issues, 113-140.

Puljak, L. & Sapunar, D. 2011. Web-based elective courses for medical students: an example in pain. Pain Med, 12, 854-63.

Reio, T. G., Rocco, T. S., Smith, D. H. & Chang, E. 2017. A critique of Kirkpatrick's evaluation model. New Horizons in Adult Education and Human Resource Development, 29, 35-53.

Rubin, A. & Babbie, E. R. 2016. Empowerment series: Research methods for social work, Cengage Learning.

Ruggeri, K., Farrington, C. & Brayne, C. 2013. A global model for effective use and evaluation of e-learning in health. Telemedicine and e-Health, 19, 312-321.
Rusingiza, E., Alizadeh, F., Wolbrink, T., Mutamba, B., Vinci, S., Profita, E. L., Rulisa, S., Delsignore, L., Solis, J., Geggel, R. & Wilson, K. 2022. An e-learning pediatric cardiology curriculum for Pediatric Postgraduate trainees in Rwanda: implementation and evaluation. BMC Med Educ, 22, 179.
Saaiq, M. & Ashraf, B. 2017. Modifying "Pico" question into "Picos" model for more robust and reproducible presentation of the methodology employed in a scientific study. World Journal of Plastic Surgery, 6, 390.
Saiboon, I. M., Zahari, F., Isa, H. M., Sabardin, D. M. & Robertson, C. E. 2021. E-Learning in Teaching Emergency Disaster Response Among Undergraduate Medical Students in Malaysia. Front Public Health, 9, 628178.
Salkind, N. J. 2010. Encyclopedia of research design, sage.
Schilling, K., Wiecha, J., Polineni, D. & Khalil, S. 2006. An interactive web-based curriculum on evidence-based medicine: design and effectiveness. Fam Med, 38, 126-32.
Schneider, A. T., Albers, P. & Muller-Mattheis, V. 2015. E-Learning in Urology: Implementation of the Learning and Teaching Platform CASUS (R) - Do Virtual Patients Lead to Improved Learning Outcomes? A Randomized Study among Students. Urologia Internationalis, 94, 412-418.
Sekine, M., Watanabe, M., Nojiri, S., Suzuki, T., Nishizaki, Y., Tomiki, Y. & Okada, T. 2022. Effects of COVID-19 on Japanese medical students' knowledge and attitudes toward e-learning in relation to performance on achievement tests. PLoS One, 17, e0265356.
Sichani, M. M., Mobarakeh, S. R. & Omid, A. 2018. The effect of distance learning via SMS on academic achievement and satisfaction of medical students. J Educ Health Promot, 7, 29.
Sikkens, J. J., Caris, M. G., Schutte, T., Kramer, M. H. H., Tichelaar, J. & Van Agtmael, M. A. 2018. Improving antibiotic prescribing skills in medical students: The effect of e-learning after 6 months. Journal of Antimicrobial Chemotherapy, 73, 2243-2246.
Smolle, J., Prause, G. & Smolle-Jüttner, F. M. 2007. Emergency treatment of chest trauma--an e-learning simulation model for undergraduate medical students. Eur J Cardiothorac Surg, 32, 644-7.
Stevens, N. T., Holmes, K., Grainger, R. J., Connolly, R., Prior, A. R., Fitzpatrick, F., O'neill, E., Bol, , F., Pawlikowska, T. & Humphreys, H. 2019. Can e-learning improve the performance of undergraduate medical students in Clinical Microbiology examinations? BMC Med Educ, 19, 408.
Tan, A., Ross, S. P. & Duerksen, K. 2013. Death is not always a failure: outcomes from implementing an online virtual patient clinical case in palliative care for family medicine clerkship. Med Educ Online, 18, 22711.
Van Bonn, S. M., Grajek, J. S., Schneider, A., Oberhoffner, T., Mlynski, R. & Weiss, N. M. 2022. Interactive live-stream surgery contributes to surgical education in the context of contact restrictions. European Archives of Oto-Rhino-Laryngology, 279, 2865-2871.
Viteri Jusué, A., Tamargo Alonso, A., Bilbao González, A. & Palomares, T. 2020. Learning How to Order Imaging Tests and Make Subsequent Clinical Decisions: a Randomized Study of the Effectiveness of a Virtual Learning Environment for Medical Students. Med Sci Educ, 31, 469-477.
Waugh, S., Devin, J., Lam, A. K. & Gopalan, V. 2022. FE-learning and the virtual transformation of histopathology teaching during COVID-19: its impact on student learning experience and outcome. BMC Med Educ, 22, 22.
Webb, A. L. & Choi, S. 2014. Interactive radiological anatomy eLearning solution for first year medical students: Development, integration, and impact on learning. Anatomical sciences education, 7, 350-360.
Yilmaz, Y., Sarikaya, O., Senol, Y., Baykan, Z., Karaca, O., Yilmaz, N. D., Altintas, L., Onan, A. & Sayek, I. 2021. RE-AIMing COVID-19 online learning for medical students: a massive open online course evaluation. Bmc Medical Education, 21.
Zare-Bidaki, M., Ehteshampour, A., Reisaliakbarighomi, M., Mazinani, R., Khodaie Ardakani, M. R., Mirabzadeh, A., Alikhani, R., Noroozi, M., Momeni, F., Samani, A. D., Mehrabi Tavana, M. M., Esmaeili, A. & Mousavi, S. B. 2022. Evaluating the Effects of Experiencing Virtual Reality Simulation of Psychosis on Mental Illness Stigma, Empathy, and Knowledge in Medical Students. Front Psychiatry, 13, 880331.