2023

EJEL Volume 21, Issue 5

Editors

Heinrich Söbke and Marija Cubric

Published by Academic Publishing International Limited
Curtis Farm, Kidmore End, Nr Reading, RG4 9AY, United Kingdom
karen.harris@academic-publishing.org

eISSN: 1479-4403

EJEL Volume 21, Issue 5

Contents

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).

 ©The Authors

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.

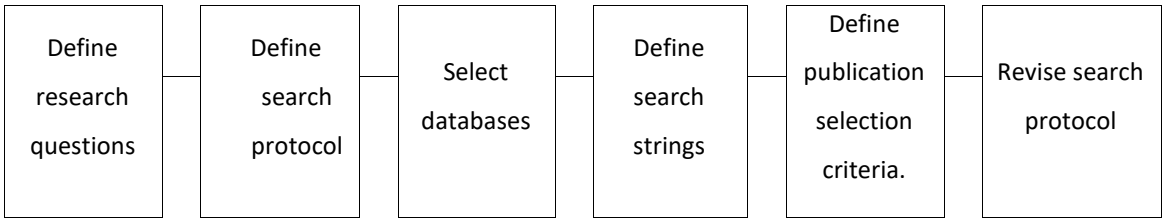

Figure 1: The Internal Process of the "Plan Review" Phase

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

 ©The Authors

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.

 ©The Authors

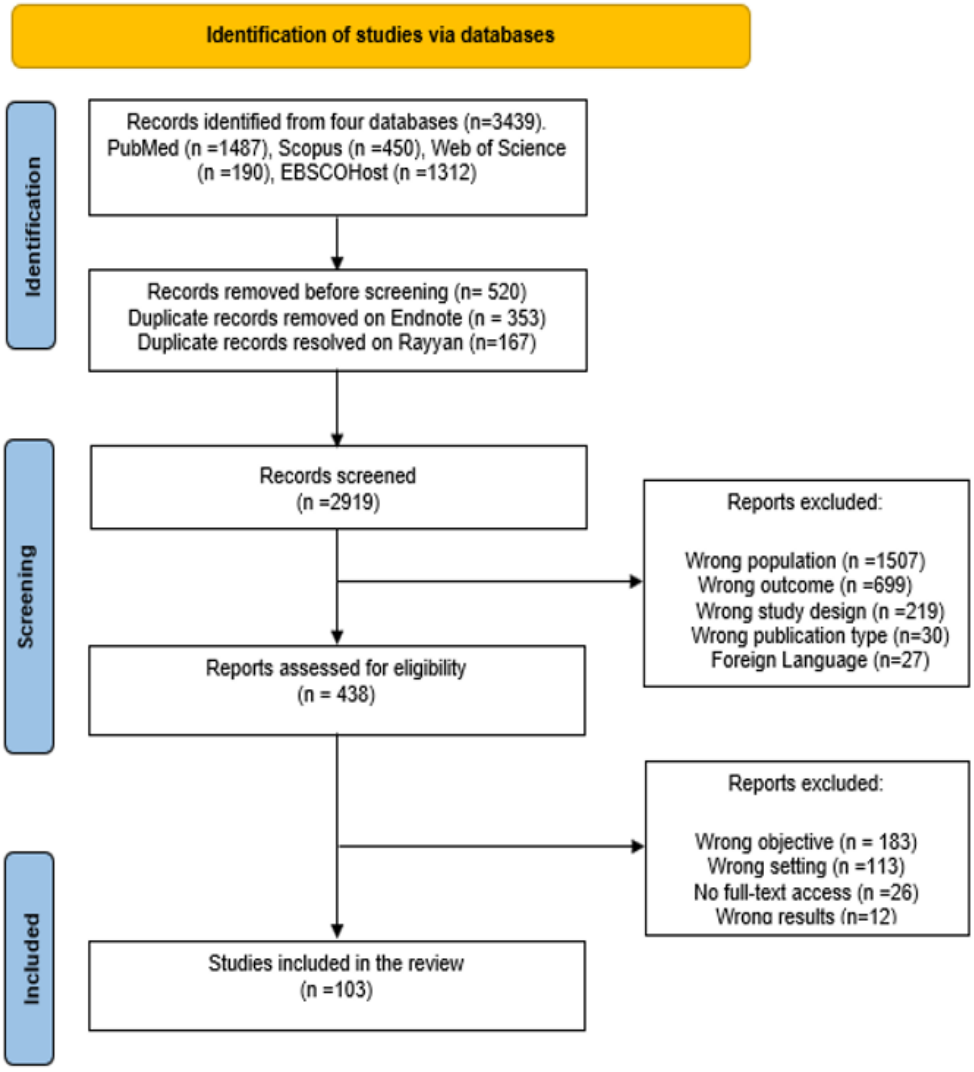

Figure 2: PRISMA Diagram

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.

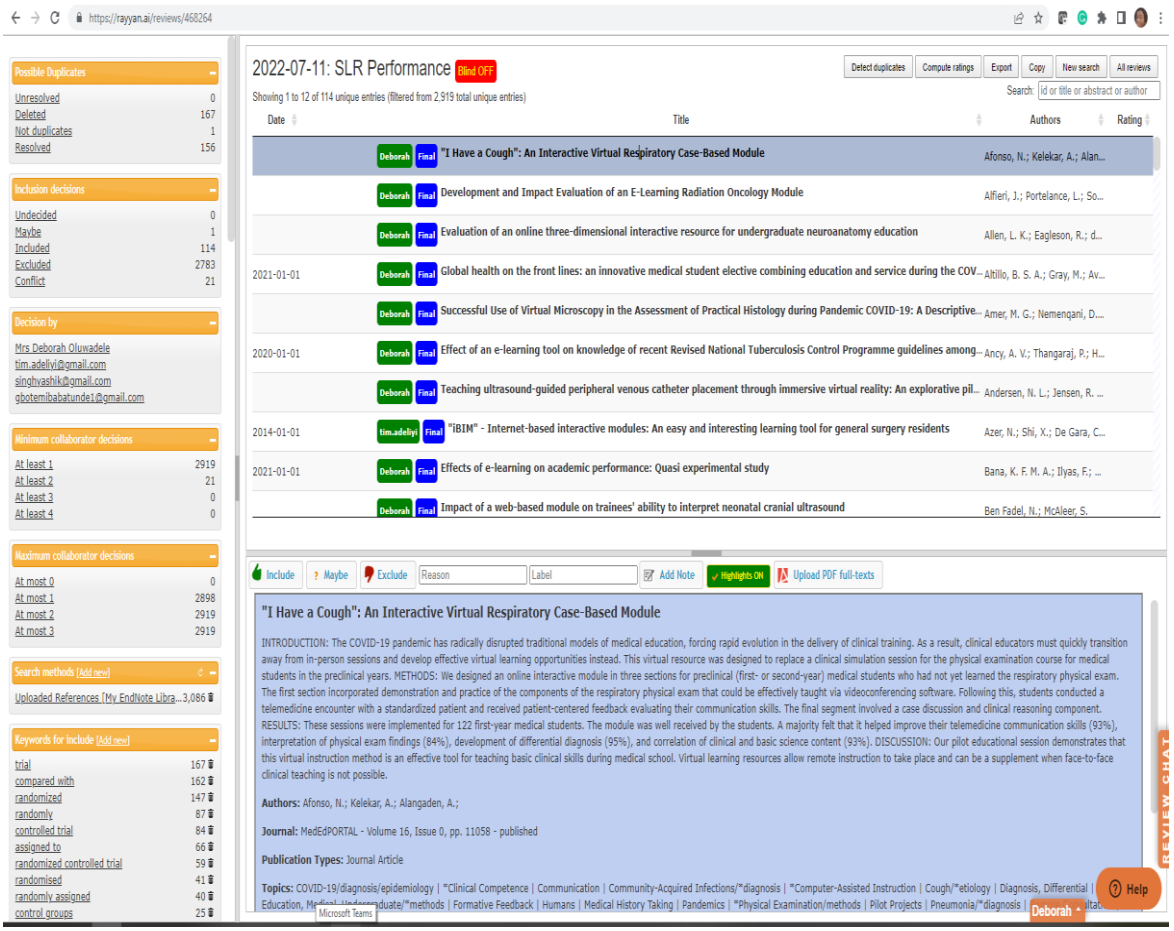

Figure 3: Quality assurance process on Rayyan

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.

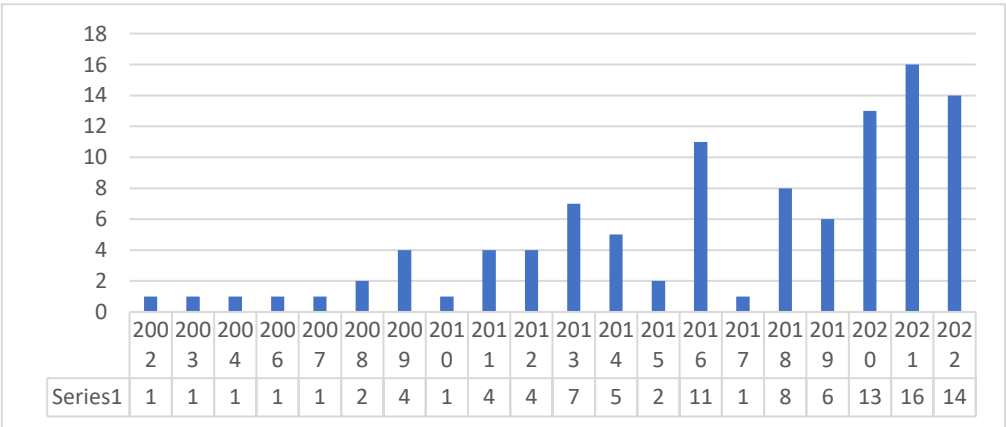

Figure 4: Historical distribution of reviewed literature

 ©The Authors

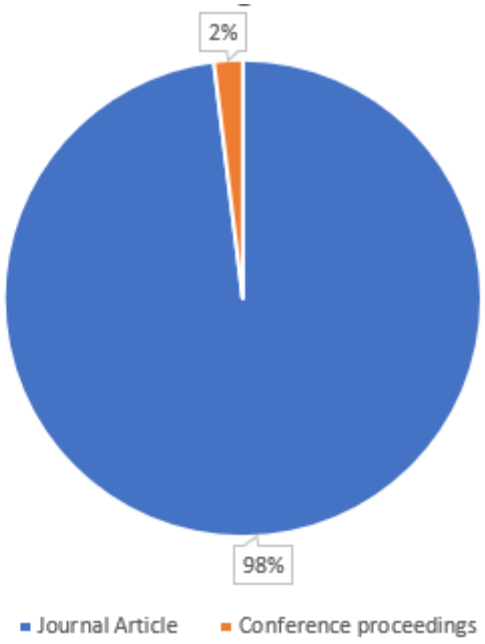

Figure 5: Publication types of the reviewed literature

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.

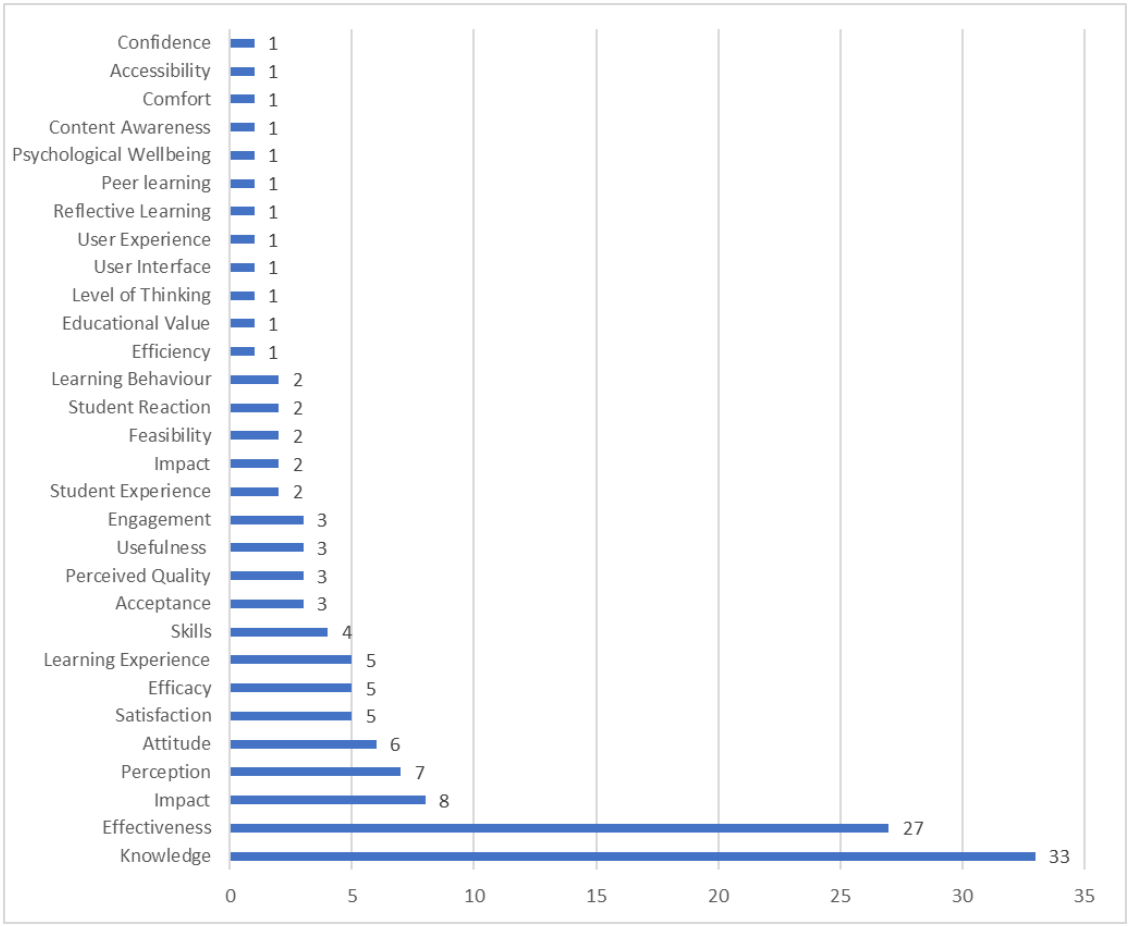

Figure 6: Constructs used to operationalize performance

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).

 ©The Authors

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.

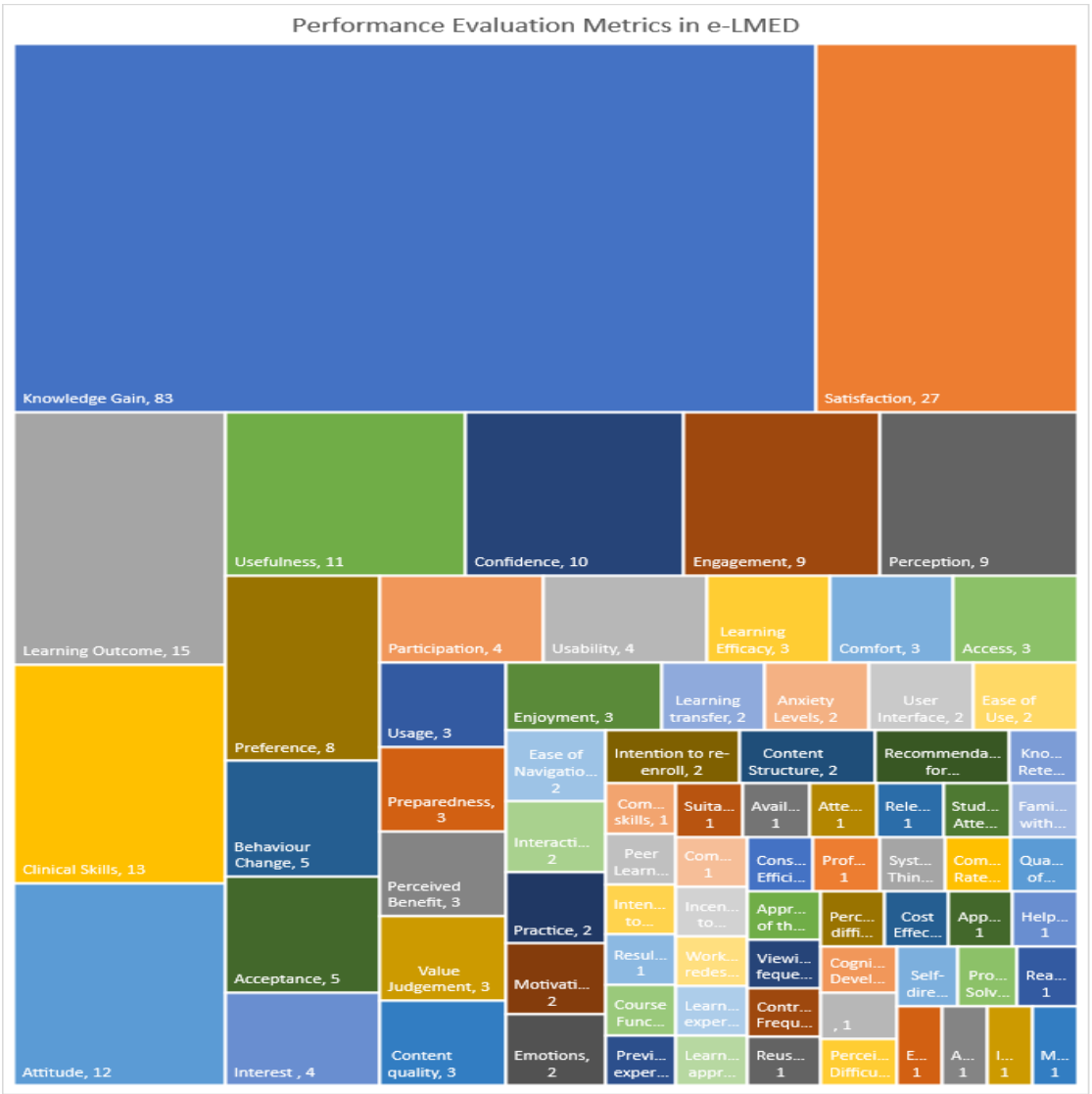

Figure 7: Metrics and factors used to measure performance

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.

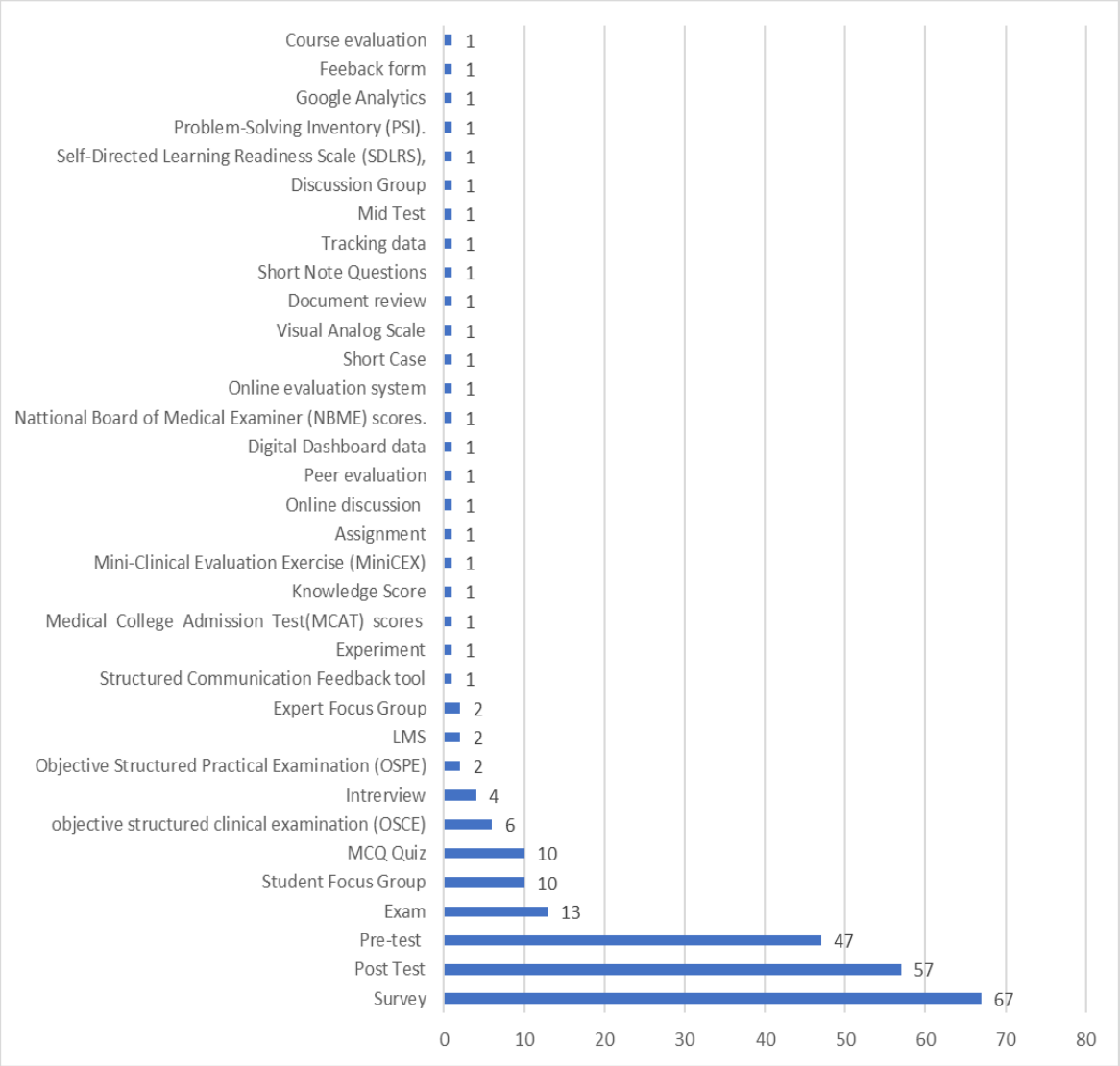

Figure 8: Research Methods used to evaluate performance in e-LMED

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

 ©The Authors

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

 ©The Authors

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.

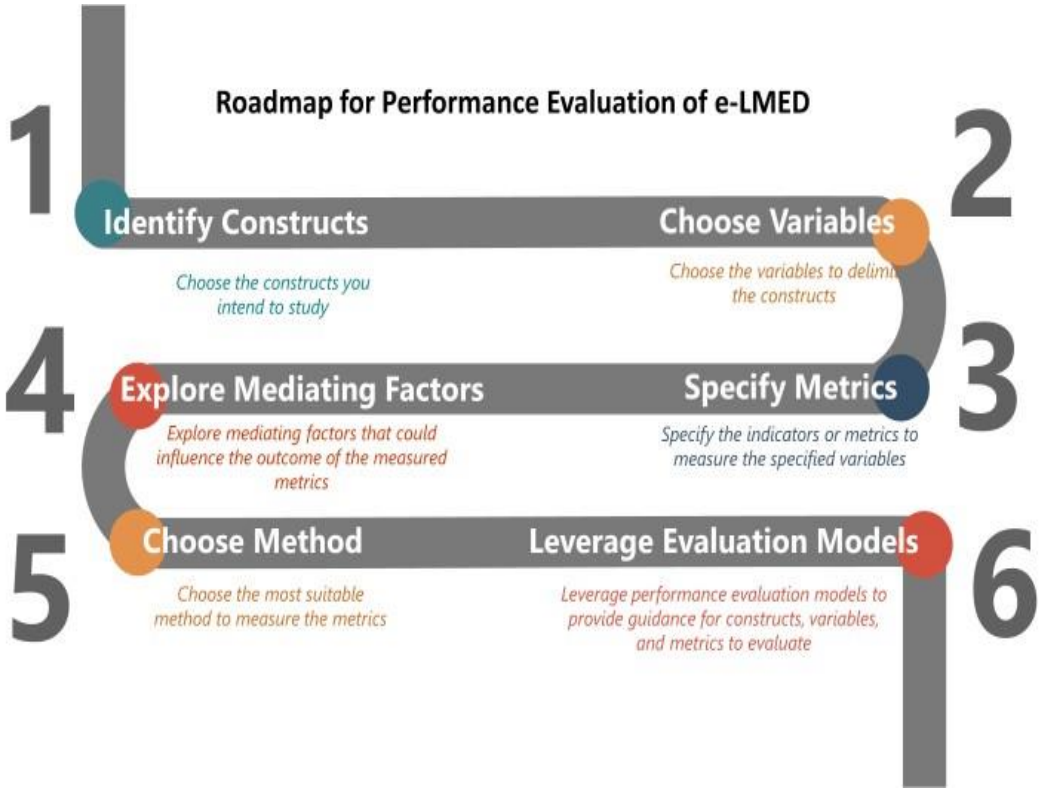

Figure 9: Proposed roadmap for performance evaluation of e-LMED

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.

 ©The Authors

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.

 ©The Authors

A Systematic Review of the Challenges of e-Learning Implementation in Sub-Saharan African Countries: 2016-2022

Thomas Mc Donald van der Merwe, Mmafani Serote and Molefe Maloma
School of Computing, University of South Africa, South Africa

vdmertm@unisa.ac.za
serotm@unisa.ac.za
malommc@unisa.ac.za

Abstract: Even a cursory review of the research literature over the past two decades will reveal many e-learning implementation challenges facing sub-Saharan Africa Higher Education Institutions. The last available systematic review of these challenges was conducted in 2017. To address this gap and by following the PRISMA-ScR guidelines, our systematic review aimed to find answers to the following two research questions: (a) What e-learning challenges were identified in research conducted in Higher Education Institutions in sub-Saharan Africa from 2016, and (b) Which e-learning challenges identified before 2016 remain and which are new? Using search terms and synonyms associated with e-learning challenges, three reviewers performed a Boolean search in Google Scholar, Web of Science, and Scopus. After duplicates (113) were removed, a total of 353 potential articles were imported into a collaboration platform and filtered using the following inclusion criteria: (1) Higher Education Institutions in sub-Saharan Africa (2) English articles from 2016 to 2022 (3) user data on real-world e-learning implementations are available (4) use of a computer/mobile device excluding m-learning approaches and (5) blended or fully online courses. After a selection process where article sets were reviewed, rotated, and conflicts resolved, 25 articles that met the inclusion criteria formed the primary evidence base. After a data collection process and using selected Grounded Theory techniques, a total of 48 challenges, synthesised in 25 challenge categories, were identified across six themes: university-related (36 citations), instructor-related (20), access-related (14), LMS-related (10), student-related (9) and computer literacy-related (9). The main challenges categories were ICT infrastructure (14 citations), technical support (10) and student bandwidth (8), while the secondary challenges were mainly related to the instructor: commitment and capacity (7), capacity (6), LMS training (6) and instructional design and support (5). Suggestions are offered to address key challenges related to ICT infrastructure, technical support, and student bandwidth. Secondary challenges, which are mainly related to the instructor, are proposed to be resolved by a skilled and energetic instructional design and support team. Although many challenges identified before 2016 were reinforced in this investigation, no new challenges were identified. Overall, the number of citations for specific challenges shows that most challenges are context-specific, that is, limited to single HEIs. However, when combined in challenge categories and reflected as themes, the review reveals many challenges to remain persistent. To ensure that an institution, its instructors, and its students are geared and supported stepwise, a holistic top-down approach that considers all challenges is required. To achieve this, a recommendation for further research is the development of an updated SSA-specific checklist that can be used to rate e-learning readiness in terms of importance and context. The research aims to provide active researchers in the field with an update on the status of research on e-learning challenges in SSA, as well as to make decision-makers aware of potential challenges when implementing e-learning in the SSA context. This may result in better institutional strategies to support e-learning practitioners, as well as to ensure the effective development and implementation of e-learning management systems.

Keywords: Systematic literature review, e-Learning challenges, Sub-Saharan Africa, 2016 – 2022

1. Introduction

E-learning, long touted as the next educational revolution, has been implemented with great success in some countries, but less so in others. Sub-Saharan (SSA), despite being considered the most dynamic e-learning market in the world (Adkins, 2016), also dominates the least developed countries (LDC) category as defined in 1971 by the United Nations (UN, 2011, 2022) with 33 of the 45 recognised LDCs situated in this region. Despite progress made, some key challenges listed by UNESCO (2021) and UNICEF (2021) that continue to exist and hinder the effective implementation of e-learning are a lack of policies to ensure connectivity, including Internet access and Wi-Fi-capable devices for student and teacher use, training, and retraining of teachers in further professional development, including using modern digital tools, and general nonavailability of distance learning. Even a cursory review of the research literature over the past two decades will reveal a multitude of e-learning implementation challenges facing SSA Higher Education Institutions (HEIs).

In his seminal paper, Selim (2007) specified a list of Critical Success Factors (CSFs) for the acceptance of e-learning in a single HEI. He proposed four main categories: 1) instructor, (2) student, (3) information technology, and (4) university support. Since then, a plethora of individual research studies in single countries or individual HEIs has revisited, added, and expanded on his list, in the process adding many factors outside of e-learning

Reference this paper: van der Merwe, T.M., Serote, M. and Maloma, M., 2023. A Systematic Review of the Challenges of e-Learning Implementation in Sub-Saharan African Countries: 2016-2022, *The Electronic Journal of e-Learning, 21(5)*, pp 413-429, available online at www.ejel.org

acceptance. Reviews of such research have been published at regular intervals (e.g., Cheawjindakarn, Suwannatthachote and Theeraroungchaisri, 2013); Elkaseh, Wong and Fung, 2015; Basak, Wotto and Bélanger, 2016; Al-Fraihat, Joy and Sinclair, 2017; Asalla, Putri and Pradipto, 2017). Whereas Selim employed the term CSF, defined as "those things that must be done if a company is to be successful" (Freund, 1988, p.20), it is often misused by e-learning researchers when they mean "challenges". Challenges are assigned to CSFs and form a basis for taking actions to control and master the various CSFs (Françoise, Bourgault and Pellerin, 2009). Therefore, the term "challenges" is a more appropriate term to describe the issues that impact the implementation of e-learning.

An in-depth search of the literature has revealed that the most recent review (scoping, rapid, narrative, systematic, or meta-analysis) on the implementation of e-learning or the adoption challenges of an e-learning management system (LMS) identified in SSA was conducted by Bervell and Umar (2017), who covered research published in the period 2007 - 2017 and is therefore outdated. This raises the broad question: Given the exponential growth of ICT and the rush by HEIs to gain a competitive advantage afforded by e-learning offerings, are the challenges identified by this review still present, and/or have any new challenges been identified?

By adopting a systematic approach, the purpose of our research is to identify, evaluate, and summarise the findings of all relevant individual studies that reported challenges of e-learning in SSA HEIs in the period 2016-2022. The reason for including the years 2016 and 2017 when Bervell and Umar (2017) also covered these years lies in a specific review exclusion criterion that disqualifies the relevance of their literature from these years.

Our specific research questions are as follows:

RQ1: What e-learning challenges were identified in research conducted in Higher Education Institutions in sub-Saharan Africa from 2016?

RQ2: Which e-learning challenges identified before 2016 remain and which are new?

Our research aims to provide active researchers in the field with an update on the status of research on e-learning challenges in SSA, as well as to make decision-makers aware of potential challenges when implementing e-learning in the SSA context. This may result in better institutional strategies to support e-learning practitioners and to ensure the effective development and implementation of e-learning management systems.

The methodology, results, and conclusions are presented next.

2. Methodology

A systematic review uses explicit and systematic methods to collate and synthesise the findings of studies that address the formulated question/s (Higgins, et al., 2019). Two of the most common methodologies used to ensure accurate, complete and transparent reviews are PICO and PRISMA. Both are typically used in clinical research reporting on health interventions, but many of the checklist items can also be used outside of this field. The reporting of our systematic review was guided by the Preferred Reporting Items for Systematic Reviews (PRISMA-ScR) Statement, exclusive of the Meta-analysis Extension. As dictated by the statement, the relevant preferred reporting items for the methodology to be employed in this review are eligibility criteria, information sources, search strategy, selection process, data collection process, data items, risk of bias assessment, and synthesis methods.

2.1 Eligibility Criteria

Inclusion and exclusion criteria were set based on the scope variables of interest, i.e., peer-reviewed research reports not included in previous systematic literature reviews; e-learning challenges identified in existing e-learning implementations and approaches, and that includes user data (feedback). Although language restrictions can result in a risk of bias and reduce generalisability, questions remain about the accuracy of translation programmes, with the result that limiting reviews to English only is common in systematic literature reviews (Jackson and Kuriyama, 2019).

The specific inclusion criteria were:

- Publications in the period 2016-2022.
- Publications in English.
- Publications on e-learning challenges in a sub-Saharan HEI.

- Publications in which challenges were identified from user data, that is, a qualitative, quantitative, or mixed methodology, and reports on the real-world use of an e-learning system/approach were reported.
- The use of a computer or a mobile device, but excluding a dedicated m-learning approach.
- Any blended or fully online courses.

The exclusion criteria were as follows:

- Publications that focused on beliefs, perceptions, and usage intentions (e.g., UTAUT and TAM).
- Publications that evaluated general perceptions of ICT use.
- Publications not published in peer-reviewed journals or conference proceedings, that is, "grey literature".
- Publications offering an e-learning framework and/or strategies only.
- Publications that focus on the architecture of an LMS.
- COVID-19-forced e-learning interventions. Whereas the pandemic forced a dramatic shift towards e-learning initiatives, that "emergency mode" implementations may result in unintended consequences (Kulikowski, Przytula and Sulkowski, 2021) are excluded.
- Testing a novel/newly designed LMS environment and reporting on its success.
- Where the focus is on a single e-learning tool (e.g., a podcast, a webinar, a discussion forum, WhatsApp, etc.).

2.2 Information Sources

To select articles for review, the three current authors, who acted as reviewers, independently searched three databases, namely Google Scholar, Web of Science, and Scopus, during the period August 2022 - October 2022. The Google Scholar index is the most comprehensive of the academic literature (Gusenbauer, 2019). Although Web of Science and Scopus only cover a small option of Google Scholar and tend to cover relatively more from the Engineering, Computer Science, Energy, Chemical Engineering, Chemistry, Veterinary, and Neuroscience fields (Gusenbauer, 2022), Tibaná-Herrera, Fernández-Bajón and De Moya-Anegón (2018) reported Scopus to index 219 journal and congress articles that included e-learning themes between 2012 and 2014. Similarly, for Web of Science, Fatima and Abu (2019) reported that one-third of the total cited e-learning references between 1989 and 2018 were published in 11 journals, the rest scattered across more than 1600 journals. It should be noted that, although Scopus returned most hits from our searches, very few met our inclusion criteria.

2.3 Search Strategy

In addition to the four primary search terms (e-learning, electronic learning, Africa, and challenges), synonyms such as online learning, distance learning, open distance learning, virtual learning, online education, and web-based learning were combined with terms often associated with challenges (CSFs, adoption, usefulness, barriers, vital, pressing, crucial, key, and lessons learnt) using Boolean operators. These additional challenge terms were identified in titles and abstracts from articles sourced from previous searches and added as additional Boolean operators. As an example, an initial search string such as ("e-learning" OR "electronic learning" OR "online learning" OR "virtual learning") AND ("Africa") AND ("critical success factors") was updated to (… AND ("critical success factors") OR "CSFs" OR "challenges" OR "barriers" OR "crucial"). Harzing's Publish or Perish tool, Version 8 (Harzing, 2007) was used to retrieve raw academic citations from the title, abstract, or body text using the title field or keyword field.

Figure 1 offers a methodology processes flow chart that shows the information sources, selection process, data item collection process, and synthesis methods, to be discussed in the following subsections.

 ©The Authors

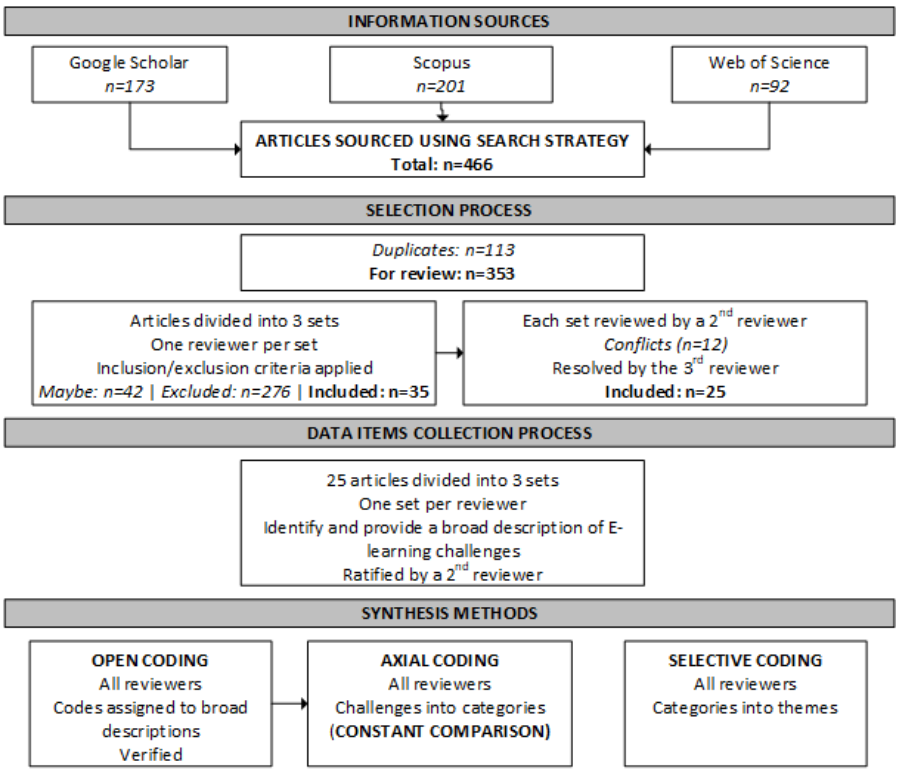

Figure 1: Methodology processes flow chart

2.4 Selection Process

A total of 466 potential citations were sourced from the three databases, exported as a RIS file, and imported into Rayyan (http://rayyan.ai), a web-based system used for collaborative systematic reviews of literature that expedites the initial selection of abstracts and titles.

Once duplicates were removed, a remaining total of 353 abstracts were divided into three sets, each set subjected to the inclusion/exclusion criteria by a single reviewer, who read the abstracts (sometimes scanning the body text if required) and labelled them as 'undecided', 'maybe', 'included' or 'excluded'. The second and third reviewers were trained by the main author, who has extensive experience in qualitative research. At the end of this preliminary process, 42 articles were labelled as 'maybe', 35 as 'included', and 276 as 'excluded'. To reduce the risk of bias and verify the tags, a second reviewer reviewed each set of abstracts and the labels assigned by the first reviewer. If the second reviewer agreed with the original label, the article was included for data collection. If there was disagreement, it was marked as a 'conflict' (12 articles), with the third reviewer required to make a final decision. In all cases, if the abstract was not clear, the full text was consulted. There was no need to contact the authors of any article for more information. After this process, the full text of the 25 articles was downloaded and included in a data collection process.

2.5 Data Item Collection Process

Each reviewer was assigned a set of articles to identify specific e-learning challenges reported in the full article text, which was recorded as a broad description in an MS Word table together with the country where the research was performed, the authors, subjects, and the methodological approach. The latter two elements were included to ensure that the specific inclusion criteria were met. To reduce the risk of bias, the sets were once again rotated with each article and its broad descriptions ratified or, if necessary, revised by a second reviewer.

2.6 Synthesis Methods

In this method, the selected Grounded Theory (GT) techniques of open and axial coding, constant comparison, and selective coding (Strauss and Corbin, 2008) were employed. The document containing the ratified table was imported into Atlas.ti (https://atlasti.com/) as a hermeneutic unit. The three reviewers then met for three sessions in which they employed the open-coding technique to assign concise codes (labels) to the descriptions of the broad challenges. This was followed by a verification session where the codes were reviewed and, if necessary, refined. In the final process, the reviewers used the Network Manager available in Atlas-ti GT and the

technique of constant comparison to sort and organise related codes into categories (axial coding). Lastly, using an adapted GT selective coding approach proposed by Van der Merwe and De Villiers (2011), categories were ordered into classes (themes) according to their attributes.

3. Results

Table 1 lists, in no specific order, the countries where the research took place, the author/s, and a reference number to be used when citing the identified challenge categories (see Table 2).

Table 1: List of studies included in the final review

Country	Authors' references	Reference no.
Tanzania	Mutisya and Makokha (2016)	1
South Africa	Olasina (2019)	2
South Africa	Patel, Kadyamatimba and Madzvamuse (2018)	3
South Africa	Qwabe and Khumalo (2020)	4
Tanzania	Raphael and Mtebe (2016)	5
Nigeria	Yakubu and Dasuki (2018)	6
Malawi	Zozie and Chawinga (2018)	7
Ghana	Ansong, Lovia Boateng and Boateng (2016)	8
Malawi, South Africa, Tanzania, Uganda	Atkins, et al. (2016)	9
South Africa	Azeez and Van Der Vyver (2018)	10
South Africa	Basitere and Ivala (2017)	11
South Africa	Pete, Coopasami and Knight (2017)	12
Sudan	Idris and Osman (2016)	13
Nigeria	Eze, Chinedu-Eze and Bello (2018)	14
South Africa	Faloye and Ajayi (2021)	15
Uganda	Gupta, et al. (2017)	16
Tanzania	Mahenge and Sanga (2016)	17
South Africa	Gani and van den Berg (2019)	18
South Africa	Mafunda and Swart (2020)	19
Namibia	Mässing (2017)	20
South Africa	Msomi and Bansilal (2018)	21
Kenya	Khavugwi (2017)	22
South Africa	Maphalala and Adigun (2021)	23
Tanzania	Mgeni, Ismail, Yunus, and Haji (2019)	24
Tanzania	Mtebe (2020)	25

All 25 included studies were published between 2016 and 2021. Most of the research originated in South Africa (12), followed by Tanzania (6) and Nigeria, Uganda, and Malawi (2 each).

Table 2 lists the themes and *challenge categories* that resulted in the themes, the author reference number/s (from Table 1), the number of citations per challenge, the total citations per category, as well as a total category citation threshold, where challenges with a single citation are considered non-persistent / contextual.

 ©The Authors

Table 2: Results of the systematic review

Themes	*Challenge categories*	Authors' references	Number of citations per category	Total Category citations	Category citation threshold (3) total
Instructor-related	*I: Instructional design and support*	5, 7, 11, 16, 25	5	20	18
	I: Capacity	1, 5, 7, 18, 20, 22	6		
	I: Commitment	7, 14, 16, 18, 20, 22, 25	7		
	I: Pedagogy	25	1		
	I: Instructor role	2	1		
Student-related	*S: Readiness*	3, 7, 12	3	9	3
	S: Capacity-	1, 2	2		
	S: Experience	2, 21	2		
	S: Finance	10	1		
	S: Familiarity	24	1		
LMS-related	*I: LMS Training*	5, 13, 14, 18, 23, 25	6	10	9
	S: User-friendly LMS	6, 10, 25	3		
	I: User-friendly LMS	6	1		
Access-related	*S: Bandwidth*	1, 4, 7, 16, 17, 18, 21, 24	8	14	14
	S: ICT ownership	11, 12, 16	3		
	I: Bandwidth	1, 5, 11	3		
Computer literacy-related	*I: Computer skills*	1, 7, 14, 18	4	9	9
	S: Computer skills	1, 3, 7, 10, 15	5		
University-related	*U: Policies*	12, 16, 17	3	36	33
	U: Environment	8, 12, 16	3		
	U: Financial constraints	1, 20	2		
	U: ICT infrastructure	1, 3, 5, 8, 10, 13, 14, 16, 17, 19, 20, 22, 23, 24	14		
	C: Copyright	1	1		
	I: Incentives	1, 13, 22	3		
	U: Technical support	3, 5, 6, 9, 12, 13, 15, 16, 22, 24	10		
Totals	25	98	98	98	86

Key: I= Instructor, S = Student, U = University, C = Course

A total of 25 challenge categories, to be deconstructed in the following section, were grouped into six themes: Instructor-related (5 categories), Student-related (5 categories), LMS-related (3 categories), Access-related (3 categories), Computer literacy-related (2 categories) and University-related (7 categories). Note that each challenge category consists of many individual and specific challenges, which will be unpacked in more detail in Section 3.

Although a specific challenge can potentially be placed in multiple challenge categories and/or themes, we followed a best-fit approach. For example, *LMS Training* requirements are arguably more system-related than what it is *Instructor-related,* i.e., when more functionality is added to an LMS, or when tools are complex and/or not user-friendly, or when new tools are added, instructors may require more training. In other words, the need for training does not necessarily originate with the instructor. Similarly, while *Instructional design and support*

can potentially fit in the University-related (support) or the LMS-related theme, the exact online teaching and learning approach is ultimately dictated by the instructor who then requires further support from an instructional design team or through workshops. For this reason, it fits best in the Instructor-related theme. In making these decisions, the authors relied on their experience in an institution that has a long history of online teaching and learning and is considered one of 11 mega-distance teaching universities in the world.

A total of 98 citations were identified. Per category, University-related challenges returned 36 citations, Instructor-related 20, Access-related 14, LMS-related 10, Student-related 9, and Computer literacy-related 9. If a minimum threshold of 3 citations is set, the total number of citations is reduced to 86. However, such a rule will only have one impact on the ranking order, with the Student-related theme moving down to last on the list (University-related 33, Instructor-related 18, Access-related 14, LMS-related 9, Computer literacy-related 9, and Student-related 3). After deliberation, all citations were included to cover potential within-theme and category challenges associations. For example, S*tudent readiness* (3 citations) can also be related to *Student experience* (2 citations). Being immersed in the total data, a best-fit approach was followed.

In ranking the number of citations per category (Figure 2), *ICT infrastructure* (14) and *Technical support* (10) are ranked highest, followed by S*tudent bandwidth* (8), *Instructor commitment* (7), *Instructor capacity* and *LMS training* (6 each). The remainder of the challenges were cited 5 times or less.

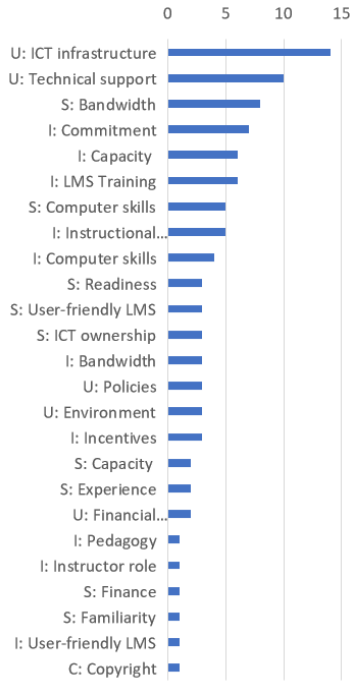

Figure 2: Number of citations per challenge

In the following sections, we unpack the specific challenges per category. Where there are potential within-category challenge associations, they are combined with other challenge categories.

3.1 Instructor-Related

3.1.1 Instructional design and support

Instructional design (Wagner, 2011 p.34) refers to "the systematic development of instructional specifications using learning and instructional theory to ensure the quality of instruction". This type of design is typically provided by a unit that supports instructors on technical and pedagogical issues, although instructors are sometimes required to design on their own.

Raphael and Mtebe (2016) reported a problem with a shortage of adequately qualified staff to support instructors on technical and pedagogical issues. The quality of instruction is also affected by the online tools used. For example, Zozie and Chawinga (2018) reported that, despite using a variety of media, training is required in other areas, such as video and audio. When using a specific tool such as video streaming, incompatibilities arise when learning hubs bypass guidelines on the hardware and software required (Gupta, et

 ©The Authors

al., 2017). Using a time-consuming platform offered by an outside company has resulted in student frustration and an unwillingness to use it (Basitere and Ivala, 2017).

Where instructors are required to design on their own, Mtebe (2020) reported that they did not have enough skills and competencies to develop quality instruction, while some instructors also display a tendency to use an LMS as a passive repository (Gupta, et al., 2017).

3.1.2 Capacity

Mutisya and Makokha (2016) found that instructors rank high workloads (the result of many students registering for online courses) as the main reason why they are not embracing e-learning. Raphael and Mtebe (2016) reported the same but added that instructors do not have the skills to manage the increased workload. Gani and van den Berg (2019) also found the workload to be prohibitive, which, although not pertinently stated, is possibly related to another finding that very few students use the LMS. That is, it offers a low return on time investment. An interesting constraint is mentioned by Mässing (2017), who noted that younger managers who govern ICT implementation in learning are overworked because the older generation of instructors does not "take orders" from them. Both Zozie and Chawinga (2018) and Khavugwi (2017) highlighted the increased workload that arises when instructors must redesign/convert the material to electronic format, while the latter noted that the complexity of online teaching can be demanding and time-consuming.

3.1.3 Commitment, pedagogy, and role

A common challenge is the underuse of the features of an LMS. Zozie and Chawinga (2018) reported that despite efforts to communicate its existence, some instructors have a negative attitude toward using technology in teaching. Both Gani and van den Berg (2018) and Eze, Chinedu-Eze and Bello (2018) found experienced instructors to see no reason to change from the analogue to the digital age, the former noting that instructors feel comfortable with the traditional way of teaching. Despite training, Mässing, 2017) found that instructor usage remained near zero, a possible solution is to make it mandatory. Mtebe (2020) found that where instructors use few LMS features, they display little pedagogical skills in using them. As noted by Khavugwi (2017), online teaching requires commitment, while faculty have a direct influence on student attitudes (Olasina, 2019). Gupta, et al. (2017) detailed requirements for suitably qualified and experienced staff but also referred to the need for motivated staff who are naturally interested in e-learning initiatives. Yakubu and Dasuki (2018) reported that information quality (effective teaching by trained instructors that will ensure relevance, comprehensibility, and accuracy) has the strongest influence on student attitudes toward the use of an LMS.

3.2 Student-Related

3.2.1 Readiness, experience, familiarity, capacity, and finance

Both Olasina (2019) and Zozie and Chawinga (2018) identified previous e-learning experience by students as a use factor – the former adding Internet experience and the latter an over-reliance on instructors as the sole provider of teaching and learning resources. Pete, Coopasami and Knight (2017) found that while students were psychologically ready, they lacked technological and equipment readiness, most likely associated with the finding that some students have not had exposure to technology at the school level (Msomi, 2018). To these factors, Mgeni, et al. (2019) added unfamiliarity to the features of an LMS, while Patel, Kadyamatimba and Madzvamuse (2018) found that students lack awareness of e-learning, or are not confident in instructors' support, skills, and usage. Student capacity for e-learning approaches was also identified as a use factor. Although ranked last by students as a problem, Mutisya and Makokha (2016) reported that students, for a variety of reasons, do not have enough time to interact online in a blended learning environment. Olasina (2019) expanded on these factors by identifying stress, workload, satisfaction, fatigue, and time management. Finally, Azeez and Van Der Vyver (2018) stated that e-learning must be affordable for students before they commit.

3.3 LMS

3.3.1 LMS training (instructor)

The need for instructors to be trained on how to use an LMS has been identified in several studies. One of the main findings of Raphael and Mtebe (2016) was the lack of instructors' skills to use the LMS features. Idris and Osman (2016) reported that two-thirds of staff have not received training in e-learning applications, with Eze, Chinedu-Eze and Bello (2018) reporting a 50% figure. Gani and van den Berg (2019) proposed that instructors must undergo training to engage creatively, intentionally, and purposefully with an LMS. Both Mtebe (2020) and

Maphalala and Adigun (2021) stated that skills and confidence in the use of an LMS and its features cannot be achieved with a few days of training, while any LMS update should be followed by further training.

3.3.2 S: User-friendly LMS (instructor and student)

The user-friendliness of an LMS dictates its subsequent use by instructors and students, which in turn is influenced by several factors such as availability, reliability, and design functionalities (Yakubu and Dasuki,2018), simplicity and ease of configuration (Azeez and Van Der Vyver, 2018), and the availability of documentation and help features that impact the user experience and thus student learning (Mtebe, 2020).

3.4 Access

3.4.1 Bandwidth (instructor and student)

Students ranked insufficient Internet connectivity as their key challenge in using an LMS (Mutisya and Makokha, 2016), which affects access time and makes it difficult to complete work (Qwabe and Khumalo, 2020). Interacting with rich media, such as videos, becomes problematic (Zozie and Chawinga, 2018; Gupta, et al., 2017), as do synchronous communications (Gupta et al., 2017). Students who live off-campus are also more affected than residential students, with the former recognised as having limited or no bandwidth (Gupta, et al., 2017; Gani and van den Berg, 2019). Basitere and Ivala (2017) found that poor student performance directly relates to student bandwidth shortages. Expensive bandwidth, of course, affects access to educational opportunities (Mahenge and Sanga, 2016). However, bandwidth limitations on campus also make it difficult for both students (Msomi, 2018; Mgeni et al., 2019) and instructors (Mutisya and Makokha, 2016; Raphael and Mtebe, 2016) to use an LMS effectively.

3.4.2 ICT ownership (student)

Without adequate hardware, e-learning will be impractical. For example, Gupta et al., (2017) and Lucas (2017) reported that most students have trouble accessing the Internet off-campus because they do not have access to computers. Basitere and Ivala (2017) noted that, unlike residential students, off-campus students do not have access to computer resources, resulting in a digital divide that leaves these students frustrated and dependent on the former to complete tests. Not having access to a computer also means that students are not equipped for e-learning (Pete, Coopasami and Knight, 2017).

3.5 Computer Literacy

3.5.1 Computer skills (instructor and students

Computer illiteracy affects both instructors and students. For instructors, Mutisya and Makokha (2016) identified limited ICT skills as the fourth most important factor that prevents them from adopting e-learning. Azeez and Van Der Vyver (2018) and Eze, Chinedu-Eze and Bello (2018) also highlighted basic computer skills as problematic, the latter identifying specific skills such as powering their laptops, composing and sending e-mails, accessing emails, attaching files, and other peripheral issues. Although a lack of computer skills makes it difficult for instructors to use an LMS (Gani and van den Berg, 2019), even those with average skills found the use of more advanced tools such as text-to-speech software, video-editing software, and basic LMS functions, such as uploading material, difficult (Zozie and Chawinga, 2018).

Students ranked computer skills as their second most important challenge (Mutisya and Makokha, 2016), with Patel, Kadyamatimba and Madzvamuse (2018) identifying basic challenges such as an inability to browse the Internet and download useful learning materials, the latter partly due to them not having access to technology. Average computer skills are also problematic when students are required to work with less common operating systems such as Linux or Ubuntu (Zozie and Chawinga, 2018). As Azeez and Van Der Vyver (2018) pointed out, a third of the students highlighted computer illiteracy as a factor that affects their decision to embrace e-learning. However, only 7% and 10% of the students highlighted the lack of basic computer skills or unfamiliarity with technology as a factor, with these students typically coming from disadvantaged schools, which not only reinforces the digital divide but also results in computer anxiety (Faloye and Ajayi, 2021).

3.5.2 University-related

Despite the existence of an e-learning policy, Pete, Coopasami and Knight (2017) concluded that all stakeholders need to analyse how e-learning can benefit a programme and the institution. Gupta, et al. (2017) pointed to policies that support traditional rather than e-learning models. Mahenge and Sanga (2016) attributed the absence of a local policy to the need for a national ICT policy that will guide adoption at the HEI level.

 ©The Authors

3.5.3 Environment and financial constraints

Ansong, Lovia Boateng and Boateng (2016) concluded that there is an inverse relationship between organisational compatibility and e-learning adoption. Pete, Coopasami and Knight (2017) concluded that various factors, most of which are referenced elsewhere in this paper, combine to create an environment that does not make it easy for students to embrace e-learning. There is also a need for a university to actively move from a non-conventional approach to a virtual approach (Gupta, et al., 2017). However, as stated by Mässing (2017), to receive continuous funding, learning professionals are forced to make ICT solutions available and do so at the expense of training programmes, which is not ideal. Mutisya and Makokha (2016) concluded that the reason for the myriad of challenges is found in the inadequate availability of university financial resources.

3.5.4 ICT Infrastructure

ICT infrastructure, including the lack of hardware (typically in computer laboratories), software, and poor internet and network connections, is highlighted by several authors. Ansong, Lovia Boateng and Boateng (2016) found that the available ICT infrastructure has a significant impact on the adoption rate of e-learning, while an inadequate number of devices leads to challenges using the LMS (Mgeni et al., 2019). The instructors in Mutisya and Makokha's (2016) study ranked the lack of computers on campus as the sixth most important factor that prevents them from adopting e-learning. Other authors who also highlight a shortage of computer hardware are Patel, Kadyamatimba and Madzvamuse (2016), Idris and Osman (2016), Mahenge and Sanga (2016), Mässing (2017), Khavugwi (2017), and Maphalala and Adigun (2021). Mahenge and Sanga (2016), Idris and Osman (2016) and Mässing (2017) identified a shortage of applicable software, while Raphael and Mtebe (2016) and Mafunda and Swart (2020) noted the significant role limited or no ICT facilities at all can play.

Campus network and/or bandwidth-related issues were identified by Azeez and Van Der Vyver (2018), Idris and Osman (2016), Eze, Chinedu-Eze and Bello (2018), Gupta, et al. (2017), Mahenge and Sanga (2016), Mässing (2017), Khavugwi (2017) and Maphalala and Adigun (2021).

3.5.5 Copyright and incentives

Mutisya and Makokha (2016) noted a demand for copyright to be relinquished to instructors for the material they develop. They also highlighted the need for incentives, as instructors felt that they were not rewarded enough (either financially or with promotion) for their increased efforts to go online. Idris and Osman (2016) reported the same, while Khavugwi (2017) added that payment delays are an inhibitory factor.

3.5.6 U: Technical support

Without providing specific details but assumed to be related to both hardware and software, many authors noted an absence of technical support in an e-learning environment (Patel, Kadyamatimba and Madzvamuse, 2018; Raphael and Mtebe, 2016; Atkins, et al., 2016; Idris and Osman, 2016; Faloye and Ajayi, 2021; Khavugwi, 2017 and Mgeni, et al., 2019). Yakubu and Dasuki (2018) are more specific, pointing to a lack of knowledge, empathy, and response of technical staff, while Pete, Coopasami and Knight (2017) mentioned the lack of support staff. Gupta, et al. (2016) identified the need for technical personnel to work together with instructors to develop attractive and easy-to-use systems.

In summary, 48 specific challenges were identified and will form part of our discussion.

4. Discussion

To best answer our research questions, the 48 specific challenges identified in Section 3 were repackaged into Bervell and Umar's (2017) themes (Table 3). Firstly, and somewhat unfortunately, the specific challenges that resulted in the identification of their themes were limited to the number of challenges presented in Column 1 and a few passing references in the body text to 5 critical challenges (i.e., internet connectivity, computers, other hardware facilities, and lack of computer literacy skills and training) and 4 seemingly minor challenges (i.e., software, institutional/managerial) support and policies and system accessibility). The primary author was approached with a request for a complete list of all challenges per theme, but given the time lapse since their review, the raw data are no longer available. Second, this lack of primary data required us to make some assumptions. However, the reader is reminded of the approach followed in constructing our challenge categories and themes: "Whereas a specific challenge can potentially be seated in multiple categories and/or themes, we followed a best-fit approach". As with the current themes, their themes emerged from their data. Since our review has more data items available, it made sense to refit these to their themes rather than the other way around. Also, since some of their themes relate to our current categories, the latter was fitted into

their themes, accompanied by a concise description of the specific challenges as they were unpacked in Section 3.

Table 3: Comparison of the Bervell and Umar review (2017) with the current review

Bervell and Umar (2017) Themes	**Total Challenges**	**Current review** ***Categories (No. of citations) and associated challenges (numbered)***	**Total Challenges**
System-related (quality)	8	*S: User-friendly LMS (3) and I: User-friendly LMS (1)* LMS is not user-friendly LMS is not always available Unreliability of LMS Lack of support documentation Too many LMS features Poor design of LMS	6
ICT infrastructure	13	*U: ICT infrastructure (14)* Lack of hardware Lack of software Poor internet connections Limited facilities *S: Bandwidth (8)* Lack of bandwidth Expensive bandwidth *S: ICT ownership (3)* Ownership	7
Skills/training	13	*I: LMS training (6)* How to use LMS *S: Readiness (3)* Over-reliance on instructors Technology readiness Lack of awareness *S: Experience (2)* Internet experience *I: Computer skills (4) and S: Computer skills (5)* Computer literacy *S: Familiarity (6)* Unfamiliarity with LMS	7
Technical support	4	*U: Technical support (10)* Lack of hardware support Lack of software support Technical staff knowledge Lack of empathy from the support team Lack of response, unavailability The need to work with technical staff	6
Leadership and management support	7	*U: Environment (3) and U: Financial constraints (2)* Lack of organisational compatibility Lack of organisational funding	6

 ©The Authors

Bervell and Umar (2017) Themes	**Total Challenges**	**Current review** ***Categories (No. of citations) and associated challenges (numbered)***	**Total Challenges**
		New functions implemented without training offered *C: Copyright (1) and I: Incentives (3)* No copyright on the work produced. No incentives for work produced Payment delays for incentives	
Policy issues	7	*U: Policies (3)* Lack of policies	1
Personal issues	3	*S: Finance (1)* Affordability impacts the commitment to e-learning *I: Commitment (7) and I: Instructor role (2)* Negative attitude towards technology use Resistance to change Low use of LMS The commitment and motivation of the instructor impact the use of the LMS by students	5
E-content and e-curriculum	2	*I: Pedagogy (1)* Pedagogical use of LSM tools is lacking *I: Instructional design and support (5)* Shortage of adequately qualified support staff Lack of training in the proper use of e-tools Time-consuming platforms result in limited use Instructors do not have enough skills to create content on their own	5
Time constraints	3	*I: Capacity (6) and S: Capacity (2)* Heavy workload due to large student numbers Lack of skills to manage large student numbers Low return on time investment The complexity of online teaching is time-consuming Demands placed by senior staff on overworked junior ICT staff	6
Totals	60	48	48

4.1 RQ1

RQ1 "What e-learning challenges were identified in research conducted in Higher Education Institutions in sub-Saharan Africa from 2016?" was unpacked in the previous section and is summarised concisely in Column 3. We focus our attention on the number of citations in our review and the challenge totals in both reviews.

Adopting the 3-citation rule and referring to Table 2, the three main and dominant themes were, in order of impact, University-related, Instructor-related, and Access-related. Secondary themes were LMS-, Computer literacy- and Student-related. In general, the challenge categories of *ICT infrastructure* (14), T*echnical support* (10), and *Student bandwidth* (8) attracted the most citations, followed by the secondary challenge categories *Instructor* C*ommitment* (4) and *Instructor Capacity* (6), *LMS training for instructors* (6), *Instructional design and support* and S*tudent computer skills* (5 each). Other than *Student bandwidth* and S*tudent computer skills,* the secondary challenge categories are all related to the instructor.

ICT infrastructure challenges can be solved by sharing limited infrastructure and investment, which will help to lower the cost burden (CRASA & ITU, 2016), as well as using technologies such as cloud computing. The latter option, as an on-demand option, is particularly appealing since it cost-effectively provides services without the

need to install and update software and hardware (Shibi, Kadiri and Aknin, 2013). If the company that provides the service is furthermore appointed to provide *Technical support*, part of the team can be reassigned and offer workshops that focus on *LMS training for instructors* and S*tudent computer skills.*

Once these challenges are resolved, the last mile problem remains - bandwidth, especially for off-campus students. Here it is worth noting that internet access in SSA has nearly doubled from 16% to 30% in the period 2015-2020 (World Bank, 2020). However, such access ranges from 68% in South Africa to 25% in Tanzania (the two countries with the most citations), to the lowest penetration of 7% in the Central African Republic (Kamer, 2022). Innovative and novel solutions such as Facebook and Google's composite terrestrial, high altitude, and satellite systems, and a wide range of advanced photonics technologies and solar power offer promising solutions (Lavery, et al., 2018), but these come at a great cost. Until bandwidth is resolved, e-learning approaches will have to be adapted accordingly, e.g., low bandwidth and asynchronous versions of key content and activities, reducing file sizes by compression and limiting videos, etc.

With reference to the instructor-related categories, employee resistance to technological change has been identified as far back as 38 years by Ewert (1984), who concluded that it is closely related to, amongst other aspects, skill levels required to use new technologies. Miller (2019), in contrast, found that employees are not necessarily resistant, and can become excited (committed) with continued experience, while a well-established LMS reinforces instructor commitment (San-Martín, et al., 2020). It is feasible to suggest that a skilled and energetic *Instructional design and support* team can go a long way in helping to resolve *Instructor Commitment* and *Instructor Capacity*, as well as offer *LMS training for instructors.* However, as noted by Kumar and Ritzhaupt (2017), a better understanding of the exact role of instructional designers is needed, including instructor and student development and support; technical responsibilities such as technical support, web design, and LMS implementation; participation in committees; and summative evaluation of courses and initiatives. The above suggests that they are often not used to their fullest.

4.2 RQ2

To answer RQ2 "Which e-learning challenges identified before 2016 remain and which are new? ", the discussion focuses on Table 3. It is noted that a large part of Bervell and Umar's (2017) review of 31 studies focused on research reporting on intentions and perceptions of LMS use (TAM and UTAUT), as well as the methodological approach used. In addition to their focus on key determinants of acceptance/adoption of LMS acceptance/adoption (Attitude and perceived usefulness, performance expectation, perceived ease of use, and Social Influence), the rest of their inclusion and exclusion criteria mirror ours. Since the current review is interested in the real-world use of an e-learning system/approach, the discussion is limited to implementation challenges identified in 18 of the 31 reviewed studies. Although rudimentary, it nonetheless offers an opportunity to view the current data from a different perspective, thereby contributing to a richer interpretation.

They identified 9 themes and 60 challenges, namely ICT infrastructure and skills/training (13 challenges each), system-related (8), leadership and management support (7), policy issues (7), technical support (4), personal issues (3), e-content/e-curriculum (2) and time constraints (3). In the current review, six themes, 25 challenge categories, and 48 specific challenges were identified. At both the theme and some of the current challenge category levels, it is evident that many challenges remain. However, the challenge totals (60 versus 48), show that fewer challenges were reported in the period 20016-2022. For most themes, the difference in total challenges is negligible. For example, in the System-related (quality) theme, they identified 8 challenges against our 6. For Technical support, it is 4 against 6.

Three observations are made. First, despite the shortcomings in the comparison, no new challenges are evident. Despite the different naming conventions, a reader who has followed research on this topic over the past two decades will immediately recognise many of the themes, challenge categories, and specific challenges. Furthermore, our challenge categories, although emerging, emulate those identified in studies elsewhere in the world. Examples are Instructor and Student Characteristics (Selim, 2007; Alhabeeb and Rowley, 2018) and Technology and/or Support (Selim, 2007; Teoh, 2011; Alhabeeb and Rowley, 2018; Miranda, et al., (2017); McPherson and Nunes, 2006); Management Support, Pedagogy and Learning Approach (Teoh, 2011); E-learning Systems and Online Learning Resources (Alhabeeb and Rowley, 2018); Content and Stakeholders (Miranda, et al., 2017), and Leadership, Structural and Cultural Issues, Design and Delivery Issues (McPherson and Nunes, 2006), and Technological, Infrastructural and Contextual categories (Oyerinde, 2014).

 ©The Authors

Second, and noteworthy, is the large reduction in skills/training and ICT infrastructure challenges (both from 13 to 7) and policy issues (7 to 1). The totals suggest that with more policies and ICT infrastructure in place, e-learning has become more mainstream, with the result that instructors and students are likely to have more computer skills, are more familiar with e-learning approaches resulting in more experience and less need for training and are therefore more ready to embrace e-learning. However, with e-learning becoming more mainstream, slight increases in challenges related to aspects of technical support, personal issues, e-content, and e-curriculum, as well as time constraints, can be expected. This notion is supported by the challenge of instructor commitment as identified in our review. Its pervasiveness (7 citations) and potential associations across and within other challenge categories, e.g., a lack of incentives and capacity due to heavy workloads, larger student numbers, and the complexity of teaching online, continue to make it difficult for instructors to fully commit to e-learning initiatives and are therefore perhaps worthy of being recognised as an explicit category that deserves specific attention. Once again, the role an active instructional design and support team can play is highlighted.

Finally, in reporting the determinants of LMS acceptance/adoption, they identified several other factors, which appear to be direct challenges, the most relevant to our results being skilled personnel, previous experience, instructor competencies, and course quality. In general, the number of citations for specific challenges reported shows that most challenges are context-specific, that is, limited to a single HEI. However, when combined in challenge categories and reflected as themes, the review reveals many challenges to remain persistent.

5. Conclusions

In concluding the systematic review, research since 2016 has reinforced many challenges identified before 2016. No new challenges were identified. However, if one considers that very few of the listed challenges after 2016 were identified in all HEIs and that all challenges were not present in a single HEI, our review suggests that the challenges identified from 2016 are more context-specific than in previous years, a conclusion that is supported by the low number of citations for many challenges (three and under). At the intersection of the order of challenge categories and the number of citations, it is appropriate to conclude that a top-down yet holistic approach is required to resolve the remaining challenges. There appears to be a move in this direction, with more policies in place to ensure that instructors and students are stepwise geared and supported for specific HEI contexts and circumstances. However, the adage remains that one simply cannot 'throw' more technology at instructors and students and expect it to solve challenges. A range of evaluation checklists, models/frameworks, and strategies for ensuring the successful implementation of e-learning in HEIs (e.g., Atwell, 2006; Fee, 2014; Anstey and Watson; 2018), and specifically for developing countries (Hadullo, Oboko and Omwenga, 2017) have been proposed. A recommendation for further research is the development of an updated SSA-specific checklist that can be used to rate e-learning preparedness in context.

Acknowledgements

This research originated in an Open Distance Learning Research Support Program at the University of South Africa.

References

Adkins, S.S. 2016. The 2016-2021 worldwide self-paced eLearning market: The global eLearning market is in steep decline. *Ambient Insight report*, pp. 1-90. Available at: <https://uhlberg-advisory.de/wp-content/uploads/2017/06/AmbientInsight_The-2016-2021_Worldwide_Self-paced-eLearning_Market.pdf> [Accessed 16 August 2022].

Al-Fraihat, D., Joy, M. and Sinclair, J., 2017, June. Identifying success factors for e-learning in higher education. In: *12th International Conference on e-Learning (ICEL 2017),* Orlando, Florida, 01-02 Jun 2017. Published in: Proceedings of the 12th International Conference on e-Learning (ICEL 2017), London, United Kingdom October 23 - 26, 2020, pp. 247-255. ISBN 9781911218364

Alhabeeb, A. and Rowley, J., 2018. E-learning critical success factors: Comparing perspectives from academic staff and students. *Computers & Education*, 127, pp.1-12.

Ansong, E., Lovia Boateng, S. and Boateng, R., 2017. Determinants of e-learning adoption in universities: Evidence from a developing country. *Journal of Educational Technology Systems*, 46(1), pp.30-60.

Anstey, L.M. and Watson, G.L.P., 2018. Rubric for eLearning tool evaluation. Available at: < https://er.educause.edu/articles/2018/9/a-rubric-for-evaluating-e-learning-tools-in-higher-education.> [Accessed 22 December 2022].

Asalla, L.K., Putri, M.R. and Pradipto, Y.D., 2017, November. The critical success factor of E-learning in higher education: A systematic literature review. *2017 International Conference on Information Management and Technology (ICIMTech),* Special Region of Yogyakarta, Indonesia, 2017, [online] pp.192-197, https://doi.org/10.1109/ICIMTech.2017.8273536

Atkins, S., Yan, W., Meragia, E., Mahomed, H., Rosales-Klintz, S., Skinner, D., Zwarenstein, M. and ARCADE consortium, 2016. Student experiences of participating in five collaborative blended learning courses in Africa and Asia: a survey. *Global Health Action*, 9(1), p.28145.

Attwell, G., 2006. Evaluating E-learning: A Guide to the Evaluation of E-learning. *Evaluate Europe Handbook Series*, *2*(2), pp.1610-0875.Azeez, N.A. and Van Der Vyver, C., 2018, November. Digital education: assessment of e-learning and m-learning adoption in tertiary institutions in South Africa. In *2018 IEEE Conference on e-Learning, e-Management and e-Services (IC3e),* [online] pp. 23-28. https://doi.org/10.1109/IC3e.2018.8632654

Basak, S.K., Wotto, M. and Bélanger, P., 2016. A framework on the critical success factors of e-learning implementation in higher education: A review of the literature. *International Journal of Educational and Pedagogical Sciences*, 10(7), pp.2409-2414.

Bervell, B. and Umar, I.N., 2017. A decade of LMS acceptance and adoption research in Sub-Sahara African higher education: A systematic review of models, methodologies, milestones and main challenges. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(11), pp.7269-7286.

Basitere, M. and Ivala, E.N., 2017, June. Implementation of Blended Learning: The experiences of students in the Chemical Engineering Extended Curriculum Program Physics course. In *Proceedings of the International Conference on e-Learning (ICEL 2017), Orlando, Florida, 1-2 June 2017* (pp. 23-30).

Cheawjindakarn, B., Suwannatthachote, P. and Theeraroungchaisri, A., 2013. Critical success factors for online distance learning in higher education: A review of the literature. *Creative Education*, 3(08), p.61.

CRASA & ITU, 2016. SADC ICT and broadcasting infrastructure sharing guidelines. Communications Regulators' Association of Southern Africa (CRASA) and International Telecommunications Union (ITU) [online] Available at: https://www.crasa.org/post-articles/sadc-ict-and-broadcasting-infrastructure-sharing-guidelines [Accessed 1 September 2023].

Elkaseh, A., Wong, K.W. and Fung, C.C., 2015. A review of the critical success factors of implementing e-learning in higher education. *The International Journal of Technologies in Learning*, 22(2), pp.1-13.

Eltahir, M.E., 2019. E-learning in developing countries: Is it a panacea? A case study of Sudan. *IEEE Access*, *7*, pp.97784-97792.

Ewert, A., 1984. Employee resistance to computer technology. *Journal of Physical Education, Recreation & Dance*, 55(4), pp.34-36.

Eze, S.C., Chinedu-Eze, V.C. and Bello, A.O., 2018. The utilisation of e-learning facilities in the educational delivery system of Nigeria: a study of M-University. *International Journal of Educational Technology in Higher Education*, 15(1), pp.1-20.

Faloye, S.T. and Ajayi, N., 2021. Understanding the impact of the digital divide on South African students in higher educational institutions. *African Journal of Science, Technology, Innovation and Development*,14(4), pp.1-11.

Fatima, N. and Abu, K.S., 2019. E-learning research papers in Web of Science: A bibliometric analysis. *Library Philosophy and Practice*, 2144, pp.1-14.

Fee, K. (2014) *Delivering E-Learning*. Philadelphia. Kogan Page Ltd.

Françoise, O., Bourgault, M. and Pellerin, R., 2009. ERP implementation through critical success factors' management. *Business process management journal*, 15(3), pp.371-394.

Freund, Y.P., 1988. Critical success factors. *Planning Review*, 16(4), pp.20-23.

Gani, F. and van den Berg, G., 2019. Lecturers' Perceptions of the Use of Learning Management Systems: A Case Study in Open Distance Learning. *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)*, 14(3), pp.15-27.

Gupta, M., Marsden, S., Oluka, T., Sharma, R. and Lucas, H., 2017. Lessons learned from implementing e-learning for the education of health professionals in resource-constrained countries. *Electronic Journal of e-Learning*, 15(2), pp.144-155.

Gusenbauer, M., 2019. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics*, 118(1), pp.177-214.

Gusenbauer, M., 2022. Search where you will find most: Comparing the disciplinary coverage of 56 bibliographic databases. *Scientometrics*, 127(5), pp.2683-2745.

Hadullo, K., Oboko, R. and Omwenga, E., 2017. A model for evaluating e-learning systems quality in higher education in developing countries. *International Journal of Education and Development using ICT*, 13(2), pp.185-204.

Harzing, A.W. (2007) Publish or Perish. Available at: <https://harzing.com/resources/publish-or-perish> [Accessed 04 December 2022].

Higgins, J.P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J. and Welch, V.A. eds., 2019. *Cochrane handbook for systematic reviews of interventions*. Chichester (UK): John Wiley & Sons.

Idris, F. and Osman, Y., 2016. Implementation of e-learning in the University of Gezira barriers and opportunities. [online] https://doi.org/10.22496/esr2016090470

Jackson, J.L. and Kuriyama, A., 2019. How often do systematic reviews exclude articles not published in English? *Journal of General Internal Medicine*, [online] 34, pp.1388-1389. https://doi.org/10.1007/s11606-019-04976-x

Kamer, L. (2022). Share of internet users in Africa as of January 2022, by country. [online] Statista. Available at: <https://www.statista.com/statistics/1124283/internet-penetration-in-africa-by-country> [Accessed 22 December 2022].

Khavugwi, B.M., 2017. *Evaluating factors explaining lecturer adoption of E-learning at the e-campus of Maseno University* (Doctoral dissertation, Maseno University).

Kulikowski, K., Przytuła, S.F and Sułkowski, Ł., 2022. E-learning? Never again! On the unintended consequences of COVID-19 forced e-learning on academic teacher motivational job characteristics. *Higher Education Quarterly*, 76(1), pp.174-189.

Kumar, S. and Ritzhaupt, A., 2017. What do instructional designers in higher education really do? *International Journal on E-Learning,* 16(4), pp.371-393.

Lavery, M.P., Abadi, M.M., Bauer, R., Brambilla, G., Cheng, L., Cox, M.A., Dudley, A., Ellis, A.D., Fontaine, N.K., Kelly, A.E. and Marquardt, C., 2018. Tackling Africa's digital divide. *Nature Photonics*, 12(5), pp.249-252.

McPherson, M. and Nunes, M.B., 2006. Organisational issues for e-learning: Critical success factors as identified by HE practitioners. *International Journal of Educational Management*, 20(7), pp.542-558.

Mafunda, B. and Swart, A.J., 2020. Determining African students' e-learning readiness to improve their e-learning experience. *Global Journal of Engineering Education*, 22(3), pp.216-221.

Mahenge, M.P. and Sanga, C., 2016. ICT for e-learning in three higher education institutions in Tanzania. *Knowledge Management & E-Learning: An International Journal*, 8(1), pp.200-212.

Maphalala, M.C. and Adigun, O.T., 2021. Academics' Experience of Implementing E-Learning in a South African Higher Education Institution. *International Journal of Higher Education*, 10(1), pp.1-13.

Mässing, C., 2017. *Success factors and challenges for E-learning information systems in the Namibian higher education system: A case study of the University of Namibia*. [Unpublished Bachelor's Degree. University of Skövde*]*.

Miller, T., 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267, pp.1-38.

Miranda, P., Isaias, P., Costa, C.J. and Pifano, S., 2017. Validation of an e-learning 3.0 critical success factors framework: A qualitative research. *Journal of Information Technology Education*, (1), pp.339-363.

Mutisya, D.N. and Makokha, G.L., 2016. Challenges affecting the adoption of e-learning in public universities in Kenya. *E-Learning and Digital Media*, 13(3-4), pp.140-157.

Mgeni, M.S., Ismail, M.J., Yunus, S.A. and Haji, H.A., 2019. The contribution of LMS to the learning environment: views from the State University of Zanzibar. In e-Infrastructure and e-Services for Developing Countries, *10th EAI International Conference, AFRICOMM 2018.* Dakar, Senegal, November 29-30, 2019. Springer International Publishing.

Msomi, A.M. and Bansilal, S., 2018. The experiences of first-year students in mathematics in using an e-learning platform at a university of technology. *South African Journal of Higher Education*, 32(5), pp.124-129.

Mtebe, J.S., 2020. Examining user experience of eLearning systems implemented in two universities in Tanzania. *Interactive Technology and Smart Education*, 17(1), pp.39-55.

Olasina, G., 2019. Human and social factors affecting the decision of students to accept e-learning. *Interactive Learning Environments*, 27(3), pp.363-376.

Oyerinde, O.D., 2014. A review of challenges militating against successful e-learning and m-learning implementations in developing countries. Available at: <https://www.researchgate.net/publication/264047648_A_Review_of_Challenges_Militating_against_Successful_E-Learning_and_M-Learning_Implementations_in_Developing_Countries > [Accessed 01 December 2022].

Patel, N.M., Kadyamatimba, A. and Madzvamuse, S., 2018, October. E-learning adoption in rural-based higher education institutions in South Africa. In *2018 Open Innovations Conference (OI)* (pp. 320-324). IEEE.

Pete, M., Coopasami, M. and Knight, S., 2017. e-Learning readiness amongst nursing students at the Durban University of Technology. *Health SA gesondheid*, 22(1), pp.300-306.

Qwabe, B.P. and Khumalo, M., 2020. Students' perspective regarding usefulness of e-learning in a South African University of Technology. *International Journal of Entrepreneurship*, 24(2), pp.1-13.

Raphael, C. and Mtebe, J., 2016. Instructor support services: An inevitable critical success factor in blended learning in higher education in Tanzania. *International Journal of Education and Development using ICT*, 12(2), pp. 123-138.

San-Martín, S., Jiménez, N., Rodríguez-Torrico, P. and Piñeiro-Ibarra, I., 2020. The determinants of teachers' continuance commitment to e-learning in higher education. *Education and Information Technologies*, 25(4), pp.3205-3225.

Selim, H.M. 2007. E-learning critical success factors: an exploratory investigation of student perceptions. *International Journal of Technology Marketing*, 2(2), pp.157-182.

Shibi, B., Kadiri, K., and Aknin, N., 2013. Towards collaborative learning process based on hybrid cloud computing and Web 2.0 tools. *International Journal of Engineering and Technology*, 5(3), pp. 2913-2919.

Strauss A. and Corbin, J. 2008. *Basics of qualitative research: Grounded theory procedures and techniques*. 2nd Edition. Sage.

Teoh, K.K., 2011. An examination of critical success factors in the implementation of ePortfolios in universities. *Journal of Academic Language and Learning*, 5(2), pp. A60-A72.

Tibaná-Herrera, G., Fernández-Bajón, M.T. and Moya-Anegón, D., 2018. Categorization of e-learning as an emerging discipline in the world publication system: a bibliometric study in SCOPUS. *International journal of educational technology in higher education*, 15(1), pp.1-23.

UN. 2011. Africa's least developed: lands of opportunity. Africa Renewal, August 2011. United Nations Publications. Geneva. Available at: <https://www.un.org/africarenewal/magazine/august-2011/africas-least-developed-lands-

opportunity#:~:text=Since%201971%20when%20the%20least,regional%20presence%20in%20the%20group>[Accessed 20 February 2023].

UN. 2022. The least developed countries report 2022. United Nations Publications. Geneva. Available at: <https://unctad.org/system/files/official-document/ldc2022_en.pdf> [Accessed 19 February 2023]

UNESCO, 2021. The Digital Learning Turn in Africa: The Role of Local Ecosystems. ED/GEC/2021/03. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000377725?posInSet=1&queryId=18b1f9bd-1b10-4893-a941-86032e55a728> [Accessed 1 December 2022].

UNICEF and African Union Commission, 2021. Transforming Education in Africa: an evidence-based overview and recommendations for long-term improvements. Available at: <https://www.unicef.org/reports/transforming-education-africa.> [Accessed 1 December 2022].

UNICEF, 2021. *Transforming Education in Africa. An evidence-based overview and recommendations for long-term improvements. Report*. Available at: <https://www.unicef.org/reports/transforming-education-africa> [Accessed 12 November 2022].

Van der Merwe, T.M.; De Villiers. M.R., 2011. The partial approach to Grounded Theory integrated with Activity Theory: A generic framework, illustrated by a base study in an e-learning context. *Alternation, Special Edition Number 4*, pp. 365-402.

Wagner, E., 2011. Essay: In search of the secret handshakes of ID. *The Journal of Applied Instructional Design*, 1(1), pp.33-37.

World Bank, 2020. *Global economic prospects, June 2020*. The World Bank.

Yakubu, M.N. and Dasuki, S., 2018. Assessing eLearning systems success in Nigeria: An application of the DeLone and McLean information systems success model. *Journal of Information Technology Education: Research*, 17, pp.183-203.

Zozie, P. and Chawinga, W.D., 2018. Mapping an open digital university in Malawi: Implications for Africa. *Research in Comparative and International Education*, 13(1), pp.211-226.

 ©The Authors

Methods for Evaluating Learning Analytics and Learning Analytics Dashboards in Adaptive Learning Platforms: A Systematic Review

Tobias Alexander Bang Tretow-Fish and Md. Saifuddin Khalid
Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark

tabtr@dtu.dk
skhalid@dtu.dk

Abstract: This research paper highlights and addresses the lack of a systematic review of the methods used to evaluate Learning Analytics (LA) and Learning Analytics Dashboards (LAD) of Adaptive Learning Platforms (ALPs) in the current literature. Addressing this gap, the authors built upon the work of Tretow-Fish and Khalid (2022) and analyzed 32 papers, which were grouped into six categories (C1-6) based on their themes. The categories include C1) the evaluation of LA and LAD design and framework, C2) the evaluation of user performance with LA and LAD, C3) the evaluation of adaptivity, C4) the evaluation of ALPs through perceived value, C5) the evaluation of Multimodal methods, and C6) the evaluation of the pedagogical implementation of ALP's LA and LAD. The results include a tabular summary of the papers including the categories, evaluation unit(s), methods, variables and purpose. While there are numerous studies in categories C1-4 that focus on the design, development, and impact assessment of ALP's LA and LAD, there are only a few studies in categories C5 and C6. For the category of C5), very few studies applied any evaluation methods assessing the multimodal features of LA and LADs on ALPs. Especially for C6), evaluating the pedagogical implementation of ALP's LA and LAD, the three dimensions of signature pedagogy are used to assess the level of pedagogy evaluation. Findings showed that no studies focus on evaluating the deep or implicit structure of ALP's LA. All studies examine the structural surface dimension of learning activities and interactions between students, teachers, and ALP's LA and LAD, as examined in categories C2-C5. No studies were exclusively categorized as a C6 category, indicating that all studies evaluate ALP's LA and LAD on the surface structure dimension of signature pedagogy. This review highlights the lack of pedagogical methodology and theory in ALP's LA and LAD, which are recommended to be emphasized in future research and ALP development and implementation.

Keywords: Adaptive Learning Platform, Learning Analytics, Systematic literature review, Methods of evaluation

1. Introduction

The field of Adaptive Learning (AL) is a relatively new area of research that spans multiple disciplines and involves numerous synonyms and definitions. While terms such as personalized learning, individualized learning, intelligent tutoring, and customized learning are sometimes used interchangeably, adaptive learning is the most used term (Shemshack and Spector, 2020). Adaptive learning platforms (ALPs) have gained significant attention in recent years, leading to the development of various methods to design and evaluate personalized activities and content. However, assessing ALPs is a complex and multi-dimensional process involving several factors such as Learning Analytics (LA), Learning Analytics Dashboards (LAD), system usability, user perception, use, and pedagogy.

While numerous reviews on adaptive learning and learning analytics exist, few studies have specifically focused on evaluating ALPs. Martin, Denne and Bonk (2020) synthesizes several systematic reviews on adaptive learning and learning analytics. The synthesis on adaptive learning includes: (1) A meta-analysis to address the question: Can students improve their knowledge when the system adapts to their profile and performance? (2) An analysis of 42 studies on source of adaptation focusing on learner and learner environment interaction; analysis of 29 studies on adaptation of content, presentation and instruction; 25 studies on rule-based, probability-based, or other adaptive pathways, (3) Content analysis involving 70 studies addressing learning styles in adaptive educational hypermedia systems. (4) Analysis of 98 studies determining characteristics for learner models in adaptive systems. (5) Document analysis of 78 studies and reviews on learning styles in adaptive systems, (6) Review of 78 studies that explored learner, dimensions of these traits, and identification techniques for these traits in adaptive learning systems, (7) 61 empirical studies were reviewed on adaptive sources based on learner models and adaptive targets based on content and instructional models, and (8) A qualitative thematic analysis of 62 studies and a subset of 12 studies on experimental designs for meta-analysis to study the effects of adaptivity in educational games. The review of reviews on learning analytics (Martin, Dennen and Bonk, 2020) include (1) Analysis of 40 studies identifying the research objectives and methods applied in learning analytics and educational datamining studies, (2) reviewed 44 studies to examine learning analytics methods, benefits, and challenges in higher education, (3) examined 43 studies on applied research design, topic of study,

Reference this paper: Tretow-Fish, T.A.B. and Khalid, M.S., 2023. Methods for Evaluating Learning Analytics and Learning Analytics Dashboards in Adaptive Learning Platforms: A Systematic Review, *The Electronic Journal of e-Learning, 21(5)*, pp 430-449, available online at www.ejel.org

educational context, learning scenario, pedagogical practices, learning platform, technology tools, and methodological techniques, (4) conducted a review on 52 studies focusing on visual learning analytics of educational data identifying approaches, audience, purposes, contexts, and data sources. (5) did a systematic review of 107 studies identifying what data was collected, modeling methods, research themes, system evaluation, and similarities and differences between open learner models and learning analytic dashboards, (6) reviewed 29 papers examining learning analytic dashboards from a self-regulated learning perspective, (7) 11 studies were reviewed to examine the efficacy of learning analytics interventions in higher education which (8) build upon where 46 studies were inspected to address whether and to what extent learning analytics were successful in providing study success in higher education.

Mousavinasab et al. (2021) conducted a review of 53 studies that investigated the variant characteristics of Intelligent Tutoring Systems across various educational fields. In their review, they raised several questions concerning the methods employed to assess these systems. For example, the importance of the learners' role in evaluating intelligent tutoring systems is evident, especially when assessing system usability. Only 5.66% of the reviewed studies evaluated intelligent tutoring systems based solely on learner experiences. However, in studies where learner experiences were combined with other variables, such as learner and system performance, learner experiences were more frequent. The review fails to provide information on the specific methods utilized to obtain the learner experience or the types of usability tests employed. Understanding the various evaluation perspectives and assessment methods utilized in evaluating Learning Analytics (LA) and Learning Analytics Dashboards (LADs) is crucial. The absence of comprehensive insights on evaluation methods for LA and LADs is the driving force behind this systematic review. So, this review expands the work of Tretow-Fish and Khalid (2022) and aims to synthesize evaluation methods used in the design, development, and implementation of ALP's LA and LAD to support pedagogical and learning-related decisions for educators and students. The review will analyze the research design, frameworks, methods, and instruments utilized to evaluate ALP's LA and LAD, the resulting impact, and the integration of students' and educators' perceptions of LAD and LA into the evaluation methods. The research will significantly contribute to the field of usability engineering, user experience, and digital learning technology. Investigating the evaluation methods applied to ALPs is crucial to enhance the quality of the learning experience and outcomes, improving educators' teaching experiences and their technology adoption, aiding the development process in companies, and ensuring the proper implementation of evaluation methods. The scope mentioned above, and motivation led us to devise the research question:

How to evaluate the Learning Analytics and Learning Analytics Dashboards of Adaptive Learning Platforms?

The objective is to identify a set of methods for evaluating the technological features' functionalities and perceived experiences and another set of methods for demonstrating the evidence of improving learning outcomes, learning experience, and teaching quality. The researchers and practitioners will be able to apply the synthesized methods and instruments for the evaluation of learning analytics and dashboards in the contexts of digital platforms and for the assessment of impact. The review will provide insights for identifying the scope of future research by providing an overview of how LA and LADs of ALPs are evaluated, with what purpose, and on which variables. Furthermore, the outcome can advance the field of interaction design, while the latter can contribute to the broader domain of service design and innovation in education and training.

2. Methods

The selection of papers and the process of analysis and synthesis is conducted using two distinct established methods.

2.1 Selection of Papers: PRISMA

The paper selection process follows the four phases of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol (Page *et al.*, 2021), which are identification, screening, eligibility, and inclusion (see Figure 1). To review the evaluation methods utilized for Learning Analytics (LA) and Learning Analytics Dashboards (LAD) on Adaptive Learning Platforms (ALPs), a range of keywords including evaluation, adaptive learning, learning analytics, learning analytics dashboards, assessment, and others are employed in different combinations. The search was limited to articles that had been peer-reviewed and published in English, Danish, and Norwegian (considering the authors' language abilities) from 2011 up to the search deadline of September 1, 2021.

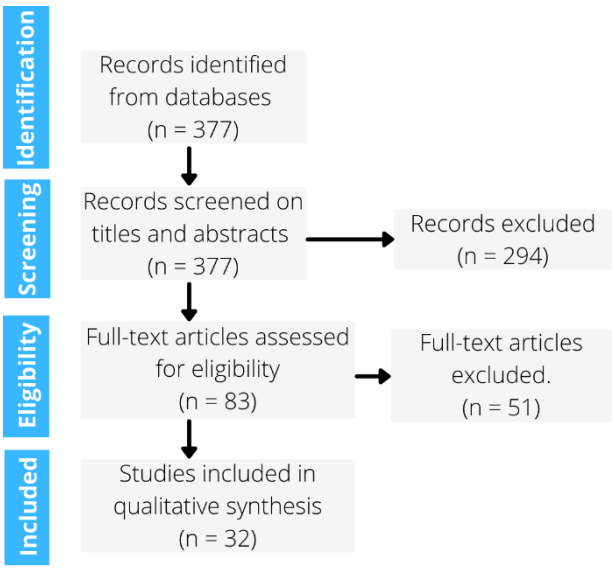

Figure 1: PRISMA Flow-Chart

After testing different keyword combinations and consulting with a librarian, four databases were selected. Various combinations of the keywords returned the following results: Scopus (n=75), ACM (n=144), ScienceDirect (n=106), ERIC (n=14), and Taylor Francis (n=38). The screening and eligibility stages involved applying specific exclusion criteria, which are as follows: (1) Papers that do not primarily investigate the LA or LAD of an ALP were excluded. (2) Papers focusing on LA and LAD in other e-learning environments that do not meet the adaptivity requirements of the learning platform were excluded. (3) Papers without empirical data examining LA or LAD on ALP or those that do not primarily focus on LA and LAD evaluation were excluded. (4) Only included papers published in the main conference proceedings, while workshop papers and posters were excluded. The two authors independently screened different databases, and only the papers selected by one author (n=83) were included in this document. For this review, 32 articles and two reviews were included for analysis and synthesis.

2.2 Constant Comparative Analysis Method

We applied the constant comparative analysis method (Hewitt-Taylor, 2001) for the analysis and synthesis. The articles were encoded according to themes and then divided into categories and subcategories. During this process, the coded sections were regularly compared to similar parts of texts containing the same codes. The intention was to connect the texts and ensure the continuity of the codes' definitions (ibid.). Each included paper was read to identify methods, variables, and purpose of evaluating LA and LAD. The data extracted from the papers are tabulated to synthesize: (1) The methods used when evaluating LA and LAD. (2) Variables measured by the methods to evaluate LA and LAD. (3) The purpose of the evaluation method applied. A thematic analysis was initiated from the identified purposes, and categories were developed.

2.3 Signature Pedagogy for Analytical Thematization in C6

To thematize the category of C6) an evaluation of the pedagogical implementation of ALP's LA Shulman's theory of signature pedagogics was applied (Shulman, 2005). The application was to identify and categorize the levels of pedagogy evaluations which coincide with signature pedagogy's dimensions of the surface, deep, and implicit structure. Surface structure describes learning activities in an educational context, such as interactions between teacher, student, and technology, or concrete learning activities, such as reading, discussing, and completing assignments on an ALP. Deep structure describes the context and pedagogic structures that learning activities exist in, such as Flipped Classroom, Inquiry-Based Learning, and Dialogic Teaching. Lastly, the implicit structure represents the inherent values associated with the surface, and deep structures of the signature pedagogy, such as Flipped Classroom's purpose of distributing the workload to student preparation implies an intrinsic value contribution to in-class learning activities.

 ©The Authors

3. Analysis and Synthesis

We report the results of a qualitative synthesis of a systematic review. Evaluation methods identified in the papers are grouped into six categories based on their main focus. These categories are: C1) evaluating design and framework, C2) performance, C3) adaptivity, C4) perceived value, C5) multimodal methods, and C6) pedagogical implementation. These categories are further described in the following sections. Each category has multiple themes to specify the studies' focus. Papers may be presented in multiple themes within a category. Some studies used multiple evaluation methods. We review and map these studies at the end of each category's section (see Table 2).

The distribution of the identified evaluation categories and how they are connected to each other is visualized in figures 2 and 3 showing a low representation of studies evaluating multimodal methods (C5) and pedagogical implementation (C6). 27 studies contribute to multiple evaluation categories and only five papers fall under single category. In addition, in literature, there is a preference for variables used to assess and describe students and platforms compared to variables assessing and describing teachers (see Table 1).

Furthermore, the most frequently applied methods of evaluation are visualized in figure 4 where Logs on user activity (n=11), Interviews (n=5), and Pre- and Post-test (n=5) are the most preferred methods used.

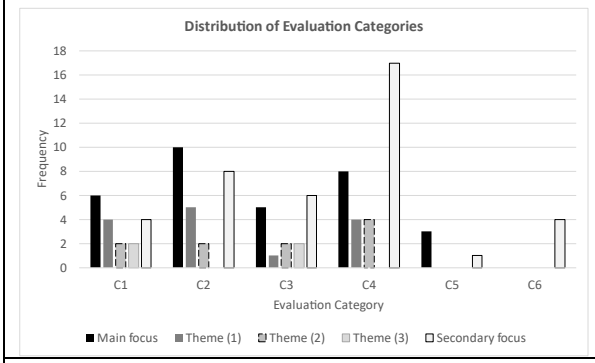

Figure 2: The distribution of evaluation categories on the evaluated papers

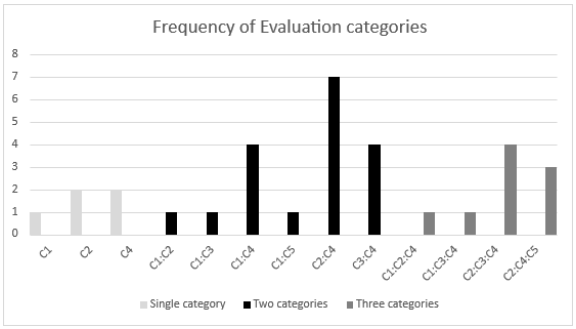

Figure 3: Frequency of primary focus evaluation categories and their combinations in studies

Table 1: Variables and their category association used by Learning Analytics and Learning Analytics Dashboards either describing or being described by students, teachers, or the platform when evaluating Adaptive Learning Platforms

	Variables describing			**Variables being described by**
Students	(C2) Time	(C2) Competence	(C2) Score	(C2) Self-assessment
	(C2) Reading competence	(C2) Completion time	(C2) Performance score	(C2:C4:C5) Perception on performance
	(C1:C4) Learning styles	(C1:C4) Learners	(C1:C4) Confidence and trust	(C2) Self-completion
	(C1:C4) User behaviour	(C1:C4) User interactions	(C1:C5) Effort	
	(C1:C5) Response time	(C1:C5) Performance	(C1:C5) Arousal	
	(C2:C4) Students' learning performance	(C2:C4) Performance	(C2:C4) Students' performance	
	(C2:C4) LAD use	(C2:C4) Satisfaction with LAD use	(C2:C4) Confidence	
	(C2:C4) Satisfaction	(C2:C4) Learning attitude	(C2:C4) Technology acceptance	

	Variables describing			**Variables being described by**
	(C2:C4) Problem solving activities	(C2:C4) Learning performance	(C2:C4) Learning experience	
	(C2:C4) Learning flow	(C2:C4) Learning performance	(C2:C4) Perceived learning experience	
	(C2:C4) Students' attitude on system	(C2:C4) Learning outcomes	(C3:C4) Linguistic category features	
	(C3:C4) Users attitudes towards the system	(C3:C4) User performance	(C3:C4) Learner's personality	
	(C3:C4) User experience of cognitive load	(C3:C4) Language complexity measures	(C1:C2:C4) Performance	
	(C1:C2:C4) Learner satisfaction	(C1:C2:C4) Motivation	(C1:C2:C4) Study time	
	(C1:C2:C4) Cognitive styles	(C1:C2:C4) Previous relevant knowledge	(C2:C3:C4) Improvement of learning scores	
	(C1:C3:C4) Learning styles dimensions	(C1:C3:C4) Learning effectiveness	(C2:C3:C4) Learner's satisfaction scores	
	(C2:C3:C4) Learning effectiveness	(C2:C3:C4) Performance scores	(C2:C4:C5) Response time	
	(C2:C4:C5) Learner physiological signals	(C2:C4:C5) Learners affective state	(C2:C4:C5) Emotions	
	(C2:C4:C5) Arousal	(C2:C4:C5) Cognitive load	(C2:C4:C5) Attention	
	(C2:C4:C5) Playing accuracy	(C2:C4:C5) Playing speed	(C2:C4:C5) Goal orientation	
Teachers	(C1:C4) Teacher perceptions	(C2:C4) Technology acceptance	(C2:C4) Teacher's performance	(C1:C4) Teacher expectations
	(C2:C4) LAD use	(C2:C4) Satisfaction with LAD use	(C2:C4) Technology acceptance	
	(C2:C4:C5) Educators' tacit experiences,	(C2:C4) Teachers' perspective on system in praxis		
Platform	(C1) User centred design principles	(C1) E-learning life cycles	(C1) Layered evaluation of adaption features	(C1:C2) Prediction of performance
	(C2) Reading Performance assessment	(C4) Feasibility for virtual mentoring	(C4) Adjustability	
	(C4) Feasibility for virtual mentoring	(C4) Satisfaction	(C4) Reliability	(C3:C4) System performance
	(C4) Delightfulness	(C1:C4) Context	(C1:C4) Learning objects	(C1:C3) Recommendations meeting requirements of learners
	(C1:C3) Accuracy of learning material estimation	(C1:C4) Threshold levels	(C1:C4) Effectiveness	(C2:C4) Homogeneity in grouping through peer collaboration features in system.

 ©The Authors

	Variables describing			**Variables being described by**
	(C1:C4) Devices	(C1:C4) Usefulness	(C1:C4) Information adequacy	
	(C1:C4) Usefulness	(C1:C4) Novelty	(C1:C4) Serendipity	
	(C1:C4) Accuracy	(C1:C4) Level of detail	(C1:C4) Richness	
	(C1:C4) Usability	(C1:C4) Context awareness	(C1:C4) Domain coverage	
	(C1:C4) Coverage	(C3:C4) Adaptability and variability	(C1:C3:C4) User centric effects	
	(C1:C4) Information diversity	(C2:C3:C4) Adaptive to learning styles	(C2:C3:C4) Accuracy of identifying students' cognitive styles	
	(C1:C3:C4) System performance	(C2:C3:C4) Adaption to learning styles	(C2:C4:C5) Learners' engagement patterns	
	(C2:C3:C4) System impact on students learning engagement	(C2:C4:C5) Level change timing		

As an example of how to read the table, the variable of Time is used to describe student performance (C2) whereas, self-assessment is a variable being described by students.

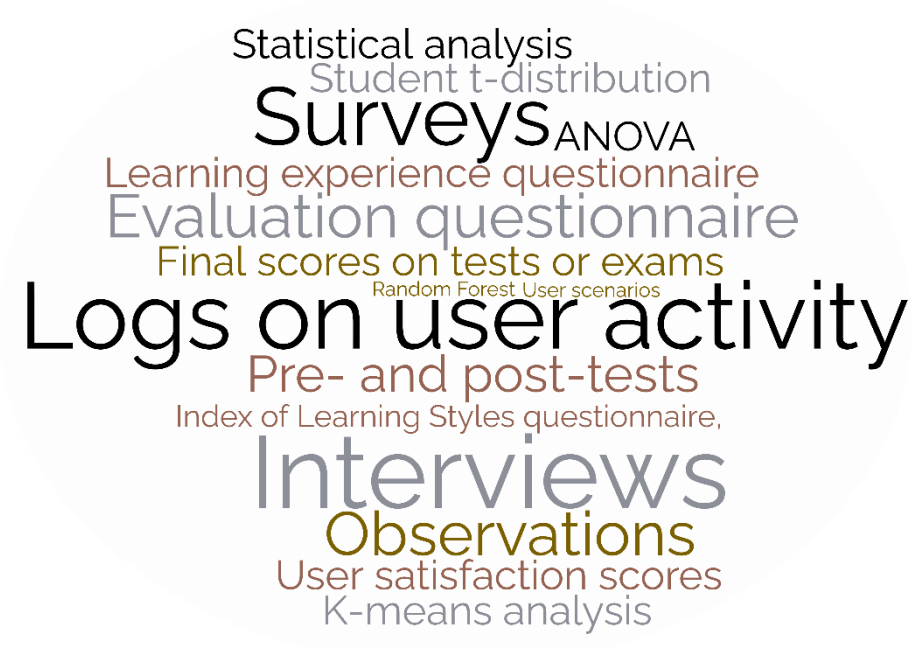

Figure 4: Word Cloud on methods from table 1 showing methods (n=235) with a frequency of <1

3.1 C1) Evaluation of LA and LAD Design and Framework

In reviewing the included studies, the category of Evaluation of LA and LAD design and Framework emerged. This category is divided into the themes of (1) Users' perception of framework and design, (2) System evaluation as a feature of the ALP, and (3) Requirement and comparative analysis of framework and design. Six papers focus

on evaluating the framework and design of LA and LAD on ALPs, whereas four papers apply the evaluation as a secondary contribution of their paper. These four will be assessed at the end of the C1) category.

3.1.1 Users' perception on framework and design

The first theme of Users' Perception on Framework and Design consists of four papers where users' perception was primarily used for evaluating the LA and LAD design and framework in ALPs. Abech et al. (2016) proposed EduAdapt, an architectural model for the adaptation of learning objects which considered device characteristics, learning style, and students' contextual information in its content recommendation to users. In evaluating the architectural model, scenarios, use cases, prototypes, a learning style survey, and a self-developed user experience survey were used. The study developed a mobile application prototype and applied it in an undergraduate course called Ubiquitous and Mobile Computing with learners (n=20) from the Computer Science area for one month of teaching. The main scientific contribution was proposing a learning object adaptation model employing inferences, rules, and learning styles in a varied context ontology. Lau, Lee, and Singh (2015) developed a recommender system for personalizing system recommendations using students' annotated metadata through a schema. The system was evaluated in two stages by students (n=92) enrolled in the course Introduction to Personal Finance, with 42% completing the evaluation. The first stage evaluated the ontology's quality using a questionnaire on 12 competency statements. The second stage used a novel seven-pillar evaluation framework, which included user-perceived accuracy, novelty, domain coverage, confidence, information adequacy, usefulness, and usability. The study's contribution was a framework that enabled learners to use their peers' opinions to locate relevant and quality resources. Tahmasebi, Ghazvini, and Esmaeili (2018) developed a feature-based educational recommender ranker system that interacts with users based on their learning styles. The system was, among other methods, assessed through a user questionnaire to evaluate the general perception of the proposals from the recommender system. The empirical study was done on science and engineering students (n=77) through a questionnaire during two semesters. The findings showed that the proposed method outperforms the general search algorithm and that the tool could be used for other systems. Fasihuddin, Skinner, and Athauda (2017) also used user perception to identify patterns and threshold levels that lead to optimal precision in detecting learning styles by tracking learners' behaviors in open environments. User perception was applied in their framework through the Indexed Learning Styles questionnaire (ILS). The empirical study was done on a prototype Cloud Adaptive Learning Courses (CALC) of an open learning environment developed and piloted on undergraduate students (n=83) taking an undergraduate IT course. It was shown that threshold values derived from literature and customized to suit open learning environments provided high accuracy in identifying learning styles.

3.1.2 System evaluation as a feature of the ALP

In the theme System Evaluation as a Feature of the ALP, two studies where the evaluation of the framework and design of the ALP is a feature of the ALP were found. Dounas, Salinesi, and Beqqali (2019) presented a framework to evaluate the INSPIREus adaptive educational hypermedia environment. The study analyzed the system's suitability for different learner types and examined if it fulfilled four requirements. The system was monitored using a novel tool, RMAS, and the collected data was checked against automatically derived constraints from the requirements. The study involved informatics and telecommunications students (n=21) over a three-month course. The results showed INSPIREus was most accurate for theorist learners and least accurate for reflector learners. It fulfilled two requirements but struggled with communication and control sharing. The system also needed more flexibility in addressing individual student needs. Fasihuddin, Skinner, and Athauda (2017) evaluated the precision of the threshold levels for identifying students' learning styles in a literature-based method. The precision was computed as the similarity of the system-identified learning style to the learning style determined by the ILS survey responded to by the students.

3.1.3 Requirement and comparative analysis of framework and design evaluations

In this section, the theme of Requirement and Comparative Analysis of Framework and Design Evaluations is presented constructed from two studies, where Abech et al. (2016) set up a comparative analysis of an ontology specified for adaptive learning systems called OntoAdapt against other ontologies. The comparative analysis was done through scenarios, analyzing the quality and fidelity of OntoAdapt. The comparative analysis was done with evaluation metrics from the Full Ontology Evaluation (FOEval) method, some additional variables provided by the software Protége and complemented with the Manchester— Web Ontology Language (OWL) tool Ontology Metric. In the study's second phase, the ontology was assessed on the platform EduAdapt to investigate if the use of ontology matches the learning objects adaptation scope. Santos and Boticario (2015)

 ©The Authors

proposed the TORMES methodology for eliciting recommendation opportunities in a recommender system using six steps: (1) Context of use, (2) User requirements, (3) Modeling of the design solution, (4) Publication of the design solution, (5) Usage to gather evaluation data, and (6) Feedback from evaluating design requirement. The study demonstrated the methodology's use in two contexts, namely the Discovering the Platform course in dotLRN LMS and the EBIFE course in the Willow free-text adaptive computer-assisted assessment system, each with two iterations. Results revealed that TORMES can detect problematic affective situations and react efficiently, but the system was intrusive in gathering affective data, causing discomfort among participants.

In addition to the three presented themes of (1) Users' perception of framework and design, (2) System evaluation as a feature of the ALP, and (3) Requirement and comparative analysis of framework and design evaluations, four papers applied the evaluation of LA and LAD design and framework as a secondary focus. Mavroudi et al. (2016) presented a framework to frame the user requirements of an adaptive system. Beckmann, Bertel, and Zander (2015) presented a framework to adapt learning material. Rincón-Flores et al. (2019) developed an algorithm from three different forecast models, and this algorithm was based on unstructured data implemented in an adaptive learning system. Lastly, Sharma, Papamitsiou, and Giannakos (2019) developed a prediction algorithm for an adaptive learning system applied in an experimental design.

3.2 C2) Evaluation of Users Performance with LA and LAD

This category emerges from the frequent use of user performance in LA is applied as a criterion for evaluating ALPs. We further thematize the category nuancing how user performance is applied. The themes are (1) sorting LA content based on user performance and (2) the effect of LA features on user performance. Ten papers focus on evaluating user performance with LA and LAD on ALPs. Eight papers also evaluate the category but as a secondary contribution of their paper.

Table 2: Review of Results

Author	Category	Evaluated unit	Methods	Variables	Purpose
Di Mascio et al. (2013)	C3, C4	Adaptive learning system TERENCE	Heuristic evaluation, expert reviewing, cognitive walk-through, observations, think-aloud and verbal protocols, controlled experiments, simulation and system performance indicators.	Users' attitudes towards the system, users' performance and system performance.	The qualitative methods such as Heuristic, expert reviewing, and cognitive walk-through evaluations are used to evaluate design choices, while simulations and system performance indicators are used to evaluate usability.
Bresó, et al. (2016)	C3, C4	A mechanism that adapts to stamina/mood	Surveys and simulations	Adaptability and variability.	The simulation evaluated system outputs, while surveys and a pilot case assessed perceived variability and adaptability levels.
Tlili et al. (2023)	C3, C4	Method for modelling to learners' personalities	Survey and LA student personality scores.	Learner personality	LA estimated learners' personalities and surveys validated personality models

Author	Category	Evaluated unit	Methods	Variables	Purpose
Hsu and Li (2015)	C2, C3, C4	Adaptive learning algorithm	Surveys, pre- and post-tests, performance scores, and user satisfaction scores.	Learner satisfaction scores and learning effectiveness	How well did the algorithm perform in terms of student performance and learning satisfaction with LAD?
Nye, et al. (2021)	C4	MentorPal, adaptive framework for virtual mentors	Formative user testing interviews, log data, pre- and post-surveys for career attitudes, and post-survey for usability.	Feasibility for virtual mentoring	The LAD's statistical evaluation was verified against users' subjective quality assurance testing.
Abech et al. (2016)	C1, C4	Ontology model for LA and LAD	FOEval, user feedback, surveys, measurements of survey reliability, user scenarios, competence questions and usage patterns.	Learners, learning objects, devices, context, context awareness, coverage, richness, and level of detail.	The ontology undergoes two phases of evaluation: the first for development and the second for comparison with other ontologies and performance in a learning context.
Mavroudi, et al. (2016)	C1, C4	Teacher-led design on the envisioned adaptive system	Evaluation questionnaire and Qualitative Comparative Analysis	Teacher perceptions and expectations	A methodology to frame requirements for critical success factors to meet user expectations of the system.
Khawaja, Chen, and Marcus (2014)	C3, C4	Adaptive multitouch tabletop interaction application.	Subjective ratings, Linguistic Inquiry and Word Count, and Advanced Text Analyzer.	User's experienced cognitive load, Language Complexity Measures, and Linguistic Category Features.	A none-intrusive, non-manipulative adaptive learning method that adjusts to users' cognitive load.
Santos, et al. (2016)	C2, C4, C5	Ambient Intelligence Context-aware Affective Recommender Platform	Tutor Oriented Recommendations Modelling for Educational Systems methodology, user-centred design methods, data mining techniques, interviews, and SUS questionnaire.	Learners' affective state, educators' tacit experiences, learner physiological signals	Exploring ambient intelligence's sensory feedback and its impact on personalized support through a recommender system.
Zhang et al. (2023)	C2, C4	Student-centred online one-to-one tutoring system	Pre- and post test of students' academic performance and system log files.	Students' learning performance (academic), teachers' performance (attracting students)	Evaluation of the system's practical value.
Kim, Jo, and Park (2016)	C2, C4	LAD for students in a virtual learning	Logs on students' frequency of use, survey on satisfaction,	Students' performance, LAD use, and satisfaction with LAD use.	Evaluating LAD use to improve

 ©The Authors

Author	Category	Evaluated unit	Methods	Variables	Purpose
		environment called Cyber Campus	and final scores on tests/exams. Analysis of the relationship among three variables.		student performance.
Lau, Lee, and Singh (2015)	C1, C4	Recommender system on user annotated metadata.	Questionnaires to experts, users, and logs on user activity	Accuracy, novelty and serendipity, domain coverage and information diversity, confidence and trust, information adequacy, usefulness and effectiveness, and usability.	Evaluating and enhancing ontology quality based on expert suggestions.
Hooshyar et al. (2018)	C2, C4	Solution-based Intelligent Tutoring System for flowchart development to improve students' problem-solving skills.	A questionnaire assessing students' learning attitude, learning interest and technology acceptance and a pre- and post-test.	Learning interest, learning attitude, technology acceptance, and problem-solving activities.	Assessing the improvement of students' problem-solving skills using the system.
Troussas et al. (2020)	C2, C4	An intelligent tutoring application over Facebook called i-LearnC# is used for learning C# programming.	Population characteristics survey, CIAO framework survey, teacher interviews, statistical hypothesis test, and system's log files.	Students' attitude on system, teachers' perspective on system in praxis, Learning outcomes, and homogeneity in grouping through peer collaboration features in system.	Evaluating social network's potential to support learners in universities and technological institutes.
Lo et al. (2012)	C2, C3, C4	An adaptive web-based learning system focusing on students' cognitive styles.	Cognitive style questionnaire based on Myers-Briggs and calculations with log files.	Accuracy of identifying students' cognitive styles and impact of the proposed adaptive web-based system on students' engagement in learning.	Unobtrusively identify students' cognitive styles through a multi-layer feedforward neural network compared to self-reported cognitive styles to provide adaptive content to students.
Dounas, Salinesi, and Beqqali (2019)	C1, C3	INSPIREus an adaptive educational hypermedia environment that provides personalized content and adaptive navigation support for each learner	Runtime Monitoring for Adaptive Systems (RMAS) tool	Accuracy of learning material estimation/recommendations meeting requirements of features used for the learning process.	To determine whether INSPIREus meets its own presented requirements.
Beckmann, Bertel, and Zander (2015)	C1, C2, C4	An adaptive framework on Moodle	Revised Verbaliser-Visualiser Questionnaire, Index of Learning Styles questionnaire, and performance scores from the test.	Learner satisfaction, motivation, Study time, Individual cognitive styles, Previous relevant knowledge, Performance	To assess whether adapting eLearning material to inter-individual differences in learning styles can improve learning

Author	Category	Evaluated unit	Methods	Variables	Purpose
					effectiveness and efficiency, learner motivation, and satisfaction.
Latham, et al. (2012)	C2, C3, C4	Adaptive online conversational intelligent tutoring system	The Index of Learning Styles Survey, Pre-post-tests, Self-developed user evaluation questionnaire, Log files.	Adaption to learning styles, Performance scores.	Assessing whether Oscar adapts to learning styles for higher student performance.
Effenberger and Pelánek (2019)	C2	RoboMission is an adaptive learning game.	Statistical analysis, Log files, and Diagnostic visualizations	Performance scores	Present discrete performance levels with universal interpretations rather than a binary failure success outcome.
Latham, Crockett, and McLean (2014)	C2, C3, C4	A conversation intelligent tutoring system called Oscar.	Index of Learning Styles, Performance score, Log files, pre- and post-test.	Adaptive to learning styles and improvement of learning scores.	Improvement of learning through an automated online conversational tutorial by presenting tutor material adapted to a student's learning style.
Dirin, Laine, and Nieminen (2017)	C4	Adaptive Mobile learning application	Emotional engagement analysis method, Web questionnaire, Diary on daily activities, Semi-structured interviews, Scenarios, Paper-based prototyping, Usability evaluation, and Observations	Delightfulness, Reliability, Adjustability, Satisfaction.	Defining a conceptual model of digital service sustainability and its measurement indicators and criteria.
Tahmasebi, Ghazvini and Esmaeili (2018)	C1, C3, C4	Feature-based educational recommender ranker system	Cronbach's Alpha Structural Equation Modelling (SEM) Fornelland Larcker's measure of average variance extracted (AVE) User perception questionnaire on the general perception of the proposals from the recommender system Index learning styles questionnaire Web crawler on meta-data of sample Open Courseware Consortium's web pages Pre- and post-test Log files.	Learning styles dimensions, System performance, learning effectiveness, and User-Centric effects.	Developing a model for a feature-based educational recommender ranker system.
Fasihuddin, Skinner and Athauda (2017)	C1, C4	CALC - open learning environments-threshold for	ILS survey Log files (Automated calculation/pattern tracking)	Learning Styles, User interactions, User behaviour, and threshold levels.	To determine whether the literature-based method

 ©The Authors

Author	Category	Evaluated unit	Methods	Variables	Purpose
		estimate Learning Styles			can achieve a satisfactory level of precision in identifying learning styles in open learning environments.
Yuksel et al., (2016)	C2, C4, C5	BACh is an adaptive brain-computer system that teaches piano.	LIBSVM, Bitwig, fNIRS data, Perception of performance, questionnaire, Interview on level change	Playing accuracy, Playing speed, Level change timings, Perception on performance.	To investigate whether BACh can dynamically adapt to increasing difficulty levels in a musical learning task based on pianists' cognitive workload.
Katuk, Kim, and Ryu, (2013)	C2, C4	A dynamic content sequencing system (DCSS)	Activity-followed-by-survey method, pre- and post-tests, learning experience questionnaire	Learning performance, Learning experience	Examining the learning experience in conjunction with learning performance to assess the adaptive learning system.
Lynch and Ghergulescu (2017)	C2, C4	Adaptemy system	Log files, Predication algorithm developed on the item Response Theory, and Questionnaire on Learning Experience.	Learning flow, Learning performance, Perceived learning experience.	Examining the learning experience and learning performance to assess the adaptive learning system.
Fadljević et al. (2020)	C2	Adaptive Moodle learning system on medical content.	K-means ANOVA, analysis, clustering, TukeyHSD, One-Way Correlation	Reading Performance assessment, Time, Competence, Score, Self-Completion	Assessing temporal behaviour as a predictor of performance on the system and whether students were fast or slow because text difficulty was unsuitable.
Papamitsiou et al. (2020)	C2, C4, C5	Multimodal self-assessment adaptive learning system	Fuzzy set qualitative comparative analysis, Pre-test goal expectations survey, Multimodal (eye-tracking, wristband, cameras, and EEG cap)	Learners' engagement patterns, Response time, Arousal, Cognitive load, Emotions, Attention, Goal-orientation	Measuring learner engagement in a multimodal learning setting.
Rincón-Flores et al. (2019)	C1, C2	Forecasting algorithm	Test and control group setup, K-nearest Neighbour, Random Forest, Logs, Grades, Photographs, Semi-structured interviews, Student t-distribution.	Prediction of performance	Using three forecast models, AI can predict student performance on unstructured data

Author	Category	Evaluated unit	Methods	Variables	Purpose
Sharma, Papamitsiou, and Giannakos (2019)	C1, C5	Developing a multimodal forecasting algorithm	Support Vector Machine algorithm, Decision trees, Gaussian process regression, Machine Learning, Principal Component Analysis, and Random Forrest.	Effort (Response time effort), Performance, Arousal	Evaluating which combination of physiological data from students explains effortful engagement and learning performance in an adaptive learning system.
Santos and Boticario (2015)	C1	A design for developing adaptive learning management systems	Brainstorming, Focus groups, Wizard of Oz, Observational study, Questionnaires, Data logs, Interviews, and Problem scenarios	User-centred design principles, E-learning life cycles, Layered evaluation of adaption features	Developing adaptive learning platforms according to ISO and other design protocols.
Al-Shanfari et al. (2020)	C2, C4	OLMlet	Bias score calculation, ANNOVA, Non-parametric tests, independent t-test, Mann-Whitney U tests, Bonferroni correction, Semi-structured interviews.	Confidence, Performance	Displaying performance and confidence levels to students to improve their performance.

3.2.1 Sorting LA content on user performance

In this section, the Sorting LA content on User Performance theme is presented which is constructed from four studies. Hsu and Li (2015) developed a new algorithm called the competency-based guided-learning algorithm (CBGLA). CBGLA-based learning system guided learners in achieving the learning objectives through personalized learning paths on the student's performance on the platform. A pilot study of the system was tested on third-year college students of electrical engineering (n=6) before an experiment on the same type of students (n=59) was conducted. The findings showed that the CBGL system supported students' learning. Yuksel et al. (2016) presented a study on an adaptive brain-computer system (BACh). The system increased the difficulty of a musical learning course aimed at the piano when cognitive workload levels became low. In a within-subject test design, study participants(n=6) undertook a training task playing 15 easy and 15 complex pieces on the piano. Pieces were learned through a typical approach (control), and two were learned afterward through BACh. Participants then played the four pieces, and performance data was used to assess the system. Results showed that learning with BACh increased accuracy and speed compared to the control setup. Fadljević et al. (2020) presented an e-learning system to support students' acquisition of health literacy with content developed in collaboration between clinical psychologists, pedagogues, and medical students. The Moodle-based system adapted text difficulty depending on students' reading competence, performance score, and self-assessment of students. From the LA, students were grouped into four competence levels. The study's participants were students [n=196] from 6th to 8th grade and worked using the system for four weeks. The results showed that at each difficulty level, students could be separated into a class of slow and a class of fast students. The text difficulty was for no students deemed unsuitable. Rincón-Flores et al. (2019) presented a predictive algorithm applied in an adaptive learning environment where instructors (n=3) taught a Physics II course. The algorithm was trained on photographs, grades, and log data from instructors' previous course students (n=182). The results showed that the algorithm provided a good forecast of the performance of each group. Al-Shanfari et al. (2020) presented an open learner modeling system (OLMlet) that used student question answers to provide adaptive feedback to students based on the correctness of their answers. The mixed method study was done with undergraduate engineering students (n=32) in an introduction to Java programming course. Students were split into a control and a test group. The control group was presented with only the skill meter (performance score). The test group was presented with the alignment between the system's evaluation of students' performance and confidence levels. The findings showed that low-achieving students benefited highly from being presented

 ©The Authors

with a visual alignment model. The visual model of alignment was associated with positive changes in their performance.

3.2.2 Effect of LA features on user performance

Five papers constitute the theme of Effect of LA Features on User Performance. Zhang et al. (2023) presented the Student-Centered Online One-to-one Tutoring system (SCOOT), where students could ask questions outside school to expand the flexibility of posing questions. The study sought to evaluate the efficiency of SCOOT and examine how students' prior knowledge and simple patterns of tutoring sessions affected student learning. The evaluation included integrating students' learning performance and behavior log files instead of conducting between-subject experiments. The empirical study comprises 40 tutoring sessions randomly selected over 50 days with a pre-test and a post-test. The study participants (n=810) were selected from Grade 7 mathematics. The results showed that the flexibility element of SCOOT was necessary and that SCOOT further increased performance differences between high- and low-achieving students. In Kim, Jo, and Park (2016), a LAD for students in the virtual learning environment Cyber Campus was implemented, which distributed video lectures and quizzes and enabled students to submit assignments. Through an experimental design on college students (n=151), the LAD was assessed on whether it would lead to higher performance. The findings showed a significant difference between the treatment and control groups. Students who had access to the LAD performed better, but an exciting finding showed that few uses of the LAD led to higher satisfaction compared to more frequent uses. Furthermore, learners who used the LAD frequently and performed well were less satisfied with the LAD. Latham et al. (2012) applied Oscar, an adaptive online conversational intelligent tutoring system that delivered a personalized natural language tutorial by predicting and adapting to students' learning styles. Participants in the study were undergraduates of science and engineering (n=70) who previously worked with the content and were to revise the topics with Oscar. Results showed a significant difference in the group with mismatched learning styles; students performed better when presented with materials matching their learning styles. In a subsequent study, Latham, Crockett, and McLean (2014) Oscar was assessed on whether it supported students' discussions in constructing knowledge. Students' learning styles were defined through Oscar's adaptation algorithms, and the validity of the system's categorization of students' learning styles was tested on student perception. The empirical study involved undergraduates (n=62) in science and engineering in assessing whether they improved their performance using Oscar. Findings showed a significant difference between match/mismatch groups. Effenberger and Pelánek (2019) presented RoboMission, which was an adaptive learning system for introductory programming. RoboMission took the form of a programming game. The case study examined the task sessions from students (n=3.800), illustrating that designing performance measures is nontrivial but possible. Three papers (Hsu and Li, 2015; Yuksel *et al.*, 2016; Al-Shanfari *et al.*, 2020) presented in C2) (1) also evaluate the effect of LA as described in C2) (1).

In addition, Santos et al. (2016) evaluated the effectiveness of supporting the learning process by, e.g., giving affective and sensory input to help calm users in a stressful learning context and whether the input was helpful in the students' performance. Hooshyar *et al.* (2018) evaluated students' improvement of learning achievement on their Solution-based intelligent tutoring system through pre- and post-tests, showing a positive impact on the learning achievements of the experimental group. Troussas *et al.* (2020) evaluated students' learning outcomes through a t-test on log data from the adaptive system. Beckmann, Bertel, and Zander (2015) used performance statistics to assess a correlation between learning styles, learning content format, and other statistics. Lo, Chan, and Yeh (2012) evaluated the students' engagement with learning on the platform, and Katuk, Kim, and Ryu (2013) assessed the students' learning performance by applying a pre and post-test setup. Lynch and Ghergulescu (2017) assessed students' performance and performance improvements with log files and developed prediction algorithms based on the Item Response Theory. Finally, Papamitsiou et al. (2020) used students' performance to compare and evaluate their self-assessment of their preparation and their engagement.

3.3 C3) Evaluation of Adaptivity

This category surfaces from the different studies that Evaluate the Adaptivity of ALPs. Evaluation of adaptivity in the ALPs is divided into three themes: (1) adaptivity based on psychological inclinations, (2) adaptivity based on users' affective capabilities, and (3) adaptivity based on users' cognitive capabilities. Five papers evaluate the adaptivity of LA and LAD on ALPs. Six additional papers mention adaptivity in their studies but do not present it as their main focus.

3.3.1 Adaptivity based on users' psychological inclinations

Only one study evaluated adaptivity on users' personalities. Tlili *et al.* (2023) assessed students (n=50) in exploratory research on building an evidence-based model to map users' personalities on Big Five Inventory dimensions to adapt learning content through the iMoodle LMS. The findings showed that the Bayesian network makes it possible to model learners' personalities compared to BFI for the three personality dimensions of extraversion, openness, and neuroticism.

3.3.2 Adaptivity on users' affective capabilities

Bresó *et al.* (2016) evaluated adaptability and variability of content from simulations and user feedback in the Personal Health System, a part of Help4Mood. The personal health system was developed to support users, so they do not relapse into depression. The system adapted its content to users' stamina or mood. Simulations (n=20.000) were done on 19 tasks and 31 subtasks; one or more subtasks could form a task. The paper concluded that the framework provided adaptive and varied sessions, improving users' use experience. Santos *et al.* (2016) presented an Ambient Intelligence Context-aware Affective Recommender Platform (AICARP) that applied Tutor Oriented recommendations Modeling for Educational Systems (TORMES) elicitation methodology to sense changes in learners' affective state from sensory communication channels. In the exploratory empirical case study, participants (n=6) completed tasks in a Wizard of Oz setting with a psycho-educational expert as the wizard. Findings showed that recommendations from an intelligent ambient system could tackle affective issues during the second language learning process.

3.3.3 Adaptivity on users' cognitive capabilities

Khawaja, Chen, and Marcus (2014) presented a model for improving performance in complex and time–critical situations by dynamically deploying more appropriate output strategies to reduce cognitive load on linguistic behavioral features. The study examined a session where participants (n=44) managed firefighting tasks as a team. The findings showed that an interaction system could apply speech and linguistic patterns to determine cognitive load and adapt system responses minimizing users' cognitive load to maintain performance. Lo, Chan, and Yeh (2012) presented an adaptive web-based learning system that adapted learning material to students' cognitive styles. In an initial study, cognitive styles were identifiable from students' (n=162) browsing behavior in the adaptive learning system. Evaluating the impact of the adaptive web-based learning system on students' engagement, another study was set up on college students (n=170) from Computer Science and Informatics. The results showed that the adaptive learning system significantly impacted temporal elements' effect on students' learning engagement. The study demonstrated that the adaptive web-based learning system based on students' cognitive styles could effectively enhance students' engagement in learning for Interpersonal and Mastery styles. In addition to the five papers synthesized above, six papers evaluated adaptivity. Di Mascio et al. (2013) presented usability associated with adaptivity and Hsu and Li (2015) assessed the effectiveness of adaptivity. Latham *et al.* (2012) and Latham, Crockett, and McLean (2014) used learning styles to adapt material delivered to students and used their performance scores to evaluate the adaptivity of their method. Tahmasebi, Ghazvini and Esmaeili (2018) used users' perception of the proposed learning material fit users learning styles as an evaluation of the adaptivity. Papamitsiou et al. (2020) used students' self-assessments to evaluate performance from multimodal data.

3.4 C4) Evaluation of ALPs Through Perceived Value

The evaluation of ALPs through perceived value is divided into the following themes: (1) users' perspectives on usability and (2) self-efficacy elements and learning styles. Eight papers focus on evaluating LA or ALPs from users' perceived value or evaluation of users' perceived value. All the papers analyzed in this review used some element of users' perspective, but most with it as a secondary focus.

3.4.1 Users' perspectives on usability

Di Mascio et al.(2013) developed the TERENCE system's Graphical User Interface (GUI) prototypes for supporting poor comprehenders and their educators. Three groups participated in the development: primary-school students (n=170), educators (n=10), and experts (n=10) such as psychologists and linguists. Usability evaluations, including expert-based evaluation, observations, think-aloud, and verbal protocol of experts and users, were conducted to identify users' requirements and context of use. Findings highlighted the importance of considering the timing and focus of users' participation and system performance during the execution of users' tasks in usability testing. Nye et al. (2021) evaluated the usability of MentorPal, a virtual mentor system that gave career advice to high school students (n=31) attending STEM internships, with the Unified Theory of

 ©The Authors

Acceptance and Use of Technology constructs (UTAUT) survey and a survey generated from variants of the CAPA system. MentorPal's career advice focused on STEM careers in the Navy, and the researchers observed the students' usage unobtrusively. Although the study had limited sample size, diversity, and impact, the students found the MentorPal experience compelling and valuable. The study's findings suggested the need to improve the diversity representation and coverage of students' main career interests by mentors. Troussas et al. (2020) studied i-LearnC, an intelligent tutoring application for learning C programming. The application overlaid Facebook, and as students made mistakes, a virtual coach (ViC) provided suitable learning material to correct misconceptions. ViC advised the pace of instruction based on students' profiles built on current and previous knowledge levels, collaboration preferences, and types of misconceptions. In an experimental design, second-year students (n=400) in an undergraduate course on Object-Oriented Development of Applications used i-LearnC, and assessment was done through a survey on the CIAO! framework, which evaluated students' use of technology-based teaching and learning. Results showed students had a positive attitude toward using Facebook for educational purposes, appreciated the communication and collaboration features, and found them helpful. Peer recommendation for collaboration showed a significant difference in acceptance of recommendations provided by i-LearnC compared to the conventional system. Dirin, Laine, and Nieminen (2017) presented an adaptive mobile learning application that provided theory and assessment for driving school students based on their learning competence and progress. The empirical study was done on participants with driver's licenses (n=7) and instructors (n=5) to identify user needs and requirements for the target application to develop a paper-based prototype. The prototype was tested in a usability laboratory on the user experience factors of delightfulness, adjustability, satisfaction, and reliability. The findings demonstrated that users' emotional attachment is essential for the target users.

3.4.2 Self-efficacy elements and learning styles

Hooshyar et al. (2018) presented a solution-based intelligent tutoring system (SITS) with an automatic text-to-flowchart conversion approach for engaging students in flowchart development aimed at improving students' problem-solving skills in an experimental design. Participants were university students (n=32) in an introductory programming course completing a self-developed questionnaire on students' learning attitude, -interest, and technology acceptance, assessing the ease of use and usefulness of SITS. From the questionnaire, it was seen that using SITS, students experienced an enjoyable learning context, were motivated to use it and experienced the content proposals from the system as helpful. Beckmann, Bertel, and Zander (2015) presented an adaptive learning system that targets performance, motivation, satisfaction, and previous knowledge to assess the effectiveness of a Moodle platform using an adaptive learning 'layer' to distribute content to students' learning styles. In a mixed-methods study design, students (n=53) of Computer Science and Media studies participated under laboratory conditions completing a questionnaire on learning motivation and satisfaction. Analysis with non-parametric statistical methods resulted in no significance of a good or bad fit between visual/verbal format and individual learning style on study time and learning outcomes. Also, there was a significant influence of matching learners' learning styles with learner satisfaction and motivation. Katuk, Kim, and Ryu (2013) presented IT Tutor, which is an adaptive learning system. In a one-way between-subject design study, participants (n=80) assessed the e-learning application, the tutorial session, and the learning experience using a learning experience questionnaire and other instruments. The results indicated that the lower or medium achievers gained certain benefits from the platform, while the high achievers in learning performance might suffer from boredom. Lynch and Ghergulescu (2017) assessed the Adaptemy system, which used curriculum-mapped content to provide personalized learning journeys to students. Secondary students participated in an objective study (n=7.614) assessing the adaptivity content and a subjective study (n=80) assessing the perceived learning experience. The subjective study revealed that the students felt increased confidence in solving math questions and an increase in enjoyment, confidence, and improved learning. In addition, 17 papers reported the evaluation of perceived value. Five studies (Abech *et al.*, 2016; Bresó *et al.*, 2016; Santos *et al.*, 2016; Tlili *et al.*, 2023; Zhang *et al.*, 2023) presented the evaluation of perceived value as a method for further informing the performance of LA. Two studies (Bresó *et al.*, 2016; Santos *et al.*, 2016) used the evaluation of perceived value to evaluate adaptability and variability and to assess the usability of the LA. Hsu and Li (2015) presented the perceived value of students to determine satisfaction levels of LA, Khawaja, Chen, and Marcus (2014) estimated the perceived level of cognitive load, Zhang et al. (2023) evaluated the practical value of the LA, and Abech et al. (2016) and Mavroudi et al. (2016) developed the application with input on the perceived value from students. Lo et al. (2012) used Myers-Briggs definitions to create cognitive style questionnaires to get students' insights on their cognitive styles. Latham et al. (2012) and Latham, Crockett, and McLean (2014) used a self-developed user evaluation questionnaire to understand students' perceived value of the intelligent conversational agent system Oscar. Tahmasebi, Ghazvini and Esmaeili (2018) used users' perceptions for the assessment of the adaptivity of their

system. In the assessment process, they also evaluated users' experienced value of the system. Fasihuddin, Skinner, and Athauda (2017) applied the Felder and Silverman Learning Styles (ILS) survey to compare the automated calculation of Learning Styles on behavior and interactions with students' perceived learning styles. Yuksel et al. (2016) used questionnaires to evaluate participants' perceived performance with BACh and used interviews to assess whether the level changes were done adequately. Al-Shanfari et al. (2020) used semi-structured interviews to understand student experiences and explain behavior identified in the data logs acquired from the system.

3.5 C5) Evaluation of Multimodal Methods

Only three papers pertained to category C5, eliminating the need for themes to further refine the studies. Papamitsiou et al. (2020) introduced a self-assessment adaptive learning system using multimodal data analysis in various configurations. The study involved 32 undergraduates who participated in an online adaptive self-assessment procedure. Multimodal data were collected using cameras, wristbands, eye-tracking, clickstreams, and EEG caps, measuring variables such as cognitive load, heart rate, blood volume pressure, temperature, EDA, attention, and emotions. Six configurations explained learners' high performance, while three configurations explained learners' medium/low performance based on engagement measures from the collected data. Sharma, Papamitsiou, and Giannakos (2019) studied an online adaptive self-assessment system for a Web Technologies course with 32 undergraduates. They collected EEG, eye-tracking, facial expressions, and wristband data, used feature selection algorithms, and employed Machine Learning techniques for prediction purposes. In contrast, Yuksel et al. (2016) used brain and MIDI data to evaluate piano performance in the BACh platform.

3.6 C6) Evaluation of the Pedagogical Implementation of ALP's LA and LAD

For the category of C6) the three dimensions (surface structure, implicit structure and deep structure) of signature pedagogy (Shulman, 2005) are used to assess the papers' level of pedagogy evaluation. We do not include the theme for evaluating ALP's LA surface structure or implicit structure in this category. All included studies examine learning activities and/or interactions between students, teachers, and ALP's LA and LAD as examined in the categories of C2-C5. No studies were exclusively categorized as a C1 category which follows that all studies evaluate ALP's LA and LAD on signature pedagogy's surface structure dimension. Moreover, the signature pedagogy dimension of ALP's LA and LAD's implicit structure was evaluated by no papers. This leaves the evaluation category of the pedagogical implementation of ALP's LA and LAD with only one theme: Evaluation of ALP's LA deep structure. Four papers have a secondary focus on evaluating ALP's LA deep structure.

3.6.1 Evaluation of ALP's LA deep structure

Four papers mentioned pedagogical theory as a contextual factor for their studies. None of the studies evaluated how pedagogical theory was evaluated in either LA, LAD, or frameworks. Di Mascio et al. (2013) applied expert evaluation consisting of 10 learning experts, who evaluated the TERRENCE system prototype, they included a pedagogical direction described as the pedagogical stimulation plan. The results from the user evaluation consisted of users (n=170) who assessed whether the expectations of the pedagogical stimulation plan were met. The evaluations were done through observational, think-aloud, verbal protocols, and controlled experiments. Abech et al. (2016)reviewed other works on an ontology that had a pedagogical approach. The reviewed ontologies were compared to their own ontology's adaption to learning styles, but their ontology was not assessed on any pedagogical parameters. Hsu et al. (2015) used competency-based learning to develop their Competency-Based Guided-Learning Algorithm (CBGLA). However, their study does not mention how CBGLA could or should be implemented in a pedagogical context, nor how CBGLA resulted in the development of users' competencies. Troussas et al. (2020) mentioned the Revised Bloom Taxonomy and collaborative learning theory as the fundamental theory for their research on the adaptive learning system i-LearnC. Still, no further description or evaluation of deep structure was included.

4. Conclusion

This study extends the research conducted by Tretow-Fish and Khalid (2022) through a systematic literature review, encompassing the analysis of 32 empirical papers and two reviews. A comprehensive assessment of Learning Analytics (LA) and Learning Analytics Dashboards (LAD) on Active Learning Platforms (ALPs) involves the examination of 27 studies contributing to multiple evaluation categories, with only five papers falling within a single category (Figure 3).

A notable trend in the literature review is the prevalence of variables employed for assessing and characterizing students and platforms, in contrast to the limited attention given to variables focused on evaluating and

 ©The Authors

describing teachers (see Table 1). The preferred methods for assessment predominantly include questionnaires (surveys), log files, and interviews (see Figure 4). Notably, the evaluations encompass ontologies, frameworks, methodologies, experimental designs, mathematical models, and LA statistics-components integral to Learning Analytics. However, assessments of pedagogical elements are conspicuously absent from the literature.

Few studies have addressed the pedagogical implementation of ALPs, with the majority treating it as a secondary focus rather than a primary concern. Despite various investigations into ALPs' LA and LADs, the existing literature predominantly evaluates the surface structure (i.e. operational acts – demonstrating, questioning, etc.), neglecting the deep structure (i.e. know-how of discipline – math by derivation practice, design by doing iterative design) and implicit structure (i.e. moral values – nursing for physical and mental health) in impact and method evaluation. Specifically, none of the reviewed studies, whether as a primary or secondary focus, systematically assess the deep and implicit structures in impact and method evaluation related to ALP's LA and LAD.

Our review reveals a noteworthy gap in the educational exploration of ALP's LA and LAD as tools for informing pedagogical or didactic decision-making among students and educators. The deficiency extends to the scarcity of studies employing established pedagogical methodologies, theories, and concepts—such as Flipped Learning, Inquiry-Based Learning, Simulation Laboratories, and Gamification. Consequently, the overall understanding of ALP's potential as a learning tool within an educational context remains underdeveloped, hindering the application of informed pedagogical or didactic choices informed by relevant theories.

For future research endeavors, a comprehensive examination of multimodal features (C5) is warranted, delving into the nuanced ways various modes of interaction influence the efficacy of LA and LADs in ALPs. Subsequent investigations should scrutinize the deep or implicit structures underpinning the pedagogical implementation of ALP's LA and LAD (C6). Methodologies ought to be developed for aligning robust pedagogical theories and concepts, thereby informing the development and assessment processes. Furthermore, future studies should incorporate evaluations encompassing teachers' perspectives to attain a holistic understanding of the impact of LA and LAD on ALPs implementation.

An essential question emerges: How can we enhance the quality of learning and teaching through LA when data collection and presentation lack a foundational methodology, framework, or theory? This query underscores the need to associate actions with data rather than merely presenting metrics such as learning objectives' difficulty, time spent on the platform, or active users. The proposed future investigations must intricately link pedagogy to LA and LAD of ALPs, providing essential support for teachers and students as they navigate cognitive and meta-cognitive impacts, behavioral changes, and social learning activities.

In essence, the envisioned assessments should integrate ontologies, frameworks, methodologies, experimental designs, mathematical models, and LA statistics—comprehensive components constituting the foundational elements of LA. Notably absent in the current landscape are evaluations of pedagogical elements, a critical gap that future research must address to comprehensively advance the field.

References

Abech, M. *et al.* (2016) 'A model for learning objects adaptation in light of mobile and context-aware computing', *Personal and Ubiquitous Computing*, 20(2), pp. 167–184. Available at: https://doi.org/10.1007/s00779-016-0902-3

Al-Shanfari, L. *et al.* (2020) 'Visualising alignment to support students' judgment of confidence in open learner models', *User Modeling and User-Adapted Interaction*, 30(1), pp. 159–194. Available at: https://doi.org/10.1007/s11257-019-09253-4

Beckmann, J., Bertel, S. and Zander, S. (2015) 'Performance & Emotion-A Study on Adaptive E-Learning Based on Visual/Verbal Learning Styles', in *International Association for Development of the Information Society*. *International Conference on e-Learning*, Las Palmas de Gran Canaria, Spain: International Association for the Development of the Information Society, pp. 27–34. Available at: https://eric.ed.gov/?id=ED562495 (Accessed: 9 November 2023).

Bresó, A. *et al.* (2016) 'A novel approach to improve the planning of adaptive and interactive sessions for the treatment of Major Depression', *International Journal of Human-Computer Studies*, 87, pp. 80–91. Available at: https://doi.org/10.1016/j.ijhcs.2015.11.003.

Di Mascio, T. *et al.* (2013) 'Design choices: affected by user feedback? affected by system performances?: lessons learned from the TERENCE project', in *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI*. *CHItaly '13: Biannual conference of the Italian chapter of SIGCHI*, Trento Italy: ACM, pp. 1–9. Available at: https://doi.org/10.1145/2499149.2499171.

Dirin, A., Laine, T.H. and Nieminen, M. (2017) 'Sustainable usage through emotional engagement: a user experience analysis of an adaptive driving school application', *Cognition, Technology & Work*, 19(2), pp. 303–313. Available at: https://doi.org/10.1007/s10111-017-0406-6.

Dounas, L., Salinesi, C. and Beqqali, O.E. (2019) 'Requirements Monitoring and Diagnosis for Improving Adaptive E-Learning Systems Design', *Journal of Information Technology Education: Research*, pp. 1–20.

Effenberger, T. and Pelánek, R. (2019) 'Measuring Students' Performance on Programming Tasks', in *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale*. *L@S '19: Sixth (2019) ACM Conference on Learning @ Scale*, Chicago IL USA: ACM, pp. 1–4. Available at: https://doi.org/10.1145/3330430.3333639.

Fadljević, L. *et al.* (2020) 'Slow is good: the effect of diligence on student performance in the case of an adaptive learning system for health literacy', in *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. *LAK '20: 10th International Conference on Learning Analytics and Knowledge*, Frankfurt Germany: ACM, pp. 112–117. Available at: https://doi.org/10.1145/3375462.3375502.

Fasihuddin, H., Skinner, G. and Athauda, R. (2017) 'Towards adaptive open learning environments: Evaluating the precision of identifying learning styles by tracking learners' behaviours', *Education and Information Technologies*, 22(3), pp. 807–825. Available at: https://doi.org/10.1007/s10639-015-9458-5.

Hewitt-Taylor, J. (2001) 'Use of constant comparative analysis in qualitative research', *Nursing Standard*, 15(42), pp. 39–42. Available at: https://doi.org/10.7748/ns2001.07.15.42.39.c3052.

Hooshyar, D. *et al.* (2018) 'SITS: A solution-based intelligent tutoring system for students' acquisition of problem-solving skills in computer programming', *Innovations in Education and Teaching International*, 55(3), pp. 325–335. Available at: https://doi.org/10.1080/14703297.2016.1189346.

Hsu, W.-C. and Li, C.-H. (2015) 'A competency-based guided-learning algorithm applied on adaptively guiding e-learning', *Interactive Learning Environments*, 23(1), pp. 106–125. Available at: https://doi.org/10.1080/10494820.2012.745432.

Katuk, N., Kim, J. and Ryu, H. (2013) 'Experience beyond knowledge: Pragmatic e-learning systems design with learning experience', *Computers in Human Behavior*, 29(3), pp. 747–758. Available at: https://doi.org/10.1016/j.chb.2012.12.014.

Khawaja, M.A., Chen, F. and Marcus, N. (2014) 'Measuring Cognitive Load Using Linguistic Features: Implications for Usability Evaluation and Adaptive Interaction Design', *International Journal of Human–Computer Interaction*, 30(5), pp. 343–368. Available at: https://doi.org/10.1080/10447318.2013.860579.

Kim, J., Jo, I.-H. and Park, Y. (2016) 'Effects of learning analytics dashboard: analyzing the relations among dashboard utilization, satisfaction, and learning achievement', *Asia Pacific Education Review*, 17(1), pp. 13–24. Available at: https://doi.org/10.1007/s12564-015-9403-8.

Latham, A. *et al.* (2012) 'Adaptive Tutoring in an Intelligent Conversational Agent System', in N.-T. Nguyen (ed.) *Transactions on Computational Collective Intelligence VIII*. Berlin, Heidelberg: Springer (Lecture Notes in Computer Science), pp. 148–167. Available at: https://doi.org/10.1007/978-3-642-34645-3_7.

Latham, A., Crockett, K. and McLean, D. (2014) 'An adaptation algorithm for an intelligent natural language tutoring system', *Computers & Education*, 71, pp. 97–110. Available at: https://doi.org/10.1016/j.compedu.2013.09.014.

Lau, S.B.-Y., Lee, C.-S. and Singh, Y.P. (2015) 'A folksonomy-based lightweight resource annotation metadata schema for personalized hypermedia learning resource delivery', *Interactive Learning Environments*, 23(1), pp. 79–105. Available at: https://doi.org/10.1080/10494820.2012.745429.

Lo, J.-J., Chan, Y.-C. and Yeh, S.-W. (2012) 'Designing an adaptive web-based learning system based on students' cognitive styles identified online', *Computers & Education*, 58(1), pp. 209–222. Available at: https://doi.org/10.1016/j.compedu.2011.08.018.

Lynch, T. and Ghergulescu, I. (2017) 'Large Scale Evaluation of Learning Flow', in *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*. *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, pp. 62–64. Available at: https://doi.org/10.1109/ICALT.2017.98.

Martin, F., Dennen, V.P. and Bonk, C.J. (2020) 'A synthesis of systematic review research on emerging learning environments and technologies', *Educational Technology Research and Development*, 68(4), pp. 1613–1633. Available at: https://doi.org/10.1007/s11423-020-09812-2.

Mavroudi, A. *et al.* (2016) 'Teacher-led design of an adaptive learning environment', *Interactive Learning Environments*, 24(8), pp. 1996–2010. Available at: https://doi.org/10.1080/10494820.2015.1073747.

Mousavinasab, E. *et al.* (2021) 'Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods', *Interactive Learning Environments*, 29(1), pp. 142–163. Available at: https://doi.org/10.1080/10494820.2018.1558257.

Nye, B.D. *et al.* (2021) 'Feasibility and usability of MentorPal, a framework for rapid development of virtual mentors', *Journal of Research on Technology in Education*, 53(1), pp. 21–43. Available at: https://doi.org/10.1080/15391523.2020.1771640.

Page, M.J. *et al.* (2021) 'The PRISMA 2020 statement: An updated guideline for reporting systematic reviews', *International Journal of Surgery*, 88, p. 105906. Available at: https://doi.org/10.1016/j.ijsu.2021.105906.

Papamitsiou, Z. *et al.* (2020) 'Utilizing Multimodal Data Through fsQCA to Explain Engagement in Adaptive Learning', *IEEE Transactions on Learning Technologies*, 13(4), pp. 689–703. Available at: https://doi.org/10.1109/TLT.2020.3020499.

Rincón-Flores, E.G. *et al.* (2019) 'Adaptive learning based on AI with predictive algorithms', in *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality*. New York, NY, USA: Association for Computing Machinery (TEEM'19), pp. 607–612. Available at: https://doi.org/10.1145/3362789.3362869.

Santos, O.C. *et al.* (2016) 'Toward interactive context-aware affective educational recommendations in computer-assisted language learning', *New Review of Hypermedia and Multimedia*, 22(1–2), pp. 27–57. Available at: https://doi.org/10.1080/13614568.2015.1058428.

 ©The Authors

Santos, O.C. and Boticario, J.G. (2015) 'Practical guidelines for designing and evaluating educationally oriented recommendations', *Computers and Education*, 81, pp. 354–374. Available at: https://doi.org/10.1016/j.compedu.2014.10.008.

Sharma, K., Papamitsiou, Z. and Giannakos, M. (2019) 'Building pipelines for educational data using AI and multimodal analytics: A "grey-box" approach', *British Journal of Educational Technology*, 50(6), pp. 3004–3031. Available at: https://doi.org/10.1111/bjet.12854.

Shemshack, A. and Spector, J.M. (2020) 'A systematic literature review of personalized learning terms', *Smart Learning Environments*, 7(1), p. 33. Available at: https://doi.org/10.1186/s40561-020-00140-9.

Shulman, L.S. (2005) 'Signature Pedagogies in the Professions', *Daedalus*, 134(3), pp. 52–59.

Tahmasebi, M., Ghazvini, F.F. and Esmaeili, M. (2018) 'Implementation and Evaluation of a Resource-Based Learning Recommender Based on Learning Style and Web Page Features', *Journal of Web Engineering*, pp. 284–304.

Tlili, A. *et al.* (2023) 'Automatic modeling learner's personality using learning analytics approach in an intelligent Moodle learning platform', *Interactive Learning Environments*, 31(5), pp. 2529–2543. Available at: https://doi.org/10.1080/10494820.2019.1636084.

Tretow-Fish, T.A.B. and Khalid, Md.S. (2022) 'Evaluating Learning Analytics of Adaptive Learning Systems: A Work in Progress Systematic Review', in E. Brooks, J. Sjöberg, and A.K. Møller (eds) *Design, Learning, and Innovation*. Cham: Springer International Publishing (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), pp. 37–52. Available at: https://doi.org/10.1007/978-3-031-06675-7_3.

Troussas, C. *et al.* (2020) 'Intelligent and adaptive tutoring through a social network for higher education', *New Review of Hypermedia and Multimedia*, 26(3–4), pp. 138–167. Available at: https://doi.org/10.1080/13614568.2021.1908436.

Yuksel, B.F. *et al.* (2016) 'Learn Piano with BACh: An Adaptive Learning Interface that Adjusts Task Difficulty Based on Brain State', in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (CHI '16), pp. 5372–5384. Available at: https://doi.org/10.1145/2858036.2858388.

Zhang, L. *et al.* (2023) 'Evaluation of a student-centered online one-to-one tutoring system', *Interactive Learning Environments*, 31(7), pp. 4251–4269. Available at: https://doi.org/10.1080/10494820.2021.1958234.

Effects of Memrise on Vietnamese EFL Students' Vocabulary: A Case Study at a College in a Rural Area

Anh Thi Nguyen[1], Thu Thuy Nguyen[2], Thanh Thao Le[1], Hoang Yen Phuong[1], Trut Thuy Pham[3], Thi Anh Thu Huynh[1] and Huong Tra Nguyen[1]

[1]School of Foreign Languages, Can Tho University, Can Tho, Vietnam

[2]Ca Mau Medical College, Ca Mau, Vietnam

[3]Faculty of Foreign Languages, Nam Can Tho University, Can Tho, Vietnam

thaole@ctu.edu.vn (corresponding author)

Abstract: Understanding the role of English language acquisition in developing the socioeconomic status of Vietnam forms the backdrop for this study, which seeks to shed light on the potential benefits of technology-assisted vocabulary learning. Based on this context, this study employed a mixed method of quantitative and qualitative design attempting to investigate the vocabulary development of college students (n=23) who were treated by the use of Memrise (Experimental group) in combination with a traditional teaching method compared to their counterparts (n=24) who were taught by a traditional teaching method (TTM) (Control group) without the use of this application. Students' thoughts about the benefits and challenges of Memrise for their vocabulary learning after the treatment were also investigated. The study yielded noteworthy results on multiple fronts. Quantitatively, data collected from the vocabulary post-tests showed that the experimental group significantly outperformed the control group in their vocabulary scores. Qualitatively, interview data indicated that Memrise positively impacted students' motivation and interest for vocabulary learning, suggesting its effectiveness in addressing the particular problems of engagement and retention in vocabulary acquisition. However, there were logistical challenges, particularly in rural settings. Factors like unstable internet connections and inadequate mobile devices adversely affected the quality of learning experiences with Memrise. By focusing on these elements, this study highlights key pedagogical implications and underscores the value of technological interventions in language learning, offering insights into addressing infrastructure barriers and tailoring e-learning methods to diverse learning environments and needs.

Keywords: Memrise, Rural area, Traditional teaching methods, Vocabulary teaching and learning, Vietnam

1. Introduction

The significant importance of vocabulary in language acquisition has been discussed in literature (Alqahtani, 2015; Asyiah, 2017). Together with methods of teaching and learning vocabulary developed by scholars and experts, applying technology to vocabulary teaching and learning also plays a crucial role (Klímová and Berger, 2018). The Memrise application, along with apps, has appeared as a powerful tool to help learners improve their vocabulary learning (Affandi and Syafi'i, 2018; Fathi, Alipour and Saeedian, 2018; Łuczak, 2017). Fadhilawati (2016), for instance, found out that Memrise had positive effects in developing learners' vocabulary, and they, therefore, responded positively toward the use of it. Izah (2019) concluded that Memrise could enlarge students' vocabulary and, at the same time, engage them in a relaxing learning environment. From these perspectives, it can be hypothesized that integrating Memrise into language teaching, particularly vocabulary teaching, can be considered a valuable approach that may be especially suitable for Vietnamese students, and by extension, potentially beneficial for other English learners globally.

In 2008, the Vietnamese government made a significant policy intervention in language education with the introduction of a reform called "Teaching and Learning Foreign Languages in the National Formal Educational System". This reform aimed to make a shift from traditional form-based teaching methods towards more modern, interactive techniques that would better prepare Vietnamese students for global engagement (Nguyen, 2018). Also, the introduction of this policy coincided with the rise of digital technology in educational settings, allowing for the incorporation of computer-assisted language learning tools (Nguyen and Vo, 2021). Therefore, investigating the suitability of e-learning practices, such as those offered by Memrise, for Vietnamese students can provide insights into addressing the digital divide and enhancing English language proficiency.

Memrise has been widely used as a powerful tool in helping learners learn English vocabulary worldwide (Melati and Herlina, 2019). However, only a few studies on the use of Memrise have been conducted in the context of Vietnam, where English language teaching has been said to be different from other contexts due to its policies and also the influence of traditional form-based teaching methods (Nguyen and Jaspaert, 2021). This underscores the need for a tailored examination of Memrise's applicability to the Vietnamese educational milieu and its relevance for Vietnamese students specifically.

Reference this paper: Nguyen, A.T. et al., 2023. Effects of Memrise on Vietnamese EFL Students' Vocabulary: A Case Study at a College in a Rural Area, *The Electronic Journal of e-Learning, 21(5)*, pp 450-460, available online at www.ejel.org

From these perspectives, it is crucial to examine whether using Memrise affects EFL learners' vocabulary development, especially in rural areas of Vietnam where applying technology to English teaching and learning still encounters several difficulties. Moreover, students' perceptions of the use of Memrise in their vocabulary learning should be addressed. By doing so, this study aims to not only address a gap in the research but also offer practical insights for various stakeholders involved in the realm of English language education in Vietnam and beyond.

1. To what extent does Memrise enhance students' vocabulary development more than TTM?
2. How do students perceive the specific advantages and disadvantages related to vocabulary learning when using the Memrise application for their English language education?

2. Literature Review

2.1 Memrise

Memrise is known as a technological tool for language learning, which has been selected for this study due to its unique approach to vocabulary acquisition. It uses memes which are known as Memes-markers. Memes can be mnemonics, etymologies, amusing videos, photos, pictures, example sentences, or anything connecting students with their learning. These memes stimulate learners' senses, imagination, and emotions to strengthen their memory and keep the learning process interesting (Zhang, 2019). Engagement and retention are also critical factors in language learning, and Memrise addresses these directly, which underscores the rationale for its selection in this context.

Hamer (2021) stated that Memrise can be easily accessed via computers and other electronic devices. The accessibility of Memrise is a crucial factor for its selection, as it aligns with the increasing penetration of digital technology in Vietnam. Thanks to various functions provided by Memrise, it supports English learning effectively, particularly vocabulary. This application allows users to take advantage of available design courses, or it is also possible for users to create their courses with multimedia added to particular items for different purposes (Kent and Sherman, 2013). The ability to customize and create courses is particularly advantageous for Vietnamese educators and learners, as it allows for the incorporation of local cultural and linguistic nuances into the learning process. Such customization is a compelling reason for choosing Memrise as a tool for this study, and it highlights the areas worth exploring in terms of personalization and effectiveness in the Vietnamese educational context.

2.2 Features of Memrise

2.2.1 Space-Repetition

Space-Repetition assists learners in identifying their strengths and weaknesses (Arnold and McDermott, 2013). The Memrise application uses Space-Repetition to facilitate memorization and recall. Furthermore, Karanfil and Özet (2021) stated that Space-Repetition is the primary teaching feature that triggers test-potential effect. Additionally, a specific algorithm is applied for testing vocabulary and helps learners discover the knowledge gap and work on inadequate knowledge (Zhang, 2019).

2.2.2 Competition

Memrise uses the feature of Competition as a gamification tool that motivates students to gain the highest position in the leaderboard (Karanfil and Özet, 2021). The board range of Memrise provides several points that learners get in every time. Learners can find their and others' results in the board range. Learners who frequently practice with items and give correct responses will get more points on the board. If they want to move on to the next level, they have to correctly provide a new word.

2.2.3 Community

Łuczak (2017) argued that Memrise allows learners to create their courses, and members of the Memrise community can also contribute their effort to the course content. It is noted that learners can interact and learn with other members and share their ideas while using Memrise (Affandi and Syafi'i, 2018). Also, learners can freely decide on the topics or courses for their learning purposes (Melati and Herlina, 2019).

2.2.4 Exercise system

The exercise system effectively enables learners to practice and strengthen their memory capacity (Tyas and Nurdiawati, 2019). This system provides several activities, such as vocabulary learning, translation, multiple-choice and word arrangement, that help engage learners in the learning process. Moreover, the sound pattern

application in the exercise system helps learners improve their pronunciation, contributing to their language acquisition. Moreover, word translation and typing challenges are other exercises provided by Memrise that help learners practice the language (Izah, 2019).

2.2.5 Reminder

The reminder function effectively helps remind students about the time to learn or review their courses (Fadhilawati, 2016). It sends an automatic email message to remind learners about the studied material to increase opportunities for further practice (Luczak, 2017).

2.2.6 Gamification elements

Memrise has been featured as an application incorporating several gamification elements into the learning process. At first, Memrise makes learning new words like a game that encourages learners to plant a flower seed. Then, every correct answer gained by learners will assist the seed to growing into fully blooming (Izah, 2019). Another task type is to require learners to match visual context cues with either spoken or written format. If the learner answers correctly, the flower seed grows, and vice versa. The items will be grown and watered to make a certain number of points each time. Feedback is given immediately through visual or musical code. Learners, therefore, can see their results and know what to prepare for the next learning stage.

2.3 Vocabulary

In literature, there are several definitions of vocabulary proposed by scholars. Hatch and Brown (1995) defined vocabulary as a list or a set of words for a specific language the speakers use to deliver their messages to their listeners. Vocabulary is categorized into two types of active and passive (Ansarin and Khabbazi, 2021). Active vocabulary refers to the words that learners can understand and use orally, while passive vocabulary refers to the words which students know through passive recognition and students, therefore, cannot use actively (Schmitt, 2008). Other scholars divide vocabulary into two types: receptive and productive (e.g., Hatch and Brown, 1995). About receptive vocabulary, Alqahtani (2015) claimed that receptive vocabulary refers to the words that learners recognize and understand as these words are used in context, but they cannot produce these words. Learners recognize receptive vocabulary in reading texts but cannot use it in speaking and writing. Dealing with productive vocabulary, it is used constructively in speaking and writing (Alqahtani, 2015) since learners can understand and pronounce or write it correctly.

2.4 The Relationship Between Memrise and Vocabulary Teaching and Learning

Vocabulary teaching requires learners to be introduced to a new word to comprehend it (Nuralisah and Kareviati, 2020). Teachers can apply certain aspects of a word that assist learners in having more opportunities to learn a word. Memrise satisfies the requirement of vocabulary teaching procedure as its features facilitate teachers' teaching. By using it, teachers can provide a new word with its pronunciation, spelling, and part of speech. Moreover, teaching vocabulary needs to apply interesting and fun activities to develop learners' interest and help them memorize words easier and longer (Asyiah, 2017). Various tasks combining flashcards and games in Memrise can train learners' memorization capacity. Hence, it makes learning vocabulary enjoyable for students.

Yugafiati and Priscila (2019) stated that learners should focus on several aspects of vocabulary, including pronunciation and orthography, grammar and meaning to master vocabulary. Specifically, Memrise helps provide input consisting of pronunciation, spelling, part of speech, and word meaning, so learners can recognize how to accurately pronounce words and understand its spelling and part of speech. Moreover, Memrise includes unique features that assist learners in practicing a word by listening to it and/or using it in a particular context through tasks, such as fill-in-the-blanks or sentence completion. Therefore, it can be said that the app is potential for vocabulary teaching and learning.

2.5 Related Studies

Fadhilawati (2016) conducted action research in which the author used Memrise in teaching vocabulary to thirty-nine learners at a university. Three levels of vocabulary related to agriculture were prepared for teaching, and the learners were required to learn the vocabulary in the classroom and review it at home via Memrise. The results revealed that using Memrise as a vocabulary learning tool improved learners' vocabulary achievement from a mean score of 60.45 (Pretest) to 86.27 (Posttest). Moreover, learning vocabulary through Memrise was an enjoyable activity, and the learners were motivated to learn vocabulary (Abarghoui and Taki, 2018). Other factors, including satisfactory responses, task achievements, and social relationships, made the participants willing to learn and review vocabulary through the application.

 ©The Authors

Similarly, Fathi, Alipour and Saeedian (2018) investigated the effects of Memrise on vocabulary learning and self-regulatory capacity in vocabulary acquisition. Fifty-nine learners were randomly assigned to an experimental group (n=33) and a control group (n=26) for the intervention. The results showed that Memrise helped learners in the experimental group outperform their counterparts in the development of vocabulary and self-regulating capacity (Esmaeili and Shahrokhi, 2020).

Tyas and Nurdiawati (2019) conducted an experimental study to determine whether Memrise is practical for tenth-grade learners' vocabulary learning. The learners were divided into two groups, consisting of an experimental group and a control group. The results showed that it was not adequate for the learners to use Memrise to learn vocabulary, mainly due to personal and technical problems (e.g., lack of electronic devices, low internet connection). Therefore, it is argued that only when the learners are well-supported with mobile devices and sufficient internet connection, Memrise can be successfully applied as a practical learning tool.

In brief, Memrise is a valuable tool that generally enriches the vocabulary learning experience by making it more interactive, efficient, and enjoyable. Nonetheless, for optimum results, it should be integrated thoughtfully into a comprehensive language learning strategy that takes into account the learner's individual circumstances, technological accessibility, and learning objectives. While Memrise appears to foster learning by using gamification and various memory techniques, there is a gap in the research in Vietnam focusing on direct comparisons of Memrise's effectiveness in vocabulary learning against traditional methods. This also extends to the qualitative aspects, as most of the literature does not deeply explore students' perceptions about the benefits and challenges of using Memrise. This lack of specific focus on comparative effectiveness and detailed user perceptions leads us to the primary aims of this current research.

3. Methods

3.1 Research Design

This experimental study employed a pretest and posttest design to investigate the students' development in their lexicon knowledge when learning with and without Memrise. The study was structured with two settings for vocabulary learning: one group would be learning vocabulary with the assistance of the Memrise app, and the other group would not use the app. This design allowed for a direct comparison of vocabulary acquisition rates and retention between the two groups, thereby addressing both lexicon knowledge and vocabulary usage. To add depth to this study, qualitative data were collected through semi-structured interviews. These interviews would provide insightful perspectives from students who were assigned to the Memrise-using experimental group, shedding light on their subjective experiences in terms of benefits and challenges.

3.2 Participants

Participants were students studying at a Medical college in the Southwest of Vietnam, specifically majoring in Nursing. Their average age was around 19 years old, and their level of English was generally intermediate according to the CEFR. Approximately 80% of the students had previously been exposed to English language learning prior to college but had limited experience with computer-assisted language learning methods.

To recruit the participants, firstly, the researcher contacted the school to gather contact information with students there. After receiving permission, the researcher sent invitations to 148 first-year students, equivalent to six classes. The number of students in each group in order was 26 (Group 1), 24 (Group 2), 24 (Group 3), 23 (Group 4), 25 (Group 5), and 25 (Group 6). Given that these students were relatively new to the college environment, they were less likely to have been influenced by the teaching methods of their teachers. Then, the students had an orientation day to discuss the topic, its importance, students' contribution to this research, and their rights if they agreed to voluntarily participate in research.

Next, the 148 students took a 150-word screening test taken from the content of the upcoming lessons to redefine their lexicon knowledge. This step aimed to select groups of similar English competence before officially becoming participants of the study and prepare a list of vocabulary that would be taught and tested in this study. All student groups had the same test time of 180 minutes, and each question was worth 1 point. The results showed that the average scores of groups from Group 1 to Group 6 were 60, 30, 52, 29, 73, and 80, respectively. There was a significant difference in the lexicon knowledge of the groups, except for Group 2 and Group 4, when compared with each other. Therefore, these two groups (Group 2: Traditional teaching method; Group 4: Memrise) were eligible to participate in the intervention.

Two volunteer lecturers were also invited as collaborators of the study. These two lecturers had relatively similar backgrounds. Specifically, they were all lecturers with years of experience teaching English at higher education

level and had all applied Memrise into their teaching in practice. Therefore, they knew how to use this app to teach vocabulary to their students. Before the intervention was implemented, the research team and these two teachers, together, had several sessions to discuss lesson plans that they would use to teach the students. The research team designed all the lesson plans used in this current study to reduce the fact that the differences in the teachers' teaching styles would affect the research results. Therefore, they were required to stick to lesson plans. Then, two teachers taught demos under the observation of the research team to evaluate and adjust their teaching to best serve the purpose of this study.

As for the selection of interviewees, this study employed a control quota sampling technique. This technique allowed the researchers to select the participants based on predetermined characteristics so that the total sample would have the same distribution of characteristics. Specifically, 9 out of the 23 students in Group 4 were intentionally selected. Among them, 2 students scored the highest, 2 scored the lowest, 2 scored average, and 3 whose scores varied on the posttest. The rationale for involving only 9 students in the interview process was influenced by a combination of methodological rigor and practical constraints. The control quota sampling technique, applied here, ensured a representation across a spectrum of performance levels within Group 4. By selecting two students with the highest scores, two with the lowest scores, two with average scores, and three students with varying scores on the posttest, the researchers aimed to capture a broad and diverse array of experiences and insights into the use of Memrise for vocabulary acquisition.

3.3 Data Collection Instruments

3.3.1 Vocabulary test

The screening test served as a critical instrument for data collection and was designed to equitably compare the efficacy of the two treatment conditions - learning with Memrise and learning without Memrise. The researchers followed a rigorous selection process which was grounded in three key criteria: 1) the words had to belong to themes presented in the students' current textbook, ensuring educational relevance; 2) they had to be rated as A2 level according to the CEFR, making them suitable for the proficiency level of the participants; and 3) the words had to be unfamiliar to the students, creating a fresh learning experience and allowing us to better measure vocabulary acquisition. To be specific, 72 words that were answered incorrectly by students in both groups were identified. These words were written on individual pieces of paper, folded, and placed in a cardboard box. Team members then took turns drawing these pieces of paper until a total of 50 words were selected.

To enhance the reliability, initially, a pilot test was conducted on a subset of 30 students who did not participate in the main study. It is noted that the demographic and educational backgrounds of the participants involed in the pilot test were closely matched to those of the main study sample. Based on the feedback received from the pilot study, 90% of the students found the questions to be clear, while 85% believed the level of difficulty to be appropriate. This feedback was instrumental in fine-tuning the questions for the main study. Three experts also reviewed the questions for appropriateness and relevance to the study's objectives. Finally, to quantitatively confirm the internal consistency of the vocabulary tests, a Cronbach's alpha was calculated post-administration of the tests. The Cronbach's alpha value was found to be 0.87, indicating a high level of internal consistency and thereby confirming the reliability of the test items.

The posttest was designed to assess students' memorisation of these 50 words through four types of questions including: 1) word completion exercises (10 words); 2) word completion exercises (10 words); 3) multiple-choice questions (15 words); and 4) word selection tasks (15 words). This diverse range of question types was chosen to assess both receptive and productive vocabulary skills, as well as the ability to understand vocabulary in context.

To address the management of written records, it is essential to highlight that all materials, including the written records such as pretest and posttest sheets, were treated with the utmost confidentiality and in accordance with ethical research standards. All written records were anonymized to remove any personal identifiers, ensuring the privacy of the participants was maintained. These records were then stored securely, accessible only to the primary researchers involved in the study. For data analysis purposes, the written responses from the tests were digitally transcribed, with each participant being assigned a unique code to further ensure anonymity. The digital records of the transcriptions were encrypted and stored on a password-protected computer. These transcriptions were used for the quantitative analysis. All physical records were to be kept for a period defined by the study's protocol, after which they would be shredded and disposed of in a secure manner. Digital records,

 ©The Authors

similarly, would be retained for a period as stipulated by the ethical guidelines before being permanently deleted.

3.3.2 Semi-structured interviews

To offer more details on the qualitative component of the study, semi-structured interviews were scheduled shortly after the treatment course was completed, ensuring that the experience of using Memrise was still fresh in participants' minds. The interviews were designed to provide in-depth, qualitative data that would complement the quantitative results. According to Kvale and Brinkmann's (2009) methodology, the aim of these interviews was not just to collect surface-level data, but to dig deeper into participants' personal experiences and viewpoints regarding the use of Memrise for vocabulary learning. In this manner, the interviews could reveal the nuanced perspectives of learners, in their own words, about the various advantages and disadvantages of using the app. The questions for these interviews were carefully crafted based on a framework proposed by Wright and Alison (2004), which aimed to investigate two key areas of interest: 1) the perceived benefits of using Memrise for vocabulary learning; and 2) the perceived challenges, which could span issues such as technical difficulties, limitations of the app's vocabulary database, or any demotivating factors.

It is worth noting that language can be a significant factor in the quality of data collected during interviews. In this respect, all interview participants were given the option to use their mother tongue, Vietnamese, for the interviews. This choice aimed to create a more comfortable and open environment (Creswell, 2012), thereby encouraging participants to fully express their opinions, share their experiences, and contribute nuanced insights into the utility and limitations of Memrise in vocabulary learning.

All interviews were audio-recorded with the explicit consent of the participants and transcribed verbatim. These transcriptions allowed for a detailed analysis of the data, ensuring that the participants' insights and nuances were accurately captured. To maintain confidentiality, all transcripts were anonymized, with any identifying information being removed or altered. As per ethical research practices, both audio recordings and their corresponding transcriptions will be retained only until the end of the research project, after which they will be permanently deleted.

3.3.3 Data analysis

For the quantitative data pool, descriptive statistics were first run to summarize the central tendency, dispersion, and shape of the dataset's distribution. This helped identify whether any noticeable differences existed in the posttest results between the two groups, those who used Memrise and those who did not. Furthermore, an Independent Samples T-test was performed to ascertain whether the observed differences in posttest vocabulary scores were statistically significant. For this study, the alpha level for determining statistical significance was set at $p<.05$. Regarding the qualitative portion, a thematic analysis method was applied to examine the interview transcripts. This analytical approach involves identifying, analyzing, and interpreting patterns within qualitative data. By employing thematic analysis, the researchers were able to systematically address the research issues and questions. In this respect, it adds a rich and contextual layer to the quantitative results.

3.3.4 Procedures

The procedures of this study are detailed in Table 1.

Table 1: Research procedures

Duration	**Stages**	**Sub-stages**	**Experiment Group**	**Control Group**
2 weeks	Pre-intervention	Pretest	See Section 3.2	
15 weeks	While-intervention	Step 1	**Introducing the lesson aims.** The teachers set the scene to teach the lesson, including its aims and objectives. To create an environment for discussion (in pairs or in groups), the students were asked leading questions and, at the same time, received suggested clues for the answers. There were many activities that the teachers used in this step to stimulate the students' contribution of ideas.	
		Step 2.1	**Teaching vocabulary**. The instructional activity was organized into two phases. In the first phase, the teachers taught vocabulary by traditional teaching method; that was, they explicitly introduced the vocabulary to students by directly translating its meaning in L1 or showing pictures representing the meaning of the words. The students were also	

Duration	Stages	Sub-stages	Experiment Group	Control Group
			instructed on pronouncing the words correctly by tentatively listening to the teachers spelling out and repeating what they heard.	
		Step 2.2	In the second stage, the students were instructed to work with **Memrise** to review the words they had learned. To start learning with Memrise, the teacher introduced the words for learning with their provided Vietnamese meanings, sound, and parts of speech.	In the second stage, the students were instructed to work with **handouts** to review the words they had learned. To start learning with the handouts, the teacher introduced the words for learning with their provided Vietnamese meanings, sound, and parts of speech.
		Step 3	**Checking students' understanding of vocabulary**. This step aimed to give students opportunities to review the words they had learned through consolidating exercises. The students were encouraged to complete the exercises in the allotted time required by the teachers. Once the students finished the activities, they could know the degree of vocabulary they had the mastery of during the lesson.	
		Step 4	**Using the words**. This step aimed to encourage the students to apply the words they had just learned to a so-called real-life situation word use. In this step, the teachers often asked the students to practice the words in particular communicative topics, customarily taken as speaking or writing activities. Students could work individually or in pairs or groups for their performance.	
2 weeks	Post-intervention	Posttest	At the end of the treatment, a posttest was run to check the students' vocabulary development in the two groups. On the test date, the students were arranged to sit separately from each other and constantly observed and managed by two examiners to ensure that any cheating activities could be prevented. Each of the performances was then scored by two independent judges for its reliability.	
		Interviews	After the treatment, nine students were invited for the interview, which was conducted via Zoom. Each lasted for approximately 60 minutes. All interviews were audio-recorded, then double-checked by an independent judge for data analysis afterward.	**x**
"X" means to do nothing.				

4. Results and Discussions

4.1 Differences in Students' Vocabulary Development Between the Memrise Setting and the Traditional Teaching Method Setting

Table 2 describes the results of the Independent Samples T-test used to check whether there was any significant difference in the posttest results between the two groups of students after being treated by two different settings.

Table 2: Posttest results

GROUP	N	Mean	Max	Min	SD	p
Group 4	23	32.5	46	15	8.8	.02
Group 2	24	25.2	49	9	12.2	

Descriptive results show that the experimental group outperformed the control group in developing their vocabulary (M_{Group4}=32.5, M_{Group2}=25.2, respectively). The results of the Independent Samples T-test indicated a significant difference in vocabulary performance between the two groups ($p<.05$) in favor of Group 4. It is therefore concluded that integrating Memrise with a traditional teaching method was influential in developing students' vocabulary compared to that of the traditional teaching method only.

The current results not only substantiate the existing literature but also shed light on the possible mechanisms that contribute to these improvements. Firstly, the enjoyment factor cannot be underestimated. Fadhilawati (2016) indicated that the engaging nature of digital learning platforms like Memrise can positively affect learner outcomes. Unlike traditional methods, Memrise employs gamified elements that make the learning process

 ©The Authors

more entertaining and less monotonous resulting in increased engagement and motivation. Secondly, Memrise offers a multifaceted approach to vocabulary learning. It incorporates multiple modes of practice, including but not limited to vocabulary recognition, spelling, and listening exercises, and reviewing mechanisms (Fadhilawati, 2016). This comprehensive feature set might have contributed to the observed difference, as it provides a richer learning experience compared to traditional methods that often focus on one or two aspects of vocabulary learning. Furthermore, Memrise incorporates spaced repetition algorithms that encourage frequent reviewing, enabling better retention. This is consistent with the findings of Zhang (2019), who pointed out that regular review and repetition are crucial for effective vocabulary memorization. The results align well with previous studies by Fathi, Alipour and Saeedian (2018) and Esmaeili and Shahrokhi (2020), thereby adding to a growing body of evidence supporting the efficacy of Memrise in enhancing vocabulary learning. These cumulative findings suggest that Memrise is not just an alternative but could be considered a robust supplement for traditional vocabulary learning methods.

4.2 Students' Perceptions of the Benefits and Challenges Using Memrise in Vocabulary Learning

Regarding the benefits of using Memrise in vocabulary learning, the findings showed that most students (n=7 out of 9) indicated that Memrise was convenient and effective for enhancing their vocabulary range. For instance, one student reported: "I could remember the words because I learned the words many times via Memrise" (Student 1) or "I could remember the words faster and longer because when learning with Memrise, I could listen, write, and see the words many times via Memrise." (Student 2).

Not only did the students find Memrise convenient for learning new vocabulary, but they also highlighted its role in reinforcing memory retention of new words through multiple exposures. The "many times" mentioned by the majority of students could be traced back to the app's accessibility - being available at any time and place, which in turn promotes frequent use and repetition. This is particularly important because vocabulary retention is often linked to frequent and varied exposures to new words (Nation and Webb, 2011). Previous studies, such as those by Affandi and Syafi'i (2018) and Aminatun and Oktaviani (2019), have highlighted the importance of "ample input" in vocabulary learning. Memrise caters to this by providing diverse modes of input, i.e., visual, auditory, and kinesthetic that contribute to a fuller understanding of a word's nuances, including its spelling, pronunciation, and contextual usage. For example, Student 2 specifically pointed out how the multimedia features of Memrise, which allowed him to "listen, write, and see the words many times" led to faster and more enduring memory retention. This mirrors findings by Deris and Shukor (2019) that audio aids are instrumental in helping students recognize and pronounce words correctly. Additionally, the students implicitly acknowledged the app's effectiveness in improving multiple aspects of vocabulary knowledge, not just recall. This encompasses understanding the word meaning, correct spelling, and proper pronunciation, which collectively contribute to a richer vocabulary skill set.

The results additionally showed that most of the students (n=7 out of 9) strongly perceived the positive effect of Memrise in promoting their learning interest. Student 1, for instance, expressed: "Well, honestly, I like learning vocabulary with Memrise" because his/her motivation and interests increased due to "the repetition function", which allows him/her to "practice the words with Memrise until I could remember the words exactly". Sharing the same view, Student 3 reflected: "I like learning vocabulary with the Memrise application because my pronunciation was improved without looking the words up in the dictionary by practicing with the listening part." Moreover, this student found the translation function given by Memrise enjoyable because "I could rewrite the words many times if I gave the wrong answer so that I would remember the words better." In addition, the data indicated the significant impact of the leaderboard on students' motivation to learn vocabulary through Memrise, as one student reported: "I want to climb to the highest position of the leaderboard." (Student 9).

It could be seen that Memrise impacted not just vocabulary retention but also the intrinsic motivation and engagement levels of the students. Starting with the notion of "repetition" as pointed out by Student 1, it is worth noting that the effectiveness of repetition is not merely anecdotal but is grounded in cognitive psychology principles concerning the "spacing effect" which postulates that information is better retained when studied in multiple sessions over time (Rogers, 2017). The repetition function in Memrise is not just a feature but a strategically designed mechanism based on cognitive learning theories, contributing to more effective vocabulary retention. Similarly, the improvement in pronunciation, as mentioned by Student 3, leverages auditory learning strategies, which supplement visual or reading-based methods. This multi-modal approach is supported by Mayer's Cognitive Theory of Multimedia Learning, suggesting that learners can understand and retain material much better when they engage with it through multiple sensory channels. The "leaderboard"

aspect reported by Student 9 echoes findings in gamification literature that competition can serve as a strong extrinsic motivator (Hamari and Koivisto, 2014). The "gamification" of Memrise, which many students found appealing, ties into a broader educational philosophy that learning can - and should - be fun to foster engagement. The game-like features of Memrise go beyond mere entertainment. They tap into pedagogical paradigms that underscore the role of engagement and motivation in effective learning. The gaming elements offer immediate feedback, rewards, and challenges, all of which are proven motivational boosts in educational settings (Kapp, 2012).

Regarding students' challenges when using Memrise to learn vocabulary, the students (n=6 out of 9) reported that they had problems when registering and installing the app because of their lack of knowledge and skills in using applications in language learning. Student 1, for instance, shared that she "had to do many times when I first signed up for the application" mainly due to her poor knowledge of the app. In a similar vein, Student 4 found it difficult to use Memeise since "this was the first time I had downloaded any English learning apps to study".

It is possible to recognize the role that the "digital divide" and pedagogical traditions play in students' ability to navigate educational technologies like Memrise. Student 1's and Student 4's difficulties with registration and initial use can be seen not merely as individual challenges but as manifestations of a broader issue of digital literacy. This links to educational literature suggesting that the effective use of technology in learning is not just about having the necessary hardware but also the skills to use it effectively (van Deursen and van Dijk, 2015). The students' difficulties in the initial setup could potentially act as an obstacle to their long-term engagement with the app, thus impacting their vocabulary acquisition adversely. This highlights the need for scaffolding support during the initial stages of technology adoption in educational settings, something that is often overlooked. It is also important to explore the sociocultural factors affecting technological adoption. The lack of experience with edtech among the student participants is not merely an individual problem but is influenced by the pedagogical culture they come from. The teacher-centered, traditional educational systems prevalent in Vietnam may not prepare students well for the self-directed learning that platforms like Memrise require. This suggests that even if edtech solutions are pedagogically sound, their effectiveness may be limited if students are not prepared to navigate them. Moreover, the students' initial struggles with Memrise highlight a potential barrier to entry that goes beyond simple usability. The students are, essentially, facing a double cognitive load, that is learning how to use the technology and learning English vocabulary simultaneously (Sweller, 2018). Taken together, these observations suggest that while Memrise has potential benefits for vocabulary acquisition, its effectiveness can be impacted by a variety of external factors including digital literacy skills, cultural approaches to pedagogy, and the cognitive demands of using the app. These are crucial aspects to consider for teachers or stakeholders who are planning to integrate technology like Memrise into their curricula. The experiences and challenges faced by the students in this study are not isolated incidents but are reflective of larger trends and challenges in the edtech landscape, particularly in settings that have traditionally been less technologically oriented.

Another difficulty is that the slow Internet connection and students' low configuration devices affected the Memrise learning process. More than half of the students (n=5 out of 9) shared that the Internet speech function in Memrise worked badly due to poor internet connection. As for clarification, Students 1 and 3 admitted that their devices could not function well because of the "low speed of the internet" (Student 3). Regarding the devices for Memrise learning, Student 4 shared: "Sometimes the network was too bad, but my phone configuration also needed to be stronger I guess […] so that the download could have been faster".

Based on the outcomes of the above interviews, it is vital to acknowledge the intricate interplay between socio-economic status, access to technology, and effective e-learning. The issue goes beyond mere technical difficulties; it is a matter of educational equity. The challenges posed by slow internet connections and low-configuration devices can act as significant barriers to effective e-learning. This aligns with the research on the digital divide which suggests that a lack of access to high-quality digital technology can exacerbate educational inequalities (Chen, Wei and Liu, 2011). For the students in this study, the technical issues they faced could be a manifestation of broader social inequalities, limiting their ability to fully benefit from educational technologies like Memrise. Moreover, the students' challenges highlight the idea of technological efficacy which refers not just to the availability of technology but its effectiveness in aiding educational outcomes. If students are to get the most out of applications like Memrise, it is not enough to have a device and an internet connection; these need to be robust enough to support the application's multimedia elements, such as Internet speech functions. As for the poor internet connectivity, attributing it merely to poor technical services oversimplifies the issue.

 ©The Authors

Infrastructure challenges, especially in rural areas, often reflect wider systemic issues, from government policy to local funding constraints (Bélanger and Carter, 2009).

5. Conclusions

This study aimed to investigate the efficacy of the Memrise app in a unique educational setting, a college in Southwest Vietnam. Additionally, the students' perceptions of the benefits and challenges when using Memrise to learn vocabulary were addressed. Based on the results from the posttest, using Memrise helped students in the experimental group outperform those in the control group. Additionally, interview data revealed that Memrise could significantly assist students with vocabulary acquisition. The application, according to the students, was an effective and enjoyable tool for learning vocabulary. However, the learning conditions in the current study were not in favor for using technology, which could significantly affect the learners' experience when learning vocabulary with Memrise. Specifically, participants experienced initial challenges in familiarizing themselves with the Memrise platform and faced technological hindrances such as unstable internet connections and suboptimal mobile devices.

Based on the results, it is implied that language learners are encouraged to learn more about the features given by the Memrise to ensure that they can use this app successfully. Due to the significant development in the learners' vocabulary range and their positive perceptions in this current study, it is implicated that teachers should use Memrise as a supplementary practice for their students. Dealing with difficult issues found, it is suggested that necessary supports given by educational leaders in terms of learning conditions such as wireless connections are essential for students in rural areas to experience and learn most effectively by applying technology, typically using Memrise for vocabulary learning. These supports address some of the unique challenges, such as limited educational resources and access to quality instruction, that students in rural Vietnam face.

Despite rigorous methodology and effort invested in the study, it is important to acknowledge that certain limitations remain such as limited time for the treatment, small sample size of participants and incapability of measuring the potential of this app in students' vocabulary learning. In addition, the study investigated the students' vocabulary performance shortly afterward the treatment. Therefore, the study only evaluated the students' memory in constant time, not for vocabulary retention.

References

Abarghoui, M. A. and Taki, S., 2018. Measuring the effectiveness of using "Memrise" on high school students' perceptions of learning EFL. *Theory and Practice in Language Studies*, 8(12), pp 1758-1765.

Affandi, M. H. and Syafi'i, A., 2018. Memrise as a builder students vocabulary. *Efektor*, 5(1), pp 32-38.

Alqahtani, M., 2015. The importance of vocabulary in language learning and how to be taught. *International Journal of Teaching and Education*, 3(3), pp 21–34. https://doi.org/10.20472/TE.2015.3.3.002

Aminatun, D. and Oktaviani, L., 2019. Using "Memrise" to boost English for business vocabulary mastery: Students' viewpoint. *Proceeding Universitas Pamulang*, 1(1), pp 590-596.

Arnold, K. M. and McDermott, K. B., 2013. Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), pp 940-945. https://doi.org/10.1037/a0029199

Asyiah, D. N., 2017. The vocabulary teaching and vocabulary learning: perception, strategies, and influences on student's vocabulary mastery. *Journal Bahasa Lingua Scientia*, 9(2), pp 293–318. https://doi.org/10.21274/ls.2017.9.2.293-318

Bélanger, F. and Carter, L., 2009. The impact of the digital divide on e-government use. *Communications of the ACM*, 52(4), pp 132-135.

Chen, N. S., Wei, C. W. and Liu, C. C., 2011. Effects of matching teaching strategy to thinking style on learner's quality of reflection in an online learning environment. *Computers & Education*, 56(1), pp 53-64.

Creswell, J. W., 2012. *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson Education, Inc.

Deris, F. and Shukor, N., 2019. Vocabulary learning through mobile apps: A phenomenological inquiry of student acceptance and desired apps features. *iJIM*, 13(7), pp 129-140.

Duong, M. C., 2020. Vietnam: COVID-19 and after: Impact and ways forward. *East Asia Forum Quarterly*, 12(2), pp 31-33.

Esmaeili, Z. and Shahrokhi, M., 2020. The impact of Memrise application on Iranian EFL learners' collocation learning and retention. *International Journal of Language Education*, 4(2), pp 221-233.

Fadhilawati, D., 2016. Learning and reviewing vocabulary through Memrise to improve students' vocabulary achievement. *JARES (Journal of Academic Research and Sciences)*, 1(2), pp 33-46.

Fathi, J., Alipour, F. and Saeedian, A., 2018. Enhancing vocabulary learning and self-regulation via a mobile application: An investigation of the memrise app. *Journal of Modern Research in English Language Studies*, 5(1), pp 27-46.

Hamari, J. and Koivisto, J., 2014. Measuring flow in gamification: Dispositional flow scale-2. *Computers in Human Behavior*, 40, pp 133-143.
Hamer, W., 2021. Utilizing Memrise application as instructional media-based technology to enrich the students' vocabulary mastery. In *UNNES-TEFLIN National Seminar*, 4(1), pp 364-375.
Hatch, E. and Brown, C., 1995. *Vocabulary, semantics, and language education*. Cambridge: Cambridge University Press.
Hoang, D. T. N. and Nguyen, N. T. B., 2019. Mobile augmented reality activities in EFL classrooms at a Vietnamese university from the students' perspective. *Journal of Asia TEFL*, 16(1), pp 411-419.
Izah, N., 2019. Upgrading students'vocabulary through "Memrise" app. *In Proceeding of 1st Conference of English Language and Literature (CELL) Innovative Teaching of Language and Literature in Digital Era* (pp 1-10).
Kapp, K., 2012. *The gamification of learning and instruction: Game-based methods and strategies for training and education*. NewYork: John Wiley and Sons.
Karanfil, F. and Özet, M. N., 2021. How EBA (Educational Informatics Network) platform and Memrise may help EFL learners: A review for state school EFL learners in Turkey. *Journal of Foreign Language Education and Technology*, 6(1), pp 76-91.
Kent, D. and Sherman, B., 2013. Pilot study for use of Memrise application by Korean junior college students studying EFL vocabulary in a blended learning context. *Yeongsang-yeongeogyoyughaghoe [Society for Teaching English Through Media]*, 14(3), pp 169-192.
Klímová, B. and Berger, A., 2018. Evaluation of the use of mobile application in learning English vocabulary and phrases–a case study. In *Emerging Technologies for Education: Third International Symposium, SETE 2018, Held in Conjunction with ICWL 2018, Chiang Mai, Thailand, August 22–24, 2018, Revised Selected Papers 3* (pp 3-11). Springer International Publishing.
Kvale, S. and Brinkmann, S., 2009. *Interviews: Learning the craft of qualitative research interviewing*. Sage.
Łuczak, A., 2017. Using Memrise in legal English teaching. *Studies in Logic, Grammar and Rhetoric*, 49(1), pp 141–152.
Melati, M. and Herlina, H., 2019. Optimizing EIL using "Memrise" as a builder students' vocabularies. *Edu-Ling: Journal of English Education and Linguistics*, 3(1), pp 190-206.
Nation, I. S. and Webb, S. A. , 2011. *Researching and analyzing vocabulary*. Cengage Learning.
Nguyen, L. A. T. and Habók, A., 2022. Digital literacy of EFL students: An empirical study in Vietnamese universities. *Libri*, 72(1), pp 53-66.
Nguyen, T. A. and Jaspaert, K., 2021. Implementing Task-based Language Teaching in an Asian context: Is it a real possibility or a nightmare? A case study in Vietnam. *ITL-International Journal of Applied Linguistics*, 172(1), pp 121-151.
Nguyen, T. N. Q. and Vo, N. T. T., 2021. The need of applying English learning apps to help Van Lang University students improve their spoken English performance. *AsiaCALL Online Journal*, 12(2), pp 72-86.
Nguyen, T.T. H. and Pham, T.T.M., 2016. Difficulties in teaching English for specific purposes: Empirical study at Vietnam universities. *Higher Education Studies*, 6(2), pp 154-161.
Nguyen, V. T., 2018. Project 2020 and professional development for high school EFL teachers in Vietnam. In *Professional development of English language teachers in Asia* (pp 95-108). Routledge.
Nuralisah, A. S. and Kareviati, E., 2020. The effectiveness of using Memrise application in teaching vocabulary. *Professional Journal of English Education*, 3(4), pp 494-500.
Rogers, C. S., 2017. Semantic priming, not repetition priming, is to blame for false hearing. *Psychonomic bulletin & review*, 24, pp 1194-1204.
Schmitt, N., 2008. Instructed second language vocabulary learning. *Language Teaching Research*, 12, pp 329-363.
Smith, K., 2004. The evolution of vocabulary. *Journal of Theoretical Biology*, 228(1), pp 127-142.
Thao, L. T. and Mai, L. X., 2020. English language teaching reforms in Vietnam: EFL teachers' perceptions of their responses and the influential factors. *Innovation in Language Learning and Teaching*, 16(1), pp 29-40.
Tyas, E. W. and Nurdiawati, D., 2019. The effectiveness of Memrise online application on vocabulary mastery of the tenth grade BDP students of SMK Al-Furqon Bantarkawung. *Jurnal Dialektika Peogram studi Pendidikan Bahasa Inggris,* 7(2), pp 159-169.
Van Deursen, A. J. and Van Dijk, J. A., 2015. Toward a multifaceted model of internet access for understanding digital divides: An empirical investigation. *The information society*, 31(5), pp 379-391.
Walker, L., 2015. The impact of using Memrise on student perceptions of learning Latin vocabulary and on long-term memory of words. *Journal of Classics Teaching*, 16(32), pp 14-20.
Wright, A. M. and Alison, L., 2004. Questioning sequences in Canadian police interviews: Constructing and confirming the course of events?. *Psychology, Crime and Law*, 10(2), pp 137-154.
Yugafiati, R. and Priscila, T., 2019. The use of mind map to improve students' vocabulary at SMP Waringin Bandung. *Eltin Journal: Journal of English Language Teaching in Indonesia*, 7(2), pp 98-105.
Zhang, X., 2019. Learning technology review: Memrise. *Calico Journal*, 36(2), pp 152-161.

 ©The Authors