The Role of Evaluation Methods in Health-Related E-learning: A Rapid Review

Jason D. Stemp
La Trobe University, Australia

Debannita Ghosh
Australian Catholic University, Australia

Urooj R. Khan
James H. Boyd
La Trobe University, Melbourne, Australia

Abstract
Training and development programs are increasingly delivered online with numerous studies reporting no differences in learning outcomes between online and traditional learning. However, there are no established standardized methods to evaluate the effectiveness of online learning. This review aims to map the state of research around health-related education to determine what e-learning evaluation methods are being used, the strengths or deficiencies of these methods, and which are appropriate for measuring the effectiveness of online education. Databases searched were PubMed, ProQuest, Education Resources Information Centre, Cumulative Index to Nursing and Allied Health Literature, Scopus, PsychInfo, and Medline. Studies were included if they were published between 2011 and 2021, reported health-related online education and included an evaluation component. Thirty studies were obtained from numerous countries with varied methodologies and designs. Participants ranged from undergraduate students to medical professionals. Evaluation methods included student participation, students' reaction to the training program, self-efficacy, knowledge assessment, long-term performance, and the Kirkpatrick Evaluation Framework. The review identified that course evaluations, such as measuring student satisfaction scores alone, are insufficient when used to quantify learning effectiveness for online education. This was particularly important as studies are reporting these single metrics as positive effects of training interventions without justification. Suggestions within the reviewed papers were to adopt and implement an appropriate validated method within the course curriculum to evaluate learning outcomes.

Keywords: assessment tools, evaluation method, learning impact measure, outcome measure, e-learning, online education, online learning, health education, healthcare

Stemp, J.D., Ghosh, D., Khan, U. R., & Boyd, J. H. (2022). The role of evaluation methods in health-related e-learning: A rapid review. *Online Learning, 26*(4), 369-398.
DOI: 10.24059/olj.v26i4.3115

Workplace educational training and development plays a critical role in staff development and organizational efficiency, helping organizations achieve goals and objectives. The way training is designed, delivered, and implemented contributes to the success or failure of these outcomes (Salas et al., 2012). The last decade has seen the workplace training function driven by a legal requirement to ensure businesses comply with regulations, such as health and safety requirements (Khan, 2011).

In many organizations, training and development opportunities have been encouraged to improve staff skills and improve operational efficiencies (Hughes et al., 2016). This has resulted in an increase in professional development opportunities to extend skills and knowledge in the workforce and allow organizations to take advantage of technological advances.

As part of quality improvement and patient safety in health, Australia introduced continuing professional development requirements in 2015 to educate staff about current advances in health and care practices and the use of innovative technologies in healthcare (Australian Health Practitioner Regulation Agency, 2019).

The recent 2019 coronavirus pandemic (COVID-19) has had a major impact on teaching and learning, with organizations and higher educational facilities worldwide shifting to online platforms instead of the traditional face-to-face learning environment (Dhawan, 2020; Pokhrel & Chhetri, 2021). In health for example, e-learning in specialized medical training, such as in surgical settings, can include virtual patient cases, digital modelling, online tutorials, and standardized videos and images (Jayakumar, 2015).

Despite large investments in workplace education and professional development activities, there is little evidence about the effectiveness of online education compared to traditional face-to-face learning (Vaona et al., 2018). There is a variety of individual metrics for measuring training effectiveness and evaluation frameworks like the Kirkpatrick evaluation model (Kirkpatrick, 1994; Kirkpatrick & Kirkpatrick, 2006) for the measurement and evaluation of learning. However, there are no agreed standardized methods to measure effectiveness and no assessment of outcomes between online and traditional learning (Vaona et al., 2018).

Background

In response to the COVID-19 pandemic, governments around the world introduced a series of phased restrictions and lockdowns to manage the spread of the disease. This included limiting face to face interactions and encouraging online work, training, and education. In health, the COVID-19 pandemic resulted in a significant increase in e-learning across many aspects of the professional development education and training program. This highlighted a need for better measurement, assessment, and evaluation of online learning.

With a significant uptake in online education and training, health organizations need to ensure that professional development training allows health care professionals to maintain and improve standards of practice through the development of knowledge, skills, and behavior. This process requires robust methods for the measurement, assessment, and evaluation of online education. In this rapidly changing environment health organizations are keen to know about the changes in e-learning practices and outcomes across all aspects of health. This rapid review identifies some of the new and emerging methods and practices for evaluating e-learning. This includes building on previous reviews that were limited in focus and identifies changes to practice, to assess what is already known about e-learning practice and gaps in evaluation methods.

Previous Systematic Reviews

In the past ten years there have been seven health-related systematic reviews undertaken to investigate the effectiveness of online training to improve participants' knowledge, skills and competencies (Barteit et al., 2020; Campbell et al., 2019; Lima et al., 2019; Moehead et al., 2020; Rouleau et al., 2019; Salter et al., 2014; Zafar et al., 2014). Looking at e-learning in nursing, pharmacy, radiology, dementia, and orthodontics. Many of the reviews identified a need for better measurement, assessment, and evaluation of online learning.

The challenges reported within these reviews highlighted a need to understand whether e-learning models can improve professional practice, professional knowledge, and the long-term effects compared to face-to-face learning. Many of the studies identified in the reviews were small-scale and short-term, often with limited granularity of reported details, overrepresentation of the effects of e-learning intervention, and underrepresentation of patient and practice outcomes.

Previous systematic reviews have focused on very specific areas within health without looking across the health landscape to identify and report different practices. This review covers ten years (including two pandemic years) during which advances in internet bandwidth, technology, and software have supported a shift to online training.

Objectives of the Review

This review aimed to identify new methods of measurement and assessment, as well as gaps and limitations to initiate discussion of valid evaluation within the health field. The objective is to map the state of research to determine what evaluation methods are currently used in health-related online education. In addition, the review aimed to summarize the strengths and limitations of these evaluation methods and recommend which of these methods could be used to measure the effectiveness of online health education.

Methods

We conducted a rapid review to identify online education evaluation methods specific to health-related training. A rapid review is an abbreviated systematic review that gathers and synthesizes study findings in a short amount of time. A rapid review can be used to address a wide range of issues and to help provide recommendations that can be used to inform policy and systems decisions (Tricco et al., 2017). Methods and results were reported using Rapid Reviews to Strengthen Health Policy and Systems (Tricco et al., 2017) and the 2020 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Page et al., 2021). See Supplementary Table 1. For this review, we define online learning and e-learning as an educational intervention that is delivered electronically through computer networks with *no physical classroom attendance.* The review does not include face-to-face or blended education models.

Protocol Development

The protocol was developed based on the population, intervention, comparison, and outcome (PICO) framework (Huang, Lin, & Demner-Fushman, 2006). Table 1 provides an overview of the protocol used to inform the search strategy.

Table 1

Protocol Development Using the PICO Framework

Parameter	Description
Population	Health care professionals or health-related students.
Intervention	Health-related courses delivered online with no face-to-face component.
Comparison	Type of evaluation method used.
Outcome	Performance, effectiveness, and limitations of the evaluation component.

Database Search

Seven databases (PubMed, ProQuest, Education Resources Information Centre [ERIC], Cumulative Index to Nursing and Allied Health Literature [CINAHL], Scopus, PsychInfo, and Medline) were searched for studies published between 2011 and early 2021. Using appropriate search strings and Medical Subject Headings (MeSH), the keywords used were related to the PICO framework and included e-learning, performance, efficiency, evaluation, assessment, and Kirkpatrick (see Table 2).

Table 2

Search Method and Number of Results per Database

Search Method	Pubmed	ProQuest	EBSCO (ERIC, CHINAHL)	Scopus	Ovid (PsychInfo, Medline)
TITLE-ABS-KEY (Kirkpatrick OR ADDIE OR Kaufman OR Brinkerhoff OR LTEM OR reflective OR "course evaluation" OR "education assessment" OR "evaluation model" OR LMS OR HRIS OR "personalised learning" OR "personalized learning" OR QILT OR "learning satisfaction") AND (elearning OR e-learning OR "electronic learning" OR "online learning" OR "online training" OR "open learning" OR "massive open online courses") AND ("return on investment" OR ROI OR performance OR efficiency OR efficacy OR cost OR financial) AND 2011-2021 AND ENGLISH AND Article OR Review AND Open Access (peer reviewed scholarly and unrestricted online access)			76	119	

(Kirkpatrick OR ADDIE OR Kaufman OR Brinkerhoff OR LTEM OR reflective OR "course evaluation" OR "education assessment" OR "evaluation model" OR LMS OR HRIS OR "personalised learning" OR "personalized learning" OR QILT OR "learning satisfaction") AND (elearning OR e-learning OR "electronic learning" OR "online learning" OR "online training" OR "open learning" OR "massive open online courses") AND ("return on investment" OR ROI OR performance OR efficiency OR efficacy OR cost OR financial) AND 2011-2021 AND ENGLISH AND Full-Text		7
AB-TI-SU((Kirkpatrick OR ADDIE OR Kaufman OR Brinkerhoff OR LTEM OR reflective OR "course evaluation" OR "education assessment" OR "evaluation model" OR LMS OR HRIS OR "personalised learning" OR "personalized learning" OR QILT OR "learning satisfaction") AND (elearning OR e-learning OR "electronic learning" OR "online learning" OR "online training" OR "open learning" OR "massive open online courses") AND ("return on investment" OR ROI OR performance OR efficiency OR efficacy OR cost OR financial)) AND 2011-2021 AND ENGLISH AND Full Text AND Peer Reviewed	139	
TITLE-ABS ((Kirkpatrick OR ADDIE OR Kaufman OR Brinkerhoff OR LTEM OR reflective OR "course evaluation" OR "education assessment" OR "evaluation model" OR LMS OR HRIS OR "personalised learning" OR "personalized learning" OR QILT OR "learning satisfaction") AND (elearning OR e-learning OR "electronic learning" OR "online learning" OR "online training" OR "open learning" OR "massive open online courses") AND ("return on investment" OR ROI OR performance OR efficiency OR efficacy OR cost OR financial)) AND 2011-2021 AND ENGLISH AND Articles OR Reviews AND Full-Text	121	

Note. Abbreviations: ADDIE, Analysis, Design, Development, Implementation, and Evaluation; CINAHL, Cumulative Index to Nursing and Allied Health Literature; ERIC, Education Resources Information Centre; HRIS, Human Resource Management System; QUILT, Quality Indicators for Learning and Teaching; LMA, Learning Management System; LTEM, Learning-Transfer Evaluation Model; ROI, Return on Investment.

Inclusion Criteria

Eligible studies were defined as scholarly, peer-reviewed articles published between 2011 and 2021 available in full text. A timeframe of 10 years was agreed upon and selected to limit the results of studies published using distance education methods described above. This included e-learning interventions relevant to health involving higher education students or healthcare personnel (i.e., continuing professional development) with an evaluation component. To understand the effectiveness of the evaluation components, eligible studies were required to report on the strengths and weaknesses of the methods used.

Exclusion Criteria

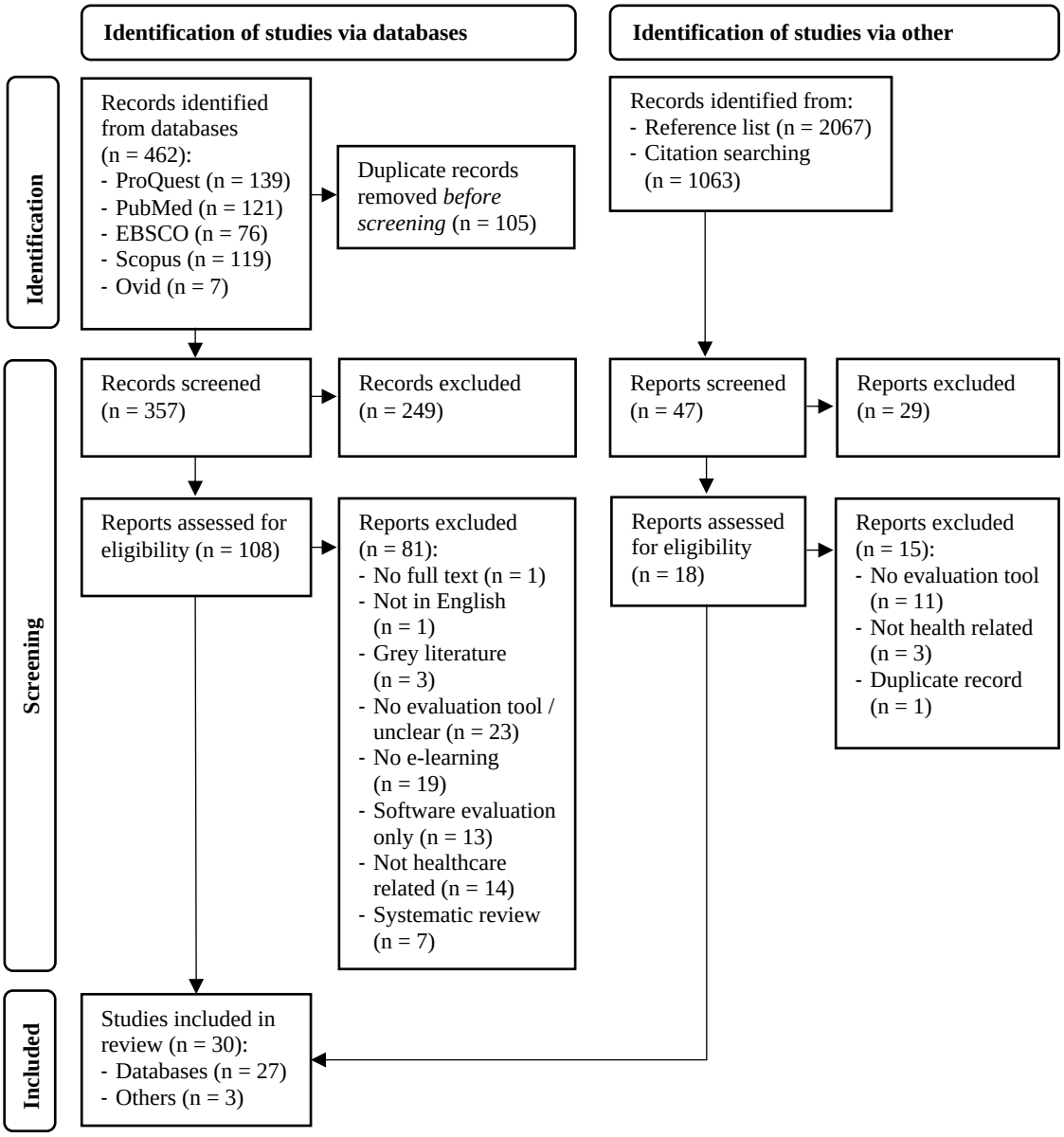
Grey literature articles, book chapters, conferences, opinions, proposals, or comment pieces were excluded from the review. In addition, technology acceptance, software evaluations concerning the e-learning platform, and medical interventions (such as clinical trials) were also removed since learning effectiveness was the focus of the review. Any blended or hybrid learning models, which included face to face or correspondence-based learning not completely delivered online, were excluded.

Screening and Study Selection

After restricting the database search to full-text, peer-reviewed articles, a total of 462 studies were retrieved and imported into Covidence Systematic Review Software (Veritas Health Innovation, 2019). Of these, 105 were duplicates leaving 357 for screening. Two reviewers screened the studies for relevance based on titles and abstracts, and then later by full text. The screening strategy was broad, looking to exclude articles that were not health-related, had no mention of an online education component or met the exclusion criteria. Uncertain articles were retained for review in the full-text screening stage. Of the 357 studies, 108 were retained for full-text screening, and 81 were finally excluded. The final 27 articles were considered appropriate and retained. Reference lists were exported from the Scopus database and citations from the retained articles were exported using the Publish or Perish software (Harzing, 2007). These articles were then imported into Excel and screened by title and abstract by the two reviewers, where three articles were retained. There were 30 articles included in the final selection. A PRISMA flow diagram shows the articles selected for inclusion and exclusion (Figure 1).

Figure 1

*PRISMA **F**low **D**iagram for **S**tudy **S**election (adapted from Page et al., 2021)*



Data Extraction

Data were extracted into Microsoft Excel using a template designed by the reviewers (see Appendix A) that included the country in which the study was undertaken, study design, education setting, course, population, evaluation methods, limitation of evaluation component, and study design considerations. The data were extracted by the two reviewers, who independently identified emerging themes and then agreed upon the outcome.

Results

Study Characteristics

A total of 30 articles published between 2011 and 2021 were included in the final review (see Appendices A and B). The number of participants within studies ranged from 16 (Adwan, 2016) to 3,752 (Hegerius et al., 2020). The studies were from 16 different countries, with the most common from the United States (9 studies), followed by Spain (3 studies) and Canada (3 studies). One multinational study, based in Sweden, used data from 137 countries (Hegerius et al., 2020). Study specifications are summarized in Appendix B.

Educational Level

The education level of participants varied, with the majority of studies from continuing professional development (15), followed by undergraduate education (9), then a combination of undergraduate, postgraduate, and continuing education (3), postgraduate education (2), and one combination of undergraduate students and teaching staff (Garrett et al., 2013).

Discipline

Several studies specified the healthcare discipline of the student population, with the majority from medicine (10), followed by nursing (8), then pharmacy (2). Ohers were from multiple disciplines (6), and a small number were from nutrition (Heuberger et al., 2019), health research (Tannenbaum & van Hoof, 2018), health informatics (Adwan et al., 2016), and global health (Lee et al., 2020).

Educational Institution

Course delivery was online, with more than half facilitated by universities (16 or 53%), followed by hospitals (6) and then a combination of universities and health centers (6), one research center (Hegerius et al., 2020), one combination research center and university (Tannenbaum & van Hoof, 2018) and one council (Willman et al., 2016).

Study Designs

The majority of studies used quasi-experimental designs (14), followed by descriptive designs (9), randomized controlled trials (3), and mixed methods (2). Others included a case study (Peterson et al., 2016) and a qualitative study (Prosser et al., 2021).

Evaluation Methods

Methods for evaluating e-learning effectiveness were the focus of this review and are summarized in Table 3. This section describes the tools and methods that were used in the literature to assess learning effectiveness. The methods include student participation, student satisfaction, performance measures, and training models, among others.

Table 3

Summary of Evaluation Methods and Their Limitations

Examples	Limitations
Student participation	
The proportion of students who participated in and completed the course Class attendance records System log data of students' interaction on the learning platform and participation in discussion forums Self-report questionnaire feedback that asks about participation	Poor measurement of learning outcomes Participation does not explain learning platform usage Unable to explain student dropouts or participation rates during the course
Students' reaction to training program	
Reactions can be used during the course and at the end to evaluate student satisfaction Self-reported questionnaires using Likert-type scales and open-ended questions It can also be obtained from focus groups It can also measure students' acceptance of the learning platform Often administered with incentives such as reminder emails and cash incentives	Poor measurement of learning outcomes and overemphasized use in the literature Difficulties obtaining adequate responses when conducted at the end of the course compared to mid-way through Most questionnaires were designed for the course with no prior validation When questionnaires are made voluntary or little incentives were given, lower response rates and response biases occur It does not allow updating of course delivery when courses were rated poorly if conducted at the end of the course Unable to explain reasons behind course withdrawals and student satisfaction over time
Performance measures: Assessment of knowledge	
Measured knowledge acquisition in the form of assessments, exams and final grades Some studies used validated knowledge-based questions to measure course-specific changes in knowledge before commencement and at the end of the course (pre-test/post-test)	Difficulty determining knowledge acquisition from assessment and exam scores alone No justification between course pass rate and knowledge acquisition Lower response rates with post-test measures when they do not count towards final grades It does not measure the long-term impact of knowledge acquisition

Performance measures: Long-term or follow-up	
Measured knowledge transfer over time using follow-up questionnaires Timeframes ranged from one month to four years after course completion, and some used multiple follow-up periods Follow-ups were identified as the most useful tool to measure knowledge transfer after course completion	Risks of low response rates when little incentives were given Requires resources including time and money to conduct compared to no follow-up
Self-efficacy	
Typically uses pre-test/post-test self-report questionnaires, validated and non-validated Used in combination with course evaluation and participation questionnaires	Poor measurement of learning outcomes Similar issues with other questionnaires, including low response rates and self-report bias
The Kirkpatrick Model	
Well-researched evaluation model with three levels: Level 1: Reaction—satisfaction and self-efficacy questionnaires Level 2: Learning—knowledge-based assessments Level 3, Behavior—follow-up questionnaires Level 4, Result—use of workplace information system data, rarely measured	Most studies use some aspects of the model in terms of Levels 1 and 2, which are poor measurements of learning outcomes Often Levels 3 and 4 are not measured without rationale, which are more robust measures of learning performance Requires follow-up evaluations or access to workplace data that may be costly Does not measure return on investment Limited research into the utility of the model for online learning
Other methods	
Focus groups Written reflections Feedback for student performance Electronic portfolios Cost-effectiveness	No standard methodology for these tools Requires resources including time and money to train staff in their use and conduct Feedback was only effective when delivered in real-time during the course and not after Electronic portfolios were only used to evaluate clinical practice skills

Student Participation

A small proportion of studies measured participation using a variety of methods including class attendance (Lee et al., 2020), interaction with class discussion forums or completing class exercises (Adwan, 2016; Carrizosa et al., 2018; dos Reis et al., 2019; Salinas et al., 2017), obtaining learning platform analytic data (Reese et al., 2021; Wlodarczyk et al., 2017), and finally student evaluation about their participation experience (Figuccio, 2020; Liaw et al., 2016; Peterson et al., 2016). While participation data reflects student reactions, it does not evaluate learning effectiveness (Lima et al., 2019) and neglects to inform teachers of how students used online platforms (Backhouse et al., 2017). Carrizosa et al. (2018) further reported that while students were participating below staff expectations, the data could not provide reasons behind the participation rates.

Students' Reaction to Training Programs

Student reaction to the course is a subjective measure (such as students' self-reported satisfaction with the course) that is typically completed mid-way (formative) or towards the end of the course (summative). Less than half of the studies measured student reaction, or acceptance of pedagogy, using either of these methods (13; e.g., Adwan, 2016). Hegerius et al. (2020) measured students' acceptance of the information system or platform used to deliver the course. Evaluation tools that were delivered as formative (mid-way) studies were found to have higher response rates (e.g., 85%; Peterson et al., 2016), compared to summative evaluations that were completed towards the end of the course, which had lower response rates (e.g., 62.4%; Backhouse et al., 2017). Questionnaires were typically voluntary, and issues included low response rates, such as 13.2% (Hegerius et al., 2020), and some report high course dropout rates (dos Reis et al., 2019). Incentives included regular reminders using email (Hegerius et al., 2020) or by earning points that contributed to their final grades (Adwan, 2016). Studies attributed low response rates when participation in questionnaires was voluntary (e.g., Whitt et al., 2016). Peterson et al. (2016) identified the advantage of early evaluations, as poorly rated courses were able to respond quickly and make changes when questionnaires were conducted mid-way through the course.

Another limitation to these methods is the inability to explain the reasons behind course withdrawals (Reese, 2021) or to capture student satisfaction with the course over time (Tannenbaum & van Hoof, 2018). Evaluations that were conducted mid-way through the course were helpful in updating course delivery when courses were rated poorly. It was reported that qualitative, open-ended surveys provided varying degrees of information, from too little to too much information, and was the least useful aspect of the course evaluation (Le Marne et al., 2020). It was identified that when questionnaires are voluntary, there is a potential problem of selection bias between those who respond and those who do not (Hegerius et al., 2020). Poor response rates can also impact the reliability of the information from questionnaires (Garrett et al., 2013). Adwan (2016) used Google Docs to conduct the evaluations and reported issues with the useability of the information system by staff and security concerns with students.

Performance Measures: Assessment of Knowledge

Several studies used grades from assessment tasks and final exams on two or more occasions to assess student knowledge (7; e.g., Annan et al., 2020). However, these articles did not discuss the value of the metrics used in their assessments or exams. For example, participants had to pass an examination to complete the course by achieving 60 percent or more (Carrizosa et

al., 2018), while another used a 100 percent pass rate (Willman et al., 2018). Types of assessments varied from multiple-choice questions (e.g., Schulz-Quach et al., 2018; Whitt et al., 2016) to oral assessments (e.g., Elzainy et al., 2020). Electronic portfolios were another method designed to assess clinical practice (Garrett et al., 2013). While studies did not report the limitations or deficiencies of their grading systems, using portfolios to measure clinical competency raised student concerns around privacy and confidentiality (Garrett et al., 2013).

Studies also varied in the format and delivery of questionnaires to measure changes in knowledge. Tannenbaum and van Hoof (2018) used a self-report questionnaire to test students' knowledge after the course. However, the authors identified that the questionnaire had not been previously validated (2018). Studies included those with externally validated questionnaires to measure students' performance (e.g., Kemper, 2017; Willman et al., 2018) and studies which had validated their own questionnaires (e.g., Heuberger et al., 2019; Schulz-Quach et al., 2018). Some questionnaires were specific to their subject content, such as stroke assessment (Gorchs-Molist et al., 2020), drug dispensing (dos Reis et al., 2019), and seizure management (Le Marne et al., 2016) and others measured self-directed learning readiness (Gagnon et al., 2015; Reviriego et al., 2014). Finally, the study by Kemper (2017) focused on questionnaires specific to measuring mindfulness (refer to Table 3) but did not measure the long-term impact on the participants.

Pre-test/post-test designs were also used to measure students' knowledge (e.g., Salinas et al., 2017) and performance (e.g., Backhouse et al., 2017) before and after training. However, some studies reported high dropout rates in the post-test phase (Annan et al., 2020; dos Reis et al., 2019; Gagnon et al., 2015; Reese, 2021; Reviriego et al., 2014), while others reported difficulties in accurately measuring the long-term impact of knowledge acquisition (Backhouse et al., 2017).

Performance Measures: Long-term or Follow-up

Various studies used follow-up questionnaires, ranging from one month to four years after course completion. Follow-up questionnaires implemented at one month had a 78 percent response rate (Wlodarczyk et al., 2017), while others implemented at eight months achieved 67 percent (Salinas et al., 2017). Gorchs-Molist et al. (2020) reported multiple follow-up periods, including after 1–2 years (71% response rate) and 3–4 years (91% response rate). Several studies identified the need to follow-up participants but could not undertake this process (Le Marne et al., 2016; Liaw et al., 2016; Simonsen et al., 2014; Uden-Holman et al., 2014).

Kirkpatrick's Training Evaluation Model

Various articles identified the need for learning evaluation methods, such as Kirkpatrick's Training Evaluation Model (Kirkpatrick & Kirkpatrick, 2006). This model has four levels of training outcomes: Level 1 (reaction) measures student responses about the quality of training; Level 2 (learning) quantifies learning using assessments and exams; Level 3 (behavior) measures the extent to which learning can be applied to the workplace; and Level 4 (results) measures how training has impacted organizational goals (Bates, 2004). Six studies reported on Kirkpatrick's evaluation model. Single measures ranged from self-reported student satisfaction at Level 1 (Hegerius et al., 2020) to course completion at Level 3 (Reese, 2021). In addition, hospital information system data were used to determine changes in compliance rates for Level 3 and clinical outcomes for Level 4 (Gorchs-Molist et al., 2020; Liaw et al., 2016). Level 3 was also measured using a 6-month post-test evaluation (Uden-Holman et al., 2014). However, one study

reported improvements at Levels 1 and 2 but not at Levels 3 or 4 (dos Reis et al., 2019). Other studies identified in the review that did not use Kirkpatrick's model have inadvertently used elements from Kirkpatrick's model (e.g., reaction, learning, behavior, and outcomes).

Self-efficacy

Bandura's Conceptual Model of self-efficacy (Bandura, 1997), part of Level 2 of the Kirkpatrick evaluation model (Kirkpatrick & Kirkpatrick, 2006), were delivered using pre- and post-test methods and included a Likert-type scale design with a validated nine-item, ten-point scale (Aper et al., 2012); a non-validated three-item, ten-point scale (Tannenbaum & van Hoof, 2018); and a one-item, five-point scale (Reese, 2021). Schulz-Quach et al. (2018) identified the need to measure self-efficacy to improve methodological quality. However, the voluntary nature of these self-report questionnaires has had response rates as low as 60 percent (Whitt et al., 2016).

Other Methods

Other themes that emerged from the review were focus groups, reflections, and cost-effectiveness. Numerous studies within the review articles used focus groups. For example, focus groups using open-ended questions can examine learning experiences within the course (Garrett et al., 2013), while software such as FocusGroupIt can address themes using a SWOT (i.e., strengths, weaknesses, opportunities, threats) analysis (Elzainy et al., 2020). In contrast, Heuberger et al. (2019) conducted focus groups before their study and used the results to pilot and validate their course satisfaction survey. Furthermore, focus group transcripts and written reflections can be combined using thematic analysis to provide student feedback (Posser et al., 2021). However, the use of focus groups and reflections was impacted by the time requirements to train staff, and written reflections provided little additional information. Finally, formative feedback delivered to students in real-time has demonstrated success at commending high performers and encouraging low performers to improve their grades (Adwan, 2016).

Few studies reported the cost-effectiveness of the evaluation methods and their outcomes. Several studies reported the need for additional cost-effectiveness research (e.g., Kemper, 2017). Other studies commented on the cost savings of delivering training online instead of face-to-face (e.g., Martinez et al., 2019). While the cost-effectiveness of the evaluation methods was not always measured, some authors evaluated cost-effectiveness from self-perception scores, increase in knowledge and self-efficacy (e.g., Carrizosa et al., 2018).

Discussion

This rapid review identified research articles that used evaluation methods to measure the impact health-related online education has on student performance. When organizations are faced with emerging technology-driven changes and digital disruptors, as with the COVID-19 outbreak, there is a need for learning and development to support improvement in workplace performance. However, training and learning design and delivery methods need to be assessed to ensure education is efficient and relevant.

This is particularly important in assessing the change from traditional face-to-face delivery to online models for teaching and learning. This review attempts to summarize research in this area and provide actionable and relevant evidence to help organizations plan learning interventions and measure the impact of student performance over time.

What were the Evaluation Methods and their Limitations?

From this review, it is apparent there is no single method that comprehensively measures the impact of learning interventions. Based on the level of data obtained, tools ranged from self-report evaluations measuring participation rates and student's satisfaction to metric data such as course grades (and in health, very specific performance metrics related to clinical information system data e.g., changes to hospital length of stay) (Liaw et al., 2016). The distinct types of data collection were self-report questionnaires that used Likert-style scoring with or without open-ended questions; assessment tasks and exams that were scored on a grading system; and focus groups, reflections and portfolios that provided qualitative information and uncovered themes. There were also various applications of these tools, including before and after the course (e.g., using a pre-test/post-test approach), mid-way and at the end of the course (in a formative and summative approach), only at the end of the course (e.g., with final grades or course evaluations), and follow-ups after the course (e.g., six-month follow-up; Wlodarczyk et al., 2017; Carrizosa et al., 2018; Salinas et al., 2017; Gorchs-Molist et al., 2020). While available tools were used with a combination of students and teachers, the Kirkpatrick's evaluation method was the only complete framework described and used in some of the studies identified by the review (Hegerius et al., 2020; Reese, 2021; dos Reis et al., 2019; Gorchs-Molist., 2020; Liaw et al., 2016; den-Holman et al., 2014).

Several systematic reviews identified deficiencies with some of these evaluation methods, such as whether the use of non-validated measurement tools affected the validity of the outcomes or whether the training design affected student performance (Campbell et al., 2019; Moehead et al., 2020). Recommendations highlighted the need for validated frameworks to better synthesize learning effectiveness and a need for more robust study designs to enhance research methodologies (Barteit et al., 2020; Salter et al., 2014; Rouleau et al., 2019).

Which Tools are Appropriate for Measuring the Effectiveness of Online Education?

Given the limited evaluation methods and limitations with study designs (outlined in Table 3), caution is needed when assessing the utility of tools used. Nevertheless, the results suggest that using a framework, such as Kirkpatrick's, enables hierarchical measurement of learning effectiveness based on research-based findings. However, this involves using a collection of several types of evaluation tools, such as self-report questionnaires and comparing final grades, all with their own strengths and weaknesses.

Even though there is little evidence to suggest that e-learning has different outcomes to traditional face-to-face learning (Vaona et al., 2018), studies using Kirkpatrick's framework tended to limit the depth of learning effectiveness to Levels 1 and 2 and do not investigate how the course impacts performance in the workplace or over time.

While there are several explanations for limited evaluation, including the time and costs associated with measuring student performance, the following summary is a breakdown of tools and how they could be improved.

Participation Rates

Participation rates are the most straightforward metric used to measure student engagement. However, they provide little evidence about learning outcomes and fail to explain the reasons behind student dropouts. The use of self-report data could supplement information about dropouts and how these can be reduced in the future.

Self-Report Evaluation Tools

Self-reported course evaluation tools were found to be appropriate in obtaining students' experiences during the course and are best conducted early to allow time for the educators to adapt their teaching towards student preferences.

Furthermore, the use of teacher evaluation tools during the teaching enables more transparency within the teaching/class environment. Additionally, the use of validated course evaluation tools was found to be more robust than teacher-designed tools, providing more standardized results and allowing comparisons between classes (Barteit et al., 2020; Salter et al., 2014; Rouleau et al., 2019). Focus groups have also been used to validate evaluation questionnaires (Heuberger et al., 2019).

Knowledge Assessment

Assessments that are completed before, during and after the course (e.g., pre-test/post-test models and self-efficacy questionnaires) provided evidence of measuring learning effectiveness. Suggestions to enhance this method include the use of mandatory, or incentive-driven, delivery of these tools (e.g., grade incentives). This can help reduce the likelihood of nonresponse errors and self-selection bias. Validated subject-specific questionnaires are also recommended.

While mid-course assessments and final exams provide data about individual attainment, they fail to measure the long-term impact of the course (Backhouse et al., 2017; Kemper, 2017). One approach to measure this is the use of student follow-up questionnaires (Garrett et al., 2013), although they face the same responsiveness challenges of surveys.

Focus Groups and Written Reflections

Some studies used focus groups to obtain qualitative information about staff and student experiences from the course (Garrett et al., 2013). However, these were time-consuming and costly. Alternatives include written reflections which are less time-consuming to administer and provide a similar level of information (Prosser et al., 2021). Suggestions to improve written reflections were around incentives for completion (e.g., grade incentives), using validated methodologies (e.g., SWOT), providing real-time feedback (i.e., immediate versus delayed feedback), and capturing long-term data (e.g., post-course follow-up).

Recommendations

The majority of studies investigating the impact of online education programs did not consider a cost analysis or return on investment. This is important because the cost of one evaluation tool compared to another is an essential factor in the decision-making process around cost and benefit. Furthermore, an analysis of clinical significance was not performed in the majority of articles, as most outcomes were based on predefined goals such as achieving a pass mark (i.e., a 60+ percent score) or positive self-evaluation scores. Recommendations for future research are for more longitudinal studies that capture the effects of training after a six-month window and methods that can compare pre- and post-test outcomes.

Limitations

The rapid review process comes with several potential limitations, including the possibility that studies may have been missed (Tricco et al., 2017). This rapid review is not exhaustive, and as such, a search was not conducted on grey literature. The selected studies were from the academic research community and excluded evaluation methods within business and

private organizations other than hospitals and universities. Articles included in the analysis were limited to English, full-text studies, which may bias studies from high-income English language countries such as the United States and the United Kingdom, thus missing studies written in different languages. Further, the data extraction has been performed on learning interventions from training programs and did not consider evaluation methods available from education platforms (e.g., Learning Management System). The focal point of this rapid review was on learning outcomes and articles that only reported on technology acceptance were excluded. Furthermore, the purpose of the rapid review was to summarize evidence rather than evaluate effects, the evaluation of reported quantitative data from the studies were not the primary focus. Lastly, a critical appraisal was not performed and inter-rater reliability of selecting articles between the two reviewers was not measured, owing to the rapid nature of this review (Tricco et al., 2017). However, Table 3 includes a column that outlines the limitations mentioned within the identified studies.

Conclusion

This rapid review investigated the various methods and types of tools used to measure learning effectiveness for online education. The review included studies of online education within the discipline of healthcare and observed studies reporting positive effects of these training interventions. Education and development opportunities were identified as an important function that allows professionals to keep "up to date" with current practices. However, providing these opportunities within and across busy work schedules is complex. Although advancements in technology offer some alternatives on how professional development can be structured and organized, there is limited evidence to support what makes online teaching and learning effective. Many of the studies identified in this review suggest that professional development should provide support over a sustained period to achieve the most effective outcomes. However, due to constraints like funding, time, organizational structure, and policy, this is not often the case. Finally, while education evaluation tools and methods are popular in assessing the effectiveness of the training programs, the evidence suggests that using any evaluation method in isolation is insufficient. Suggestions were to adopt previously validated frameworks (not limited to the Kirkpatrick model) and appropriately implement them within the course curriculum. Developing a framework which identifies 'best practices' in the organization, development, delivery and evaluation of training can help support effective and sustainable education programs.

Declarations

Authors declare no known conflicts of interest.

This research was supported by the Digital Health CRC. Digital Health CRC Limited is funded under the Commonwealth Government's Cooperative Research Centres Program. The authors would like to thank Dr Melanie Haines, Education Manager, Digital Health CRC.

References

Adwan, J. (2016). Dynamic online peer evaluations to improve group assignments in nursing e-learning environment. *Nurse Education Today*, *41*, 67–72. https://doi.org/10.1016/j.nedt.2016.03.028

Ahmed, H., Allaf, M., & Elghazaly, H. (2020). COVID-19 and medical education. *The Lancet Infectious Diseases*, *20*(7), 777–778. https://doi.org/10.1016/S1473-3099(20)30226-7

Annan, R. A., Aduku, L. N. E., Kyei-Boateng, S., Yeun, H. M., Pickup, T., Pulman, A., Monroy-Valle, M., Ashworth-Hill, A., Jackson, A. A., & Choi, S. (2020). Implementing effective eLearning for scaling up global capacity building: Findings from the malnutrition elearning course evaluation in Ghana. *Global Health Action*, *13*(1). https://doi.org/10.1080/16549716.2020.1831794

Aper, L., Reniers, J., Koole, S., Valcke, M., & Derese, A. (2012). Impact of three alternative consultation training formats on self-efficacy and consultation skills of medical students. *Medical Teacher*, *34*(7), e500–507. https://doi.org/10.3109/0142159X.2012.668627

Australian Health Practitioner Regulation Agency. (2019, December 3). *Continuing professional development*. https://www.ahpra.gov.au/Registration/Registration-Standards/CPD.aspx

Backhouse, M., Fitzpatrick, M., Hutchinson, J., Thandi, C. S., & Keenan, I. D. (2017). Improvements in anatomy knowledge when utilizing a novel cyclical "Observe-Reflect-Draw-Edit-Repeat" learning process. *Anatomical Sciences Education*, *10*(1), 7–22. https://doi.org/10.1002/ase.1616

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Barteit, S., Guzek, D., Jahn, A., Bärnighausen, T., Jorge, M. M., & Neuhann, F. (2020). Evaluation of e-learning for medical education in low- and middle-income countries: A systematic review. *Computers & Education*, *145*, 103726. https://doi.org/10.1016/j.compedu.2019.103726

Bates, R. (2004). A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning*, *27*(3), 341–347. https://doi.org/10.1016/j.evalprogplan.2004.04.011

Campbell, K., Taylor, V., & Douglas, S. (2019). Effectiveness of online cancer education for nurses and allied health professionals: A systematic review using Kirkpatrick evaluation framework. *Journal of Cancer Education*, *34*(2), 339–356. https://doi.org/10.1007/s13187-017-1308-2

Carrizosa, J., Braga, P., Albuquerque, M., Bogacz, A., Burneo, J., Coan, A. C., Contreras, G., Guilhoto, L., Izquierdo, A., Ladino, L., Lin, K., Manreza, M. L., Morales, L., Rios, L., Solarte, R., Valente, K., Venegas, V., Uscategui-Daccarett, A., & Yacubian, E. M. (2018). Epilepsy for primary health care: A cost-effective Latin American E-learning initiative. *Epileptic Disorders*, *20*(5), 386–395. https://doi.org/10.1684/epd.2018.0997

Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, *49*(1), 5–22. https://doi.org/10.1177/0047239520934018

dos Reis, T. M., Baldoni, A. d. O., Campos, A. M., Girotto, E., Guidoni, C. M., Obreli-Neto, P. R., & Pereira, L. R. L. (2019). A distance-learning course to improve drug-dispensing behaviors among Brazilian community pharmacists. *American Journal of Pharmaceutical Education*, *83*(8), 1703–1711. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6900818/

Elzainy, A., El Sadik, A., & Al Abdulmonem, W. (2020). Experience of e-learning and online assessment during the COVID-19 pandemic at the College of Medicine, Qassim University. *Journal of Taibah University Medical Sciences*, *15*(6), 456–462. https://doi.org/10.1016/j.jtumed.2020.09.005

Figuccio, M. J. (2020). Examining the efficacy of e-service-learning. *Frontiers in Education*, *5*. https://doi.org/10.3389/feduc.2020.606451

Gagnon, J., Gagnon, M. P., Buteau, R. A., Azizah, G. M., JettÉ, S., Lampron, A., Simonyan, D., Asua, J., & Reviriego, E. (2015). Adaptation and evaluation of online self-learning modules to teach critical appraisal and evidence-based practice in nursing: An international collaboration. *CIN: Computers, Informatics, Nursing*, *33*(7), 285–294. https://doi.org/10.1097/CIN.0000000000000156

Garrett, B. M., MacPhee, M., & Jackson, C. (2013). Evaluation of an eportfolio for the assessment of clinical competence in a baccalaureate nursing program. *Nurse Education Today*, *33*(10), 1207–1213. https://doi.org/10.1016/j.nedt.2012.06.015

Gorchs-Molist, M., Sola-Munoz, S., Enjo-Perez, I., Querol-Gil, M., Carrera-Giraldo, D., Nicolas-Arfelis, J. M., Jimenez-Fabrega, F. X., & Perez de la Ossa, N. (2020). An online training intervention on prehospital stroke codes in Catalonia to improve the knowledge, pre-notification compliance and time performance of emergency medical services professionals. *International Journal of Environmental Research and Public Health*, *17*(17), 6183. https://doi.org/10.3390/ijerph17176183

Harzing, A. W. (2007). *Publish or Perish* (version 7). https://harzing.com/resources/publish-or-perish

Hegerius, A., Caduff-Janosa, P., Savage, R., & Ellenius, J. (2020). E-Learning in pharmacovigilance: An evaluation of microlearning-based modules developed by Uppsala Monitoring Centre. *Drug Safety*, *43*(11), 1171–1180. https://doi.org/10.1007/s40264-020-00981-w

Heuberger, R., & Clark, W. A. (2019). Synchronous delivery of online graduate education in clinical nutrition: An inquiry into student perceptions and preferences. *Journal of Allied Health*, *48*(1), 61–66. https://pubmed.ncbi.nlm.nih.gov/30826832

Huang, X., Lin, J., & Demner-Fushman, D. (2006). Evaluation of PICO as a knowledge representation for clinical questions. *AMIA Annual Symposium proceedings.* 359–363. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839740/

Hughes, A. M., Gregory, M. E., Joseph, D. L., Sonesh, S. C., Marlow, S. L., Lacerenza, C. N., Benishek, L. E., King, H. B., & Salas, E. (2016). Saving lives: A meta-analysis of team training in healthcare. *Journal of Applied Psychology*, *101*(9), 1266–1304. https://doi.org/10.1037/apl0000120

Jayakumar, N., Brunckhorst, O., Dasgupta, P., Khan, M. S., & Ahmed, K. (2015). E-Learning in surgical education: A systematic review. *Journal of Surgical Education*, *72*(6), 1145–1157. https://doi.org/10.1016/j.jsurg.2015.05.008

Kemper, K. J. (2017). Brief online mindfulness training: Immediate impact. *Journal of Evidence-Based Complementary and Alternative Medicine*, *22*(1), 75–80. https://doi.org/10.1177/2156587216639199

Khan, R. A. G., Khan, F. A., & Khan, M. A. (2011). Impact of training and development on organizational performance. *Global Journal of Management and Business Research*, *11*(7). https://swsu.ru/sbornik-statey/pdf/8-Impact-of-Training-and-Development-on-Organizational-Performance.pdf

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels* (vol. 1). Berrett-Koehler Publishers.

Kirkpatrick, J. D., & Kirkpatrick, D. (2006). *Evaluating training programs: The four levels*. Berrett-Koehler Publishers. https://search.alexanderstreet.com/view/work/bibliographic_entity%7Cbibliographic_details%7C2341903

Le Marne, F. A., McGinness, H., Slade, R., Cardamone, M., Balbir Singh, S., Connolly, A. M., & Bye, A. M. (2016). Evaluation of an e-learning resource on approach to the first unprovoked seizure. *Journal of Paediatrics and Child Health*, *52*(9), 896–900. https://doi.org/10.1111/jpc.13277

Lee, S. J., Park, J., Lee, Y. J., Lee, S., Kim, W. H., & Yoon, H. B. (2020). The feasibility and satisfaction of an online global health education course at a single medical school: A retrospective study. *Korean Journal of Medical Education*, *32*(4), 307–315. https://doi.org/10.3946/kjme.2020.178

Liaw, S. Y., Wong, L. F., Lim, E. Y., Ang, S. B., Mujumdar, S., Ho, J. T., Mordiffi, S. Z., & Ang, E. N. (2016). Effectiveness of a web-based simulation in improving nurses' workplace practice with deteriorating ward patients: A pre- and post-intervention study. *Journal of Medical Internet Research*, *18*(2), e37. https://doi.org/10.2196/jmir.5294

Lima, M. S., Tonial, F. G., Basei, E., Brew, M. C., Grossmann, E., Haddad, A. E., Rivaldo, E. G., Vargas, I. A., & Bavaresco, C. S. (2019). Effectiveness of the distance learning strategy applied to orthodontics education: A systematic literature review. *Telemedicine Journal and e-Health*, *25*(12), 1134–1143. https://doi.org/10.1089/tmj.2018.0286

Martinez, C., Castellano, Y., Andres, A., Fu, M., Feliu, A., Anton, L., Ballbe, M., Fernandez, P., Cabrera, S., Riccobene, A., Gavilan, E., Baena, A., Margalef, M., Tigova, O., Quiros, N., Guillen, O., Company, A., & Fernandez, E. (2019). Impact of an online training program in smoking cessation interventions in hospitals. *Journal of Nursing Scholarship*, *51*(4), 449–458. https://doi.org/10.1111/jnu.12469

Moehead, A., DeSouza, K., Walsh, K., & Pit, S. W. (2020). A web-based dementia education program and its application to an Australian web-based dementia care competency and training network: Integrative systematic review. *Journal of Medical Internet Research*, *22*(1), e16808. https://doi.org/10.2196/16808

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Peterson, J. L. (2016). Formative evaluations in online classes. *Journal of Educators Online*, *13*(1). https://doi.org/10.9743/JEO.2016.1.8

Pokhrel, S., & Chhetri, R. (2021). A literature review on impact of COVID-19 pandemic on teaching and learning. *Higher Education for the Future*, *8*(1), 133–141. https://doi.org/10.1177/2347631120983481

Prosser, M., Stephenson, T., Mathur, J., Enayati, H., Kadie, A., Abdi, M. M., Handuleh, J. I. M., & Keynejad, R. C. (2021). Reflective practice and transcultural psychiatry peer e-learning between Somaliland and the UK: A qualitative evaluation. *BMC Medical Education*, *21*(1), 58. https://doi.org/10.1186/s12909-020-02465-y

Reese, D., Dolansky, M. A., Moore, S. M., Bolden, H., & Singh, M. K. (2021). Quality improvement education innovation: Evaluation of Coursera MOOC "Take the Lead on Healthcare Quality Improvement." *Journal of Research in Nursing*, *26*(1–2), 62–78. https://doi.org/10.1177/1744987120982644

Reviriego, E., Cidoncha, M. A., Asua, J., Gagnon, M. P., Mateos, M., Garate, L., de Lorenzo, E., & Gonzalez, R. M. (2014). Online training course on critical appraisal for nurses: Adaptation and assessment. *BMC Medical Education*, *14*, 136. https://doi.org/10.1186/1472-6920-14-136

Rouleau, G., Gagnon, M. P., Cote, J., Payne-Gagnon, J., Hudson, E., Dubois, C. A., & Bouix-Picasso, J. (2019). Effects of e-learning in a continuing education context on nursing care: Systematic review of systematic qualitative, quantitative, and mixed-studies reviews. *Journal of Medical Internet Research*, *21*(10), e15118. https://doi.org/10.2196/15118

Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Science in the Public Interest*, *13*(2), 74–101. https://doi.org/10.1177/1529100612436661

Salinas, J., Munoz, C., Albagli, A., Araya, G., & Vio, F. (2017). The contribution of distance education to health promotion in Chile. *Health Promotion International*, *32*(5), 913–921. https://doi.org/10.1093/heapro/daw023

Salter, S. M., Karia, A., Sanfilippo, F. M., & Clifford, R. M. (2014). Effectiveness of e-learning in pharmacy education. *American Journal of Pharmaceutical Education*, *78*(4), 83. https://doi.org/10.5688/ajpe78483

Schulz-Quach, C., Wenzel-Meyburg, U., & Fetz, K. (2018). Can elearning be used to teach palliative care?—medical students' acceptance, knowledge, and self-estimation of competence in palliative care after elearning. *BMC Medical Education*, *18*(1), 82. https://doi.org/10.1186/s12909-018-1186-2

Simonsen, B. O., Daehlin, G. K., Johansson, I., & Farup, P. G. (2014). Improvement of drug dose calculations by classroom teaching or e-learning: a randomised controlled trial in nurses. *BMJ Open*, *4*(10), e006025. https://doi.org/10.1136/bmjopen-2014-006025

Tannenbaum, C., & van Hoof, K. (2018). Effectiveness of online learning on health researcher capacity to appropriately integrate sex, gender, or both in grant proposals. *Biology of Sex Differences*, *9*(1), 39. https://doi.org/10.1186/s13293-018-0197-3

Tricco, A. C., Langlois, E. V., & Straus, S. E. (2017). *Rapid reviews to strengthen health policy and systems: A practical guide*. World Health Organization. https://www.who.int/alliance-hpsr/resources/publications/rapid-review-guide/en/

Uden-Holman, T., Bedet, J., Walkner, L., & Abd-Hamid, N. H. (2014). Adaptive scenarios: A training model for today's public health workforce. *Journal of Public Health Management and Practice*, *20 Suppl 5*, S44–48. https://doi.org/10.1097/PHH.0000000000000067

United Nations. (2020). *Policy brief: Education during COVID-19 and beyond.* https://www.un.org/development/desa/dspd/wpcontent/uploads/sites/22/2020/08/sg_policy_brief_covid-19_and_education_august_2020.pdf

Vaona, A., Banzi, R., Kwag, K. H., Rigon, G., Cereda, D., Pecoraro, V., Tramacere, I., & Moja, L. (2018). E-learning for health professionals. *Cochrane Database of Systematic Reviews*, 8. https://doi.org/10.1002/14651858.CD011736.pub2

Veritas Health Innovation. (2019). *Covidence systematic review software*. https://www.covidence.org

Whitt, K. J., Macri, C., O'Brien, T. J., & Wright, S. (2016). Improving nurse practitioners' competence with genetics: Effectiveness of an online course. *Journal of the American Association of Nurse Practitioners*, *28*(3), 151–159. https://doi.org/10.1002/2327-6924.12282

Willman, B., Grankvist, K., & Bolenius, K. (2018). Evaluation of the clinical implementation of a large-scale online e-learning program on venous blood specimen collection guideline practices. *Clinical Chemistry and Laboratory Medicine*, *56*(11), 1870–1877. https://doi.org/10.1515/cclm-2018-0051

Wlodarczyk, D., Chylinska, J., Lazarewicz, M., Rzadkiewicz, M., Jaworski, M., Adamus, M., Haugan, G., Lillefjell, M., & Espnes, G. A. (2017). Enhancing doctors' competencies in communication with and activation of older patients: The Promoting Active Aging (PRACTA) computer-based intervention study. *Journal of Medical Internet Research*, *19*(2), e45. https://doi.org/10.2196/jmir.6948

Zafar, S., Safdar, S., & Zafar, A. N. (2014). Evaluation of use of e-Learning in undergraduate radiology education: A review. *European Journal of Radiology*, *83*(12), 2277–2287. https://doi.org/10.1016/j.ejrad.2014.08.017

Appendix A
Characteristics of Included Studies

First Author Year Country	Study design Intervention	Setting Course Population Sample size	Evaluation methods	Limitations of evaluation methods	Study limitations or recommendations
Adwan 2016 USA	Delayed feedback versus immediate feedback.	University Health informatics Undergraduate students n = 16	Course evaluations (Questionnaires) and performance (Final assessment grade).	Nonstandard scale used. Bias from scale with high rated self-report scores. Use of Google Docs unfamiliar with some staff. Security concerns with survey platform.	High dropout rate. Groups were formed based on peer groups.
Annan 2020 Ghana	Compares four course delivery methods	University Malnutrition e-learning course Undergraduates n = 931	Pre- and post-test assessments, self-reported questionnaires, and course completion.	Self-reported course completion, limitations for the other evaluation techniques were not mentioned.	Low post-study participation rate. No significant improvement between pre- and post-assessments were found.
Aper 2012 Belgium	Three course delivery methods	Online training Medical student competencies Postgraduates n = 186	Self-efficacy was measured using a validated questionnaire. Competencies measured by examining assessment responses.	The quality of students' performance was not measured.	Longer studies are recommended to study long-term impacts of the alternative training formats. Suggest that a qualitative study may help to validate results. Future studies could also focus on the long-term development of the leaning outcomes.
Backhouse 2017 UK	Pre-test/post-test design comparing online and face-to-face	University Anatomy (medicine) Undergraduate n = 209	Performance, Student perceptions student test scores Self response survey measuring - timing, delivery, guidance, technical, others.	Evaluations tools did not identify how students used the online platform and how they engaged with the activities. Long-term impact of knowledge acquisition was also not measured.	The differences between the two methods—online and face-to-face were discussed as limitations for comparing the findings. Evaluate student use of the teaching method.

Carrizosa 2018 Uruguay	Pre-test/post-test design Evaluation of an e-learning course	Moodle Epilepsy training Primary care physicians n = 105	Participation (forum contributions), course completion (final grades), course evaluation (questionnaires), cost-effectiveness (student investment), long-term learning (questionnaires after 6-months).	Participation in forums were below staff expectations and overlapped other modules for some students due to the short duration of each module, thus students may not have benefited from this tool.	No limitations or bias discussed.
dos Reis 2019 Brazil	Pre-test/post-test evaluations of an e-learning course	Moodle Drug-dispensing Pharmacists n = 472	Course effectiveness used Kirkpatrick's model levels 1 to 3. Participant satisfaction (5-item questionnaire), learner outcomes (pre-post-test), performance improvement (simulated practice). Patient or health outcomes (level 4) was not measured.	There were positive results obtained from Kirkpatrick's levels 1 (satisfaction) and 2 (knowledge) analyses. No improvement occurred in the conduct (level 3) of the skills and abilities assessed in simulated dispensing practice.	High level of dropouts, authors suggest higher course fees may reduce the level of dropouts. Mystery shopper technique minimized bias. Educational strategies may address the lack of practical activities in distance learning.
Elzainy 2020 KSA	Compares face-to-face learning and online	University Various medical courses Undergraduates n = 250	Final assessment scores, student satisfaction survey, weekly staff perception reports, and staff learning experiences (focus groups).	None discussed.	No limitations or bias discussed.
Figuccio 2020 USA	Compares face-to-face learning and online	University Atypical Development Undergraduates n = 58	Student experience (end-of-course questionnaire), course evaluations (questionnaire), student reflection papers (coded by tutors).	None discussed.	No limitations or bias discussed.

Gagnon 2015 Canada, Spain	Pre-test/post-test design	University Critical appraisal Nurses n = 86	Various questionnaires (knowledge acquisition, self-directed learning readiness, and satisfaction with training program).	None discussed.	Lack of control and randomization. High dropout with no reason for withdrawing from course.
Garrett 2013 Canada	Effect of and e-portfolio on clinical skills	University Science in nursing Students, n = 36 Staff, n = 18	Clinical placement experience (portfolios), LMS use analytics, instructor / student surveys, and focus groups.	Differences in instructor use of the assessment tools (access to portfolio during assessments). Poor return of questionnaires.	Small sample size and smaller. Issues with data transparency from portfolio's (suggest restricting access during assessments).
Gorchs-Molist 2020 Spain	Pre-test/post-test evaluations of an e-learning course	Hospital Stroke assessment Medical professionals n = 30	Kirkpatrick's model levels 1 to 4: 1, satisfaction survey. 2, pre-post knowledge test. 3, compliance rates with clinical system. 4, proportion of codes and prehospital care times.	Data collected was limited to prehospital setting, so effectiveness data post clinical care remains unknown. Unable to capture data on the clinical outcome of the patients.	Data collected was limited to prehospital setting, so clinical significance not directly measured. Future studies should seek to include further in-hospital clinical variables.
Hegerius 2020 Sweden, multinational	Evaluation of an online course	Research center Pharmacovigilance Health professionals n = 3752 from 137 countries	E-Learning evaluation as overall satisfaction (Kirkpatrick's evaluation model level 1). Use of LLMS system (survey and logged usage data).	No measure of the impact and cost effectiveness of the training. Low response rate may have come from a delay in post-course survey. Addressed the selection bias from those who responded to surveys compared to those who did not.	Results may not be relevant to other learning fields. Recommendations to evaluate knowledge to determine if there was any behavior change after course.
Heuberger 2019 USA	Satisfaction of synchronous and asynchronous learning	University Clinical nutrition Master's students n = 176	Evaluate student satisfaction for courses delivered in synchronous and asynchronous modes using open-ended surveys and focus groups.	None discussed.	Future research suggestions were continuing to gauge student preferences for satisfaction with the emerging education technologies.

Kemper 2017 USA	Improvements in course outcomes	Health center Mindfulness training Health professionals n = 146	10-item Cognitive and Affective Mindfulness Scale–Revised (CAMS-R) 15-item Mindful Attention Awareness Scale (MAAS) 39-item Five Facet Mindfulness Questionnaire (FFMQ).	Did not measure long-term impact and cost-effectiveness.	Unable to be generalized since there was no randomization, was conducted at one institution, and course training was voluntary.
Le Marne 2020 Australia	Pre-test/post-test evaluations of an e-learning course	Hospital Pediatric seizures Medical specialists n = 50	Performance from assessment scores Course satisfaction open-ended survey Self-rated clinical knowledge and self-efficacy.	The level of detail from self-reported qualitative feedback varied between extremes of too little information or too much detail and was reported as least useful aspects of evaluating the course.	Suggests for longitudinal follow-up to determine transference of knowledge into clinical practice of management of pediatric seizures.
Lee 2020 South Korea	Compares face-to-face learning and online	University Global health Undergraduates n = 146	Participation rate Satisfaction of the course Student preferences online and face-to-face Academic achievement.	Difficulty making comparisons with final exams scores between two years since exams differed in content and delivery. Limitations for other outcome measures not mentioned	Not generalizable to all medical students since the course was targeted at second year students at one medical institution. Written final exams scores were incomparable.
Lesser 2019 USA	Pilot study of different songs and analysis of user data	University Introductory statistics Undergraduate n = 77	Student performance from course assessment Learning system usage from analysis of log records.	None discussed.	Recommends introducing student feedback to improve completion of tasks. Implementing randomized controlled trials to compare performance under varied treatment conditions.
Liaw 2016 Singapore	Pre-test/post-test evaluations of an e-learning course	Hospital Deteriorating patients Ward nurses n = 99	Increase in knowledge from post-test scores Perceived attitudes to learning transfer Hospital length of stay from cohort workplaces Evaluation of course was guided by Kirkpatrick's evaluation model.	Unable to evaluate the effect of patient outcomes beyond the scope of the study.	Chance of missing other mitigating factors since results were analyzed based on documented outcomes. Lack of a control group to improve robustness of study outcomes.

Martinez 2019 Spain	Pre-test/post-test evaluations of an e-learning program	Hospital Tobacco intervention Hospital clinicians n = 127	Comparison of pre-post self-reported questionnaire of 63-items on a 10-point discrete scale internal reliability, a = 0.77.	None mentioned, however, the timing of the delivery of post-evaluation questionnaires would be important to consider clinical significance of the study. Focus was on the self-reported use of the intervention rather than success of program.	Focus was on the success of clinicians implementing the techniques and not the success of the program. Lack of comparison group. Results relied on self-reported responses. The sample were mainly female (85.7%), registered nurses (63%) and 45.7% had never smoked with physicians accounting for 7.9% of the sample size.
Peterson 2016 USA	Case study analysis of two online courses	University Medical terminology and pathophysiology Undergraduates n = 55	Open ended evaluations delivered at first half and at the end of the course, student performance measures (course assessments and exam grades).	Students did not use the feedback from the second evaluation, which was designed to inform them about improvements. The second course, pathophysiology, was rated poorly and received many complaints. It was decided to continue the course face-to-face.	Limitations were the short time frame to transition the face-to-face course to the online format.
Prosser 2021 UK, Somaliland	Thematic analysis of post-program in reflective writing	University Clinical cases (psychiatry) Medical students n = 33	Thematic analysis of written reflections and post-program focus groups.	Program and thematic analysis was conducted in English which was not the primary language spoken with participants from Somaliland.	Unable to evaluate the 27% of participants who dropped out of the program.
Reese 2021 USA	Pre-test/post-test evaluations of an e-learning course	Healthcare quality improvement Self-selected learners n = 88	Uses Kirkpatrick's model Levels 1 to 3 to evaluate satisfaction, learning outcomes, and knowledge.	No data from participant withdrawals from course, heavy evaluation burden placed on learners, embedded evaluations may have contributed to increased dropout rates.	Longitudinal analysis suggested for future studies to examine learning sustainability and behavior change outcomes.
Reviriego 2014 Spain	Pre-test/post-test evaluations of an e-learning course	Hospital Critical appraisal Nurses n = 50	Questionnaires to measure knowledge, satisfaction, and self-learning ability.	Identified that some participant dropouts were due to difficulty of content within the course.	Limitations were a lack of control group and random assignment. Small sample size. Difficulty determining success or failure of course.
Salinas 2017 Chile	A quali-quantitative evaluation	University Primary healthcare Postgrad course, n = 162 Technician course, n = 172	Evaluation of learning measured by participation and assessment task Program evaluation using pre-post and follow-up surveys.	None discussed.	No limitations or bias discussed.

Schulz-Quach 2018 Germany	Descriptive proof of concept study	University Palliative care Medical students n = 670	Evaluates the acceptance of eLearning and self-efficacy using a questionnaire Learning from exam of palliative care competencies.	Standard limitations from questionnaire-based evaluation.	No baseline measurements in palliative care prior to the eLearning course.
Simonsen 2014 Norway	Randomized controlled parallel design	Hospital Medication calculations Nurses n = 183	Knowledge on medication calculations using an exam. Questionnaires to evaluate the course (perceived difficulty of the course, learner satisfaction, usefulness of course).	None discussed.	Controlled test conditions may be regarded as a limitation (reflective of real-life clinical environment). Higher dropout in online course compared to face-to-face course.
Tannenbaum 2018 Canada	Pre-test/post-test evaluations of an e-learning course	Various Sex and gender science Research staff n = 543, 463, 435	Pre- and post-questionnaires to measure knowledge, self-efficacy, and self-reported behavior change intent.	Knowledge questions were not previously validated. Behavioral intent was self-reported and not indicative of actual changes in behavior, assessments were directly after completion of course and may not capture effects over time.	Participants were recruited via email and may have led to enrolment of a biased sample of researchers already interested in the course.
Uden-Holman 2014 USA	Descriptive design to evaluate two adaptive scenarios	University Psychological First Aid Public health personnel n = 112	Unspecified evaluation data on user satisfaction, content relevancy, and knowledge (Kirkpatrick's model Level 1).	None discussed.	Future recommendations include conducting a follow-up evaluation that addresses Kirkpatrick's level III, which measures transfer of learning within the work setting.
Whitt 2016 USA	Pre-test/post-test evaluations of an e-learning course	University Genetics course Nurse Practitioner students n = 140	5-point Likert scale, 65-item self-report pre-test/post-test survey measuring genetic competence and confidence plus a 21-item pre-post course multiple choice test to measure knowledge of genetics.	Measurement of genetic competencies were self-reported and not objectively measured and did not evaluate student outcomes relating to other areas such as legal, social, leadership or research.	Students were obtained from a single university and therefore not generalizable. Only 60 per cent of students responded to surveys. Finally, only an online course was evaluated, suggestions for a comparative studding face-to-face.

Willman 2018 Sweden	Course outcomes compared across groups over time	County Council Venipuncture Various technicians n = 879	Venipuncture skills questionnaire and pre-post course evaluation survey. Short answer questions (qualitative content analysis).	None discussed, the venipuncture questionnaire was previously validated. Information was not provided on number of students who failed compared to those who passed the course.	Poor participation rates over time. Future research on follow-up participants' practices and educational program efficiency.
Wlodarczyk 2017 Norway	Random assignment and control group	Primary Care facilities Active Aging General Practitioners n = 225	Self-reported scales administered on course completion and at 1-month follow-up: Communication Scale, Patients Expectations Scale, Scale, Attitude Toward Treatment and Health Scale, and Self-Efficacy Scale.	Outcome variables were self-reported.	There were unsatisfactory power sample calculations as most facilities approached declined to participate and there was more dropout rates during the progress of the study. Recommendations to consider eLearning satisfaction among doctors.

Appendix B

Tabulated List of Included Studies (n = 30)

First author	Year	Country	Study design	Sample size	Institution	Study level	Discipline	Evaluations
Adwan	2016	USA	Descriptive	16	U	UG	HI	P, RE, F
Annan	2020	Ghana	Descriptive	931	U	UG	V	K
Aper	2012	Belgium	RCT	186	U	PG	M	K, SE
Backhouse	2017	UK	Quasi-experimental	209	U	UG	M	K
Carrizosa	2018	Uruguay	Descriptive	105	U, HC	CPD	M	P, RE, K, CE
dos Reis	2019	Brazil	Quasi-experimental	472	U, HC	CPD	PH	P, RE, K, KM
Elzainy	2020	KSA	Quasi-experimental	250	U	UG	M	RE, K, FG
Figuccio	2020	USA	Quasi-experimental	58	U	UG	SW	P, RE
Gagnon	2015	Canada, Spain	Quasi-experimental	86	U	CPD	N	RE, K
Garrett	2013	Canada	Mixed methods	18	U	UG, T	N	RE, K, FG
Gorchs-Molist	2020	Spain	Quasi-experimental	30	HO	CPD	V	K, FU, KM
Hegerius	2020	Multinational	Descriptive	3752	R	CPD	PH	RE, KM
Heuberger	2019	USA	Descriptive	176	U	PG	N	K
Kemper	2017	USA	Descriptive	146	HC	CPD	V	K
Le Marne	2020	Australia	Quasi-experimental	50	HO	CPD	M	K, FU
Lee	2020	South Korea	Quasi-experimental	146	U	UG	GH	P, RE
Liaw	2016	Singapore	Quasi-experimental	99	HO	CPD	N	P, FU, KM
Martinez	2019	Spain	Quasi-experimental	127	HO	CPD	V	RE
Peterson	2016	USA	Case study	55	U	UG	M	P, RE
Prosser	2021	UK, Somaliland	Qualitative	33	U	UG	M	FG, RF
Reese	2021	USA	Descriptive	88	U, HC	CPD	N	P, K, KM, SE
Reviriego	2014	Spain	Quasi-experimental	50	HO	CPD	N	K
Salinas	2017	Chile	Mixed methods	334	U	PG, CPD	N	P, K, FU
Schulz-Quach	2018	Germany	Descriptive	670	U	UG	M	K
Simonsen	2014	Norway	RCT	183	HO	PG, CPD	N	RE, FU
Tannenbaum	2018	Multinational	Quasi-experimental	543	U, R	CPD	HR	K, SE
Uden-Holman	2014	USA	Descriptive	112	U	CPD	V	RE, FU, KM
Whitt	2016	USA	Quasi-experimental	140	U	PG, CPD	N	K, SE
Willman	2018	Sweden	Quasi-experimental	879	C	CPD	V	K
Wlodarczyk	2017	Norway	RCT	225	HC	CPD	M	P, FU

Note. C = county council; CE = cost-effectiveness; CPD = Continuing Professional Development; F = feedback; FG = focus groups; FU = follow-up; GH = global health; HC = health center; HI = Health Informatics; HO = hospital; HR = health research; K = knowledge; KM = Kirkpatrick model; M = medicine; N = nursing; NT = nutrition; P = participation; PH = pharmacy; R = research center; RCT = Randomized Controlled Trial; RE = reaction; RF = reflections; SE = self-efficacy; SW = Support Work; T = teachers and instructors; U = University; UK = United Kingdom; USA = United States of America; V = various disciplines; KSA = Kingdom of Saudi Arabia.