E-learning modules to improve clinical reasoning and practice: a prospective comparative study [version 2; peer review: 2 approved, 2 approved with reservations]

Fabiola Stollar1,2, Bernard Cerutti3, Susanne Aujesky1, Daniel Scherly3, Mathieu Nendaz3,4, Annick Galetto-Lacour2,5

1Department of Pediatrics, Gynecology & Obstetrics, Geneva University Hospitals, Geneva, 6 Rue Willy-Donzé, Geneva 1211, Switzerland
2Faculty of Medicine, University of Geneva, Geneva, 1 Rue Michel-Servet 1206, Switzerland
3Unit for Development and Research in Medical Education (UDREM), Faculty of Medicine, University of Geneva, Geneva, 1 Rue Michel-Servet, 1206, Switzerland
4Service of General Internal Medicine, University Hospitals of Geneva, Geneva, 4 Rue Gabrielle-Perret-Gentil, 1205, Switzerland
5Pediatric Emergency Division, Children's Hospital, University Hospitals of Geneva, Geneva, 6 Rue Willy-Donzé, Geneva 1211, Switzerland

Abstract

Background

Controversy remains about whether e-learning can improve clinical competences. Our study aimed to compare the effects of e-learning versus traditional education on medical students' reasoning and how they applied their knowledge to clinical competences, assess factors associated with e-learning that might influence exam scores, and evaluate medical students' satisfaction with these two learning methods.

Methods

Prospective study of 299 medical students in two fourth-year pediatric clerkship cohorts (2016–17 and 2017–18) in Switzerland.

Results

We found no evidence of a difference in students' reasoning or how they applied their knowledge to competences in clinical case resolution, whether they had followed e-learning modules or attended traditional lectures. The number of quizzes taken and being female
were factors associated with better scores. Even though overall satisfaction with the two learning methods was similar, students claimed that they learned more in e-learning than in traditional lectures and that e-learning explained learning objectives better.

Conclusions

E-learning could be used as a supplement or alternative to traditional face-to-face medical teaching methods without compromising teaching quality. E-learning modules should be better integrated into medical students’ curricula but avoid the risk of curriculum overload, especially in case of repeated COVID-like context.

Keywords

knowledge acquisition, e-learning, traditional lectures, clinical skills in pediatrics, medical education, learning method, student satisfaction
Introduction

The processes of teaching and learning are domains in constant development. Previous studies have shown that traditional educational models have several limitations, mainly temporal and spatial\(^1\). The United Kingdom’s General Medical Council has made significant contributions to shaping today’s undergraduate medical curriculum, which had been considered overloaded\(^2\). For over a decade, medical schools have been working to transform teaching by reducing the number of lectures, implementing team-based and self-directed learning, or promoting individualized, interprofessional education\(^3\).

Computers and the internet are used more and more in educational settings. Although e-learning means different things to different people, in its broadest sense, it is the use of the internet for education. E-learning’s pedagogical approach typically aspires to being flexible, engaging, and learner-centered. It overcomes students’ and teachers’ geographical and temporal constraints and encourages collaboration and communication\(^4\). Virtual learning programs, available any time, facilitate access to knowledge and add interactivity to study, encouraging reflection and problem solving and stimulating independent learning and self-discipline\(^5\).

The emergence of coronavirus disease (COVID-19) significantly disrupted medical education and required the intensive, prompt attention of medical educators. The University of Geneva’s Faculty of Medicine quickly transferred the theoretical parts of the curriculum to an online format. Small groups convened online in virtual team settings, and clinical skills sessions either occurred online or were deferred\(^6\). Most small group lectures and tutorials had to become live remote courses. This often added to the workload of doctors and faculty already burdened with the large influx of COVID-19 patients\(^8\)-\(^10\).

The present study was carried out before the coronavirus pandemic, but it now has important implications given that the current context is forcing us to adapt teaching formats. A recent systematic review found that distance education was not different from traditional education in terms of effectiveness\(^11\). A Cochrane review also compared e-learning to traditional learning and showed albeit with low certainty of evidence that e-learning made little or no difference to health professionals’ skills and knowledge\(^12\). However, these reviews did not test competences in clinical case resolution, which is a core competence of medical education, but the outcomes measured were health professionals’ knowledge retention and behaviors. To compare these two teaching methods, this study analyzed level 1 of Kirkpatrick’s model by means of a satisfaction questionnaire sent to students, and level 2 by assessing the success score of questions testing clinical case resolution\(^13\).

It thus seemed appropriate to move beyond testing learning equivalence and important to evaluate competences in clinical case resolution, assess the factors associated with better learning, and investigate whether some groups succeed more than others. Furthermore, recent studies suggest that combining face-to-face training with e-learning is more flexible than other educational methods\(^14\)-\(^15\). The challenge for the sector will be to identify optimal ways of using e-learning as part of a blended approach to learning in clinical settings\(^16\).

The present research was carried out to compare the effects of e-learning and traditional lecture-based training on medical students’ reasoning and how they applied their knowledge to clinical case resolution by evaluating their performance in a written pediatrics exam. Our secondary aims were to evaluate medical students’ satisfaction with these two different learning methods and assess the factors associated with e-learning that might influence exam scores, such as the number of quizzes taken online, the quiz performance, and gender.

Methods

Study design and subjects

This was a prospective study involving two cohorts of medical students in the fourth year of their six-year curriculum...
and participating in an eight-week pediatric clerkship rotation (2016–17 and 2017–18 intakes) at the University of Geneva’s Faculty of Medicine, Switzerland. Three students repeating a grade due to failing on the previous year were excluded.

Ethics and consent

The study protocol was submitted to the Cantonal Research Ethics Commission of Geneva and was exempted from a full review process as it does not fall under the Article 2 of the Swiss Federal Law on research on human beings as it involved existing anonymized data, and was designed to evaluate the quality of undergraduate or postgraduate educational programs (authorised by the President of the Committee Dec 22nd 2015 via email). According to a 2009 decision made by the Ethics Committee of Geneva and the University of Geneva’s Faculty of Medicine Teaching Committee Office, research projects in the field of medical education, dealing with existing anonymized data, and designed to evaluate the quality of undergraduate or postgraduate educational programs, are exempt from the need for a full review process and formal approval by the Ethics Committee. The participant consent was waived by the committee. All the data were completely de-identified in accordance with the Safe Harbor method. All files and records were stored in local institutional data servers. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for observational studies.

Learning activities

Traditional learning

Medical students participate in several supervised clinical activities during their pediatrics clerkship and attend a standardized program of traditional case-based tutorials and lectures held for each clerkship rotation.

Online e-learning course format

Geneva Children’s Hospital has set up an educational website that allows medical students to achieve various learning objectives in pediatrics through individual learning using case-based scenarios. These deal with the most common situations faced in general pediatrics, neonatology, pediatric orthopedics, and pediatric surgery. Many of the learning objectives are covered exclusively in e-learning modules available on the educational website, which is available 24/7. All students were well familiar with e-learning because from their baccalaureate they have access to e-learning.

E-learning modules use the Moodle™ (Moodle Pty Ltd, West Perth WA 6872, Australia) platform. Each learning activity is structured as follows:

- An introduction including statements about specific learning objectives;

- A topic covered using a step-by-step approach, including one or several case-based vignettes; question–answer–feedback sections alternate with theoretical learning content throughout the resolution of the case;

- A quiz section enabling students to review the learning content; key features of the clinical case's diagnosis and management are tested using multiple-choice questions; several attempts per quiz are allowed, and students can review their attempts throughout their clerkship; quiz settings also allow students to see the correct answers, their scores, and get some feedback immediately after their attempt.

Pediatric exam evaluation method

Students take a written exam once, after they have completed their clerkship: students who did their clerkship between January and April (two groups of 30 students each) take the May exam; those who did their clerkship between June and December (three groups of 30 students each) take the January exam. The latter students thus have a more extensive overall clinical and theoretical background since they have completed clerkships in other fields in the meantime.

Exams are online and use either the CAMPUS (for computers) or iEXAM (for tablets) software provided by the Umbrella Consortium for Assessment Networks UCAN, Heidelberg, Germany). Exam content is designed to test clinical reasoning skills and theoretical knowledge. Students must deal with the step-by-step management of several clinical situations presenting different common pediatric complaints. Supplemental patient information, given sequentially, allows them to move towards case resolution. (Figure 1)

Clinical situations and question content are based on the learning objectives of the clerkship. Questions are developed by pediatric hospital faculty and reviewed in a multi-step procedure by a panel of experienced pediatric faculty. Before its utilization, each exam is tested on the computer/tablet by several members of the reviewer panel.

Comparing e-learning and traditional lectures

To compare the e-learning and lecture formats, two e-learning modules (on constipation and gastroesophageal reflux) were taught exclusively using traditional lectures (the 2016–17 academic intake) and exclusively using e-learning modules (the 2017–18 academic intake). Lectures were given by the authors of the e-learning modules so as to maximize similarities between the learning content in the lectures and e-learning modules. Students from the 2016–17 academic intake could not access any parts of the constipation and gastroesophageal reflux e-learning modules (introduction section, case-based vignette, or quiz) (Figure 1). We named the questions relating to these two subjects the teaching format experiment questions (TFEQ) as they tested the two learning formats (e-learning and traditional), and we analyzed and compared the TFEQ scores for these two academic intake groups.

Factors influencing scores in subjects taught using e-learning

Questions on subjects taught exclusively using e-learning (seven modules; Figure 1) were named the e-learning experiment
questions (EEQs). The following factors were considered: exam scores (excluding TFEQs and EEQs), the number of quizzes taken, quiz score, gender, and the cohort effect.

Access to every e-learning module made by individual students using their institutional login was documented. We considered the number of attempts that each individual made for each quiz, as well as the highest quiz score they obtained for each quiz.

Experimental questions (Figure 1) were asked in the same session of the following year to avoid a “session effect”, as we had previously shown in a retrospective study that our students were more likely to get higher scores in exams taken in the second session (January) than in the first (May). This difference could be explained by the fact that medical students taking exams in May have less overall clinical experience.

Exam sessions (May exam and January exam)
Each exam session also included 34–36 other items (i.e., non-experimental) covering a range of related topics. These items were different in each exam session. The topics tested using these items were taught in lectures, problem-based learning tutorials, and/or e-learning chapters. Thus, corresponding scores could be used as controlling factors for estimating each student’s level.

Figure 1. Subjects of questions by teaching format. Definitions: Mixed subjects: Different subjects and different questions used for each exam, these subjects were taught by lectures, Problem Based Learning-vignettes and/or e-learning chapters. Teaching format experimental questions (TFEQ): the computer-based written examination (CBWE) questions related to the constipation and gastroesophageal reflux, the two subjects taught either by traditional lectures in 2016/2017 or by e-learning in 2017/2018. E-learning experimental questions (EEQ): the CBWE questions on the seven subjects taught exclusively by e-learning. All experimental questions were kept identical during both May sessions as well as during both January sessions, to compare students with the same background. * These questions were both analyzed as TFEQ and EEQ.

Evaluation of student satisfaction
All the students were invited to evaluate teaching activities (on knowledge acquisition, the clarity of learning objectives, achieving learning objectives, curriculum adequacy, teacher preparedness, e-learning, and traditional lecture or tutorial learning activities) at the end of their eight-week pediatrics clerkship. Four-point Likert scale items were used in all our institutional surveys. We also analyzed the free text comments left in the questionnaires, without the use of assistive software.

Analysis
Scores (0 to 100) were given as the percentage of points scored divided by the maximum points achievable. Data were summarized using mean, median, and inter-quartile range (IQR), and Student’s t-tests or Kruskal–Wallis rank sum tests were used to compare the two different groups’ scores.

We used multivariate linear regressions to investigate associations between EEQ scores and all the exam’s other item scores, numbers of quizzes taken, mean scores obtained for the different quizzes (if a quiz was attempted several times,
Results

Study group

The study group included 299 (43.5% men; 56.5% women) fourth-year medical students from the 2016–17 and 2017–18 academic intakes; median age on the exam date was 24.1 years old (IQR 23.4–24.9). The average exam score was 77.0 (± standard deviation 8.8) for the 2016-17 academic intakes, and 79.7 (± 8.8) for the 2017-18 academic intakes; when excluding TFEQs and EEQs the average exam score was 76.0 (± standard deviation 9.4) for the 2016-17 academic intakes, and 79.6 (± 8.3) for the 2017-18 academic intakes. Regarding the topics always addressed with e-learning (lumbar pain, enuresis, sudden infant death syndrome, meningococcal meningitis, scoliosis) the number of quizzes taken was 2.0 (±2.8) for the 2016-17 academic intakes, and 2.8 (± 3.2) for the 2017-18 academic intakes.

Comparing e-learning with traditional lectures

Students exposed to interactive e-learning modules had almost the same TFEQ scores as students exposed to traditional lectures (mean 79%, median 80%, IQR 67%–100% vs. mean 80.3%, median 80%, IQR 67%–100%; p = 0.255) (Figure 2).

Factors influencing exam scores in subject areas taught using e-learning

A linear regression model showed that EEQ scores were associated with both the number of quizzes taken and gender, irrespective of the scores on non-experiment items (Table 1).

Survey regarding the two learning methods

Table 2 compares students’ evaluation questionnaire responses regarding e-learning and traditional lectures. Even though overall evaluations were almost identical, students reported learning more during e-learning modules than in traditional lectures and that their learning objectives were better explained in the e-learning modules. They felt, however, that traditional lectures were better integrated into the curriculum. Some examples of the positive comments written in the free text sections of the evaluation questionnaire included:

- Web site quality and organization: “Great website. Perfect for learning”; “The website is a real source for studying”; “The learning information is present and easily accessible”; and “Super site, very complete, and very educational”.

Figure 2. Total combined scores from tests on gastroesophageal reflux and constipation by teaching format. Teaching format experimental questions (TFEQ).
Table 1. Linear regression model of exam scores for items associated with e-learning subjects (EEQ) ($R^2 = 0.1814$).

| Covariate                                      | Univariate analysis |                     | Multivariate analysis** |                   | p       | p       |
|------------------------------------------------|---------------------|---------------------|-------------------------|-------------------|---------|---------|
|                                                | Coefficient | Standard error | p               | Coefficient | Standard error | p       |
| Control score (non-experiment items)*         | 0.564      | 0.113            | < 0.0001               | 0.568           | 0.114            | < 0.0001 |
| Number of quizzes taken**                     | 0.715      | 0.289            | 0.0142                 | 0.717           | 0.289            | 0.0140  |
| Quiz score***                                 | 3.029      | 1.788            | 0.0921                 | -0.213          | 1.778            | 0.9048  |
| Cohort effect§                                | -0.120     | 2.026            | 0.9530                 | -2.691          | 1.927            | 0.1645  |
| Gender§                                       | -3.735     | 2.017            | 0.0658                 | -4.140          | 1.880            | 0.0290  |

* range 49–99 (score)
** range 1–18
*** range 2.4–6.0 (points)
§ Men compared with women.

* Second cohort (2016–17 academic intake) compared with first cohort (2017–18 academic intake).
** Fisher statistic: 7.492 on 5 and 169 degrees of freedom ($p < 0.0001$), $R^2 = 0.1814$

Table 2. Satisfaction questionnaire comparing e-learning and traditional lectures (N = 201 respondents)*.

| E-learning | Traditional lectures |
|------------|----------------------|
|            | Mean ± SD            |
|            | Mean ± SD            |
| P          |
| I learned a lot during this activity | 3.70 ± 0.46 | 3.37 ± 0.56 | < 0.0001 |
| The learning objectives are clearly explained | 3.34 ± 0.73 | 3.21 ± 0.65 | 0.0259 |
| I was able to achieve the learning goals | 3.22 ± 0.61 | 3.30 ± 0.57 | 0.2166 |
| The activity is well integrated into the curriculum | 3.16 ± 0.79 | 3.40 ± 0.59 | 0.0002 |
| The teachers are well prepared | 3.54 ± 0.62 | 3.65 ± 0.48 | 0.0522 |
| Overall mean score | 3.38 ± 0.49 | 3.38 ± 0.48 | 0.69074 |
| SD = standard deviation |            |

*Data could not be desagregated by gender since the teaching evaluation forms used by the institution do not include a Gender item.

- Quality of e-learning: "E-learning is great!"; "The complementary e-learning site is a positive point".

However, some examples of the negative comments are:

- Size of the website/e-learning: "Website is very long"; "A complete website, but difficult to approach"; "Too much e-learning, lots of details, not clear what you need to know".

- Lack of time for e-learning: "Given the amount of e-learning we are asked to do, there is not enough time planned for it"; "There’s not enough time for e-learning".

- Poor integration of e-learning: "Lectures and e-learning scenarios sometimes overlap. It is not easy to get an overview".

Discussion
Summary of principal findings
The present study provided evidence to suggest that online learning is at least as effective as traditional teaching formats. Regarding the factors associated with better learning, as expected, exam scores for the items associated with e-learning subjects depended on students’ intrinsic capacities and skills. This was illustrated by the fact that the TFEQ and EEQ scores
were associated with the scores for non-experimental items. The number of quizzes taken and being a woman student were factors associated with better scores.

**Equivalence between e-learning and traditional learning**
We found no differences in the acquisition of reasoning competencies when comparing undergraduate e-learning with traditional lecture-based educational methods. Two recent reviews comparing e-learning with traditional learning also found them to be similarly effective, but they evaluated a different population of learners that was a mix of health professionals, nurses, doctors, and pediatric health consultants, and the Cochrane review even excluded trials involving undergraduate health professionals.11,12

Another strength of our study was that we compared the two teaching formats through clinical case resolution competences, whereas the existing reviews had only assessed simple knowledge retention and acquisition.13,14 We wanted to test e-learning in a real-life context, with modules integrated into the curriculum to provide a more flexible form of self-learning; one where the students chose the most appropriate moment to study. The format’s impact was tested during end-of-semester exams. This was in contrast to He et al., whose systematic review and meta-analysis only included studies using synchronous distance education (i.e., studies that tested knowledge retention for every learner in a synchronized manner, just after the teaching periods).15

**Beyond testing learning equivalence—factors influencing exam scores**
Our study moved beyond testing learning equivalence and evaluated the factors associated with better learning and whether some groups succeed more than others experiencing different teaching methods. The positive impact (better scores) of completing more quizzes was one interesting result. One study evaluating factors influencing examination scores among preclinical medical students showed that students who were highly motivated to study medicine had higher scores.16 Those who spent more daily hours preparing for exams, using the internet for academic purposes, and achieving their study targets also ranked higher in the ‘excellent’ group (scores ≥ 80%).17 Moreover, several publications supported by neuroscience have shown that giving students frequent tests as part of their training is an effective learning method.18

Our study also found evidence that women students got higher EEQ scores, even when controlled for the student’s intrinsic quality and the number of quizzes taken. Men and women show differences in examination performance across the spectrum of medical education, with woman medical students consistently outperforming their men counterparts, and this trend continues into some postgraduate specialties. Woman doctors are known to perform better in pediatrics, as tested in our study, and in sex-specific domains like obstetrics and gynecology. Pope et al. also showed that woman trainee general practitioners outperformed their men peers in overall clinical skills assessment, in each individual assessment domain, and in every area of the curriculum.

The difference in performance was most marked in the areas of women’s health and sexual health and least marked in cardiovascular health, rheumatology, and musculoskeletal health. Some differences have also been found regarding the use of certain Moodle resources, men’s perceptions that e-learning activities interfere with their social life, and women’s greater sense of duty.

**Student satisfaction and evaluation questionnaire**
Our study participants gave traditional learning and e-learning similar overall evaluations, and students were satisfied with both learning methods. Liyun He et al., however, found that the pooled effect size of overall satisfaction significantly favored e-learning over traditional education.19

It is, nevertheless, important to highlight that our medical students subjectively felt that they learned more during e-learning than during traditional lectures. This may be because e-learning modules repeated important information several times, initially in the specific subject’s question format, then in the theoretical explanations following the question, and then in the module’s associated quiz. Schiekirkka S et al. in a systematic review of the factors influencing students’ ratings in undergraduate medical education course evaluations indicated that overall course ratings were mainly influenced by students’ satisfaction with teaching and exam difficulty, rather than the objective standards of high-quality teaching.20 Furthermore, our medical students claimed that they lacked time for e-learning. They may have thought that traditional lectures were better integrated into the curriculum because they occurred at set times in set places, whereas, with e-learning, students had to decide when to study by themselves.

**Study limitations**
Our study had some limitations. It was carried out in a single faculty and the results might not be generalizable to others. We included all the students, however, hence avoiding the selection bias of studies conducted among volunteers. We did not evaluate costs or other economic aspects linked to the two teaching methods, but this was beyond the study’s scope. We could not evaluate the time students spent online accessing modules because their access was discontinuous; interruptions cannot easily be measured, preventing a reliable quantification of time spent online. We did not pre-assess students’ baseline knowledge and how it had developed. Finally, the number of experimental items was small. Further research could examine the issue of students’ feelings of isolation during e-learning and its consequences, which this study was unable to do.

Since there were no differences in the success rates for the items used in both sets of exams, this suggests that there was no transmission of information about the questions asked between the 2016–17 intake students and the 2017–18 intake students. However, we cannot be sure that there was no communication between the two groups.

Lastly, it would have been interesting to evaluate clinical reasoning development as a measured outcome of specific
teaching strategies using more authentic assessment design principles, but it was beyond the scope of our study specially because we evaluated pre-graduate students, it is too early in their journey to evaluate certain clinical aspects and reasoning.

Implications for future practice and research
The COVID-19 pandemic has caused unprecedented disruption to medical education and healthcare systems worldwide. This context led us to start discussing potential new ways of replacing traditional learning that would have minimal effects on the progression of students’ overall medical training. Transforming tutorials and lectures into e-learning modules could be a long-term solution for freeing up tutors’ and students’ time without losing quality. However, knowing that this would lead to fewer contacts and interactions between students, it might be wise to encourage them to follow their e-learning modules in groups so as to enhance exchanges between them, boost discussions about clinical case resolution, and mitigate their worries about the e-learning model.

Our study illustrated the importance of completing quizzes that put students into situations that mimic real-life practice using case-based scenarios. It also showed the importance of giving students the opportunity to self-assess after completing a stand-alone e-module. During interactive face-to-face courses, tutors give constant feedback based on discussions with students and the repetition of key points. During an e-learning module done alone, an end-of-module quiz in some way mimics the same role (repeating key points and enabling students to check whether they have understood them). It seems important that e-learning modules be designed so that a clinical case resolution quiz truly contributes to improve students’ clinical competencies. This involves students creating an intellectual process based on theoretical clinical skills and knowledge—training for later implementation in their medical practice.

Timetables should, nevertheless, reserve dedicated time slots for autonomous learning. Traditionally taught and timetabled activities should be linked to e-learning modules so that the latter are better integrated into the curriculum.

Another interesting feature of e-learning is the potential to externalize learning activities. Teaching and learning could be shared with other centers, universities, and hospitals: the technology for more flexible solutions is readily available. Further research should evaluate how to integrate e-learning with traditional methods to avoid curriculum overload and a lack of time for autonomous learning. It could also evaluate differences in examination performance amongst groups of trainees studying different medical specialties, how those differences affect planned educational activities so they can be adapted to students’ particular needs, and the new opportunities e-learning offers for learning and teaching.

Conclusions
We found no evidence of differences in medical students’ reasoning or how they applied their knowledge to competences in clinical case resolution between cohorts using e-learning and traditional lecture-based methods. E-learning could be used as a supplement or an alternative to traditional face-to-face medical teaching methods without compromising teaching quality. The number of quizzes taken, and being a woman were the two factors associated with better exam scores, highlighting the importance of self-assessment. Students were satisfied with both learning methods, but they claimed that they learned more through e-learning. We should find better ways to integrate e-learning modules into the medical student curriculum, paying special attention to avoiding curriculum overload, especially if there were to be a repeat of the COVID-19 context.

Data availability
Repository: E-learning modules to improve clinical reasoning and practice
https://doi.org/10.6084/m9.figshare.21769973

This project contains the following underlying data:
- FIGSHAREPaediatricELearningData.csv (Dataset regarding the e-learning quizzes and exam scores of the students)
- FIGSHAREPaediatricELearningDataREADME.txt (Description of the variables of the dataset regarding the e-learning quizzes and exam scores of the students)
- FIGSHAREPaediatricEvaluationSurveyData.csv (Dataset regarding the evaluation of the teaching activities by the students)
- FIGSHAREPaediatricEvaluationSurveyDataREADME.txt (Description of the variables of the dataset regarding the evaluation of the teaching activities by the students)

This project contains the following extended data:
FIGSHAREPaediatricNonExperimentalTopicsDuringExams (Exam topics of non experimental questions.)

Data are available under the terms of the Attribution 4.0 International (CC BY 4.0)

Acknowledgements
We thank Primoz Nerima for his help with the e-learning website and his precious, invaluable help in the data collection process.
References

1. Reis LO, Ikari O, Taha-Neto KA, et al.: Delivery of a urology online course using moodle versus didactic lectures methods. Int J Med Inform. 2015; 84(2): 149-54. PubMed Abstract | Publisher Full Text

2. Kopelman P: The future of UK medical education curriculum-what type of medical graduates do we need? J Med Educ. 2014; 1(1): 41-46. PubMed Abstract | Free Full Text

3. Irby DM, Cooke M, O'Brien BC: Call for reform of medical education by the Carnegie Foundation for the Advancement of teaching: 1910 and 2010. Acad Med. 2010; 85(1): 220-7. PubMed Abstract | Publisher Full Text

4. Skochelak SE, Stack SJ: Creating the Medical Schools of the Future. Acad Med. 2017; 92(1): 16-19. PubMed Abstract | Publisher Full Text

5. Masters K, Ellaway R: e-Learning in medical education Guide 32 Part 2: Technology, management and design. Med Teach. 2008; 30(5): 474-89. PubMed Abstract | Publisher Full Text

6. Smolle J: Virtual medical campus: the increasing importance of E-learning in medical education. GMS Z Med Ausbild. 2010; 27(2): Doc29. PubMed Abstract | Publisher Full Text | Free Full Text

7. Rose S: Medical Student Education in the Time of COVID-19. JAMA. 2020; 323(21): 2131-2132. PubMed Abstract | Publisher Full Text | Free Full Text

8. Wurth S, Sader J, Cerutti B, Meynard L, et al.: The pedagogical value of near-peer feedback in online OSCEs. BMC Med Educ. 2022; 22(1): 232. PubMed Abstract | Publisher Full Text | Free Full Text

9. Sader J, Cerutti B, Meynard L, et al.: Correction to: Medical students’ perceptions and coping strategies during the first wave of the COVID-19 pandemic: studies, clinical implication, and professional identity. BMC Med Educ. 2022; 22(1): 232. PubMed Abstract | Publisher Full Text | Free Full Text

10. Lavalard V, Cerutti B, Audetat-Voilard MC, et al.: Formative assessments during COVID-19 pandemic: an observational study on performance and experiences of medical students [version 1; peer review: 1 approved with reservations]. MedEdPublish (2016). 2023; 13: 7. PubMed Abstract | Publisher Full Text | Free Full Text

11. He L, Yang N, Xu L, et al.: Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences. Med Educ. 2021; 55: 293-308. PubMed Abstract | Publisher Full Text

12. Vaona A, Banzi R, Kwag KH, et al.: Cochrane Library Cochrane Database of Systematic Reviews E-learning for health professionals (Review). Cochrane Database Syst Rev. 2018. PubMed Abstract | Publisher Full Text

13. Kirkpatrick DR: Revisiting Kirkpatrick’s four-level-model. Training Dev. 1996; 1: 54-57.

14. Sadeghi R, Sedaghat MM, Sha Ahmadi F: Comparison of the effect of lecture and blended teaching methods on students’ learning and satisfaction. J Adv Med Educ Prof. 2014; 2(4): 146-50. PubMed Abstract | Free Full Text

15. Garrison DR, Kansaka H: Blended learning: Uncovering its transformative potential in higher education. Internet High Educ. 2004; 7(2): 95-105. PubMed Abstract | Publisher Full Text

16. Perkins GD, Fullerton JN, Davis-Gomez N, et al.: The effect of pre-course e-learning prior to advanced life support training: A randomised controlled trial. Resuscitation. 2010; 81(7): 877-81. PubMed Abstract | Publisher Full Text

17. Cerutti B, Blondin K, Galetto A: Long-menu questions in computer-based assessments: a retrospective observational study. BMC Med Educ. 2016; 16(1): 55. PubMed Abstract | Publisher Full Text | Free Full Text

18. Stollard F, Cerutti B, Aujesky S, et al.: Evaluation of a best practice approach to assess undergraduate clinical skills in Paediatrics. BMC Med Educ. 2020; 20(1): 46. PubMed Abstract | Publisher Full Text | Free Full Text

19. Lertwilaiwittaya P, Sitticharoon C, Maikaw P, et al.: Factors influencing the National License Examination step 1 score in preclinical medical students. Adv Physiol Educ. 2019; 43(3): 306-316. PubMed Abstract | Publisher Full Text | Free Full Text

20. Green ML, Moeller JJ, Spak JM: Text-enhanced learning in health professions education: A systematic review. BEME Guide No. 48. Med Teach. 2018; 40(4): 337-350. PubMed Abstract | Publisher Full Text

21. Wiskin CM, Allan TF, Skelton JR: Gender as a variable in the assessment of final year degree-level communication skills. Med Educ. Enhanced Reader, 2004; 38(2): 129-37. PubMed Abstract | Publisher Full Text | Free Full Text

22. Cuddy MM, Swanson DB, Dillon GS, et al.: A multilevel analysis of the relationships between selected examinee characteristics and United States Medical Licensing Examination Step 2 Clinical Knowledge performance: revisiting old findings and asking new questions. Acad Med. 2006; 81(10 Suppl): S103-7. PubMed Abstract | Publisher Full Text | Free Full Text

23. McDonough CM, Morgan A, Codd MIB, et al.: Gender differences in the results of the final medical examination at University College Dublin. Med Educ. 2000; 34(1): 30-34. PubMed Abstract | Publisher Full Text

24. Henderson JT, Weisman CS: Physician gender effects on preventive screening and counseling: an analysis of male and female patients’ health care experiences. Med Care. 2001; 39(12): 1281-1292. PubMed Abstract | Publisher Full Text

25. Sirwardena AN, Irish B, Asghar ZB, et al.: Comparing performance among male and female candidates in sex-specific clinical knowledge in the MRCP GP. Br J Gen Pract. 2012; 62(599): e464-50. PubMed Abstract | Publisher Full Text | Free Full Text

26. Biemstock J, Martin S, Tzou W, et al.: Medical students’ gender is a predictor of success in the obstetrics and gynecology basic clerkship. Teach Learn Med. 2002; 14(4): 240-243. PubMed Abstract | Publisher Full Text | Free Full Text

27. Pope L, Hawkridge A, Simpson R: Performance in the MRCPGP CSA by candidates’ gender: differences according to curriculum area. Educ Prim Care. 2014; 25(4): 186-193. PubMed Abstract | Publisher Full Text

28. Cuadrado-García M, Ruiz-Molina ME, Montoro-Pons JD: Are there gender differences in e-learning use and assessment? Evidence from an interuniversity online project in Europe. Procedia Soc Behav Sci. 2010; 2(2): 367-371. PubMed Abstract | Publisher Full Text | Free Full Text

29. Schiekirka S, Raupach T: A systematic review of factors influencing student ratings in undergraduate medical education course evaluations. BMC Med Educ. 2015; 15: 30. PubMed Abstract | Publisher Full Text | Free Full Text

30. Alsoufi A, Alsoufi A, Mohseni A, et al.: Impact of the COVID-19 pandemic on medical education: Medical students’ knowledge, attitudes, and practices regarding electronic learning. PLoS One. 2020; 15(11): e0242905. PubMed Abstract | Publisher Full Text | Free Full Text
Open Peer Review

Current Peer Review Status:  ?  ✓  ?  ✓

Version 2

Reviewer Report 11 June 2024

https://doi.org/10.21956/mep.21469.r35769

© 2024 Tombs M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Michal Tombs
Cardiff University, Cardiff, UK

Many thanks for providing such an extensive response and for making and highlighting all the necessary changes.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Evaluation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 28 May 2024

https://doi.org/10.21956/mep.21469.r36846

© 2024 Ergonul E. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Esin Ergonul
Department of Medical Education, Dokuz Eylul Universitesi, Izmir, İzmir, Turkey

Article summary
The article 'E-learning modules to improve clinical reasoning and practice: a prospective comparative study', by Fabiola Stollar et al, investigates the effectiveness of e-learning compared to traditional face-to-face teaching in improving medical students' clinical reasoning skills. The study, which was conducted over two academic years (2016-17 and 2017-18), involved 299 fourth-year medical students. The primary aim of the study was to assess the impact of an e-learning
module on students' application of knowledge and reasoning in the clinical setting. Secondary aims included assessing student satisfaction with the two learning methods and identifying factors influencing exam scores. The study concluded that there were no significant differences in clinical reasoning skills between students who used e-learning and those who attended traditional lectures. However, the number of quizzes taken and gender (female) were associated with higher scores. Students expressed satisfaction with both methods, but many reported that they felt they learned more from the e-learning modules.

Detailed checking
Is the work clearly and accurately presented and does it cite the current literature?
Yes: The paper is well written, clearly presented and extensively referenced to relevant and current literature. It provides a thorough review of existing studies and places the research within the broader context of medical education.

Is the study design appropriate and is the paper technically sound?:
Partially: The study design is generally appropriate, but there are areas for improvement:
Conceptual framework: While the revised version introduces the Kirkpatrick model, this model should be more fully integrated into the design and discussion of the study to provide a clearer theoretical basis.
Methodological details: More information is needed on the validity and reliability of the all assessment tools.

Are sufficient details of methods and analyses provided to allow replication by others?
Yes

Is the statistical analysis and its interpretation adequate?
Yes

Are any limitations of the research acknowledged?
Yes: The article acknowledges several limitations, including the single-faculty study setting, potential selection bias, lack of pre-assessment of baseline knowledge, and the small number of experimental items. It also mentions the difficulty of accurately measuring students' online time and the lack of authentic assessments of clinical reasoning.

Are all source data underlying the results available to ensure full reproducibility?
Yes.

Are the conclusions drawn adequately supported by the results?
Yes: The conclusions are well supported by the results. The study finds no significant differences in reasoning skills between e-learning and traditional methods, which is consistent with the data presented.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Have any limitations of the research been acknowledged?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Medical Education, Public Health

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 20 May 2024

https://doi.org/10.21956/mep.21469.r36847

© 2024 Olivares Olivares S. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Silvia Lizett Olivares Olivares
Tecnologico de Monterrey, Monterrey, Nuevo Leon, Mexico

The article compares learning results from different delivery methods: online vs. face-to-face. The information is clear and complete regarding the test and satisfaction questionnaire.

The content of the article has been discussed for many years in the field of education. The word practice in the title is misleading. I was expecting to have distance clinical practices with patients. In the section "Learning activities, Traditional learning" does include supervised actual clinical activities. However, during the online e-learning course, the program only includes hypothetical case-based scenarios.

I recommend clarifying how the online e-learning course included practices as the title describes.
Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Have any limitations of the research been acknowledged?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
No source data required

Are the conclusions drawn adequately supported by the results?
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Medical education

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.
The manuscript presents a comparative study between two teaching methods on clinical reasoning. The context in which it was conducted is well described, and the rationale for conducting the study was presented. This was in essence a quasi-experimental design, where the researchers compared exam performance and evaluation data (satisfaction) between two cohorts in two different years of study. The sample size was appropriate and the analyses accurate. However, certain aspects of the manuscript can be better improved.

To begin with, clarity and accuracy of the work can be developed further and this can be achieved by placing the study within a conceptual framework. To this end, it is important to provide a synthesis of the literature on the topic and clearly articulate in what way the research has the potential to add to the body of knowledge and to educational practice. A research question should also be provided and this should be driven by the literature and the identification of a problem or research gap / need. What is the evaluation framework that guided the study? e.g., Kirkpatrick?

In terms of methodology, more information on the evaluation instrument would have been useful, how it was developed and by whom. This will then provide opportunity for criticality in the discussion and the limitations of such evaluation methods.

The discussion is quite extensive but the results should not be reported here. Indeed, the reporting of linear regressions should be better reported and the rationale for using this statistical methods should be provided. Regressions are normally used when a prediction is made, but there are no predictions in this study. If the study is explorative in nature, a Pearson correlation would suffice, as no predictions were made regarding the direction or strength of the relationship between variables.

In summary, this study has the potential to contribute to the body of literature, but the manuscript requires some revisions.

**Is the work clearly and accurately presented and does it cite the current literature?**
Partly

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Have any limitations of the research been acknowledged?**
Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Partly
Are the conclusions drawn adequately supported by the results?
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Evaluation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 23 Nov 2023
Fabiola Stollar

Please find below all the responses, point by point, regarding the different comments that you have made. We also attach a copy of the manuscript revised version in which all the changes have been highlighted (track changes). Yours sincerely, Fabiola Stollar and the co-authors

○ “The discussion is quite extensive, but the results should not be reported here. Indeed, the reporting of linear regressions should be better reported and the rationale for using these statistical methods should be provided. Regressions are normally used when a prediction is made, but there are no predictions in this study. If the study is explorative in nature, a Pearson correlation would suffice, as no predictions were made regarding the direction or strength of the relationship between variables.”

Response: The estimated coefficients (included the corresponding estimated variabilities) and associated significance levels are reported in Table 1. For better exhaustivity we added the \( R^2 \) coefficient, and the Fisher statistic associated with the multiple regression in the caption of Table 1. Pearson correlation and the linear regression both rely exactly on the same underlying linear model. Computing a regression coefficient is nothing else but investigating the linear association between the two variables. Thank you for pointing out that the main objective of the study was not to do make predictions. We have modified the manuscript and replace “predictor” by “associated factor”.

○ “In terms of methodology, more information on the evaluation instrument would have been useful, how it was developed and by whom. This will then provide opportunity for criticality in the discussion and the limitations of such evaluation methods.”

Response: thank you for your comment. We completed the information required in the Methods section (page: 7).

○ “To begin with, clarity and accuracy of the work can be developed further and this can be achieved by placing the study within a conceptual framework. To this end, it is important to provide a synthesis of the literature on the topic and clearly articulate in what way the research has the potential to add to the body of knowledge and to educational practice. A research question should also be provided and this should be driven by the literature and the identification of a problem or research gap / need. What is the evaluation framework that guided the study? e.g., Kirkpatrick?”

Thank you very much for this very interesting proposal. Kirkpatrick framework concept is
indeed appropriate for analyzing the impact of two different teaching methods, which is the main objective of this study. This study analyzed level 1 of Kirkpatrick model through a satisfaction questionnaire sent to students, as well as level 2 through the evaluation of the success score of questions testing clinical case resolution. A description of the framework has been added to the introduction. Summaries of the current literature on the subject, highlighting the problem gap concerning the lack of studies comparing e-learning to traditional learning in the context of case-based competence, which is a core competence of medical education, are now described in the fourth paragraph of the introduction. The main research question was: Is there a difference between e-learning and traditional learning in terms of medical students' reasoning and how they apply their knowledge to clinical case resolution? This question is transcribed in the last paragraph of the introduction (page 5). We would also like to include this reference: Kirkpatrick, D. (1996). Revisiting Kirkpatrick's four-level-model. Training & Development, 1, 54-57.

**Competing Interests:** No competing interests were disclosed.

This was an interesting paper to read which seemed to show some convincing results about the efficacy of e-learning on the teaching of clinical reasoning. The authors make some valid points about the importance of integrating content and having e-learning modules which are well designed. The method used here included case scenarios in both the traditional component and e-learning modules and the modules, in the limited description provided, meet most of the key design principles of authentic and high quality e-learning. There is a vast literature on teaching clinical reasoning using e-learning but the problem is always with the assessment of this key skill.

This research showcases a good study design to further evaluate our knowledge of the usefulness of e-learning as a complementary activity to the standard teaching of clinical reasoning in small groups or at the bedside. More information is needed however to replicate this study and evaluate it's validity. For example, did all of your study cohorts agree or consent to be included in the research? What were the characteristics of each of the cohorts? Whilst steps were taken to ensure equivalent learning experiences by having the same staff provide the lectures as who wrote the content, I am wondering if there was a test effect. For example, could the students who completed the e-learning modules have performed better in the exam by virtue of their familiarity with the case study design used in their modules and experience with the self-assessments? The authors
have described using the Mixed Subjects results as a control variable, or factor, but this was not
evidenced in the reporting of the results. It would be useful to see some reporting of the
differences in the Control score, number of tests taken etc for each cohort, other than in the
regression. I am also wondering if there were any differences in the May vs January period, and
any differences based on the topic studied. I cannot comment on the efficacy of using regression
for this sample size but I am wondering if it was the most appropriate analysis to use.

I am also wondering about the possibility of using more authentic assessments in these types of
studies. Further mention could be made of the need to evaluate clinical reasoning development as
a measured outcome of specific teaching strategies using more authentic assessment design
principles. As yet there remains little research in relation to this.

Overall, this was an interesting read and contributes more knowledge to the area in a broad
sense, but more information is needed about the methodology and reporting to enhance this for
the reader.

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Partly

**Are sufficient details of methods and analysis provided to allow replication by others?**
No

**If applicable, is the statistical analysis and its interpretation appropriate?**
I cannot comment. A qualified statistician is required.

**Have any limitations of the research been acknowledged?**
Partly

**Are all the source data underlying the results available to ensure full reproducibility?**
Partly

**Are the conclusions drawn adequately supported by the results?**
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Problem based learning, medical education, evaluation

I confirm that I have read this submission and believe that I have an appropriate level of
expertise to confirm that it is of an acceptable scientific standard, however I have
significant reservations, as outlined above.

Author Response 23 Nov 2023
Please find below all the responses, point by point, regarding the different comments that you have made. We also attach a copy of the manuscript revised version in which all the changes have been highlighted (track changes). Yours sincerely, Fabiola Stollar and the co-authors

○ “The authors have described using the Mixed Subjects results as a control variable, or factor, but this was not evidenced in the reporting of the results. It would be useful to see some reporting of the differences in the Control score, number of tests taken etc for each cohort, other than in the regression. I am also wondering if there were any differences in the May vs January period, and any differences based on the topic studied. I cannot comment on the efficacy of using regression for this sample size but I am wondering if it was the most appropriate analysis to use.”

Response: The different covariates used in the analyses were all reported on Table 1. Thank you for highlighting the fact that the basic results (control score, number of tests taken) regarding both groups would be useful for the reader. These results were added at the beginning in the Results section. Globally speaking we have frequently observed a difference between the May and the January scores in our institution. Students taking the exam in January have gone through more clinical rotations and generally perform better, likely because their level of general knowledge and skills is slightly more advanced. The linear regression model is simple (which is particularly important because of the limited among of data in this study), usually well understood, and offers the advantage to allow one to introduce of several covariates (including potential confounding factors).

○ “More information is needed however to replicate this study and evaluate it’s validity. For example, did all your study cohorts agree or consent to be included in the research? What were the characteristics of each of the cohorts? Whilst steps were taken to ensure equivalent learning experiences by having the same staff provide the lectures as who wrote the content, I am wondering if there was a test effect. For example, could the students who completed the e-learning modules have performed better in the exam by virtue of their familiarity with the case study design used in their modules and experience with the self-assessments?”

Response: We have already reported in the Methods (page: 6) that “The participant consent was waived by the committee”. All students were well familiar with e-learning because from their baccalaureate they have access to e-learning. we added this phrase in the Methods (page: 6) “I am also wondering about the possibility of using more authentic assessments in these types of studies. Further mention could be made of the need to evaluate clinical reasoning development as a measured outcome of specific teaching strategies using more authentic assessment design principles. As yet there remains little research in relation to this.” Response: thank you for your comment, indeed it would be interested to evaluate it, but it was beyond the scope of our study. We included a phrase in the limitations of the study (page 13) : “it would have been interesting to evaluate clinical reasoning development as a measured outcome of specific teaching strategies using more authentic assessment design principles, but it was beyond the scope of our study specially because we evaluated pre-graduate students.

Competing Interests: No competing interests were disclosed.