

RESEARCH ARTICLE

Open Access



Evaluation of an international medical E-learning course with natural language processing and machine learning

Aditya Borakati

Abstract

Background: In the context of the ongoing pandemic, e-learning has become essential to maintain existing medical educational programmes. Evaluation of such courses has thus far been on a small scale at single institutions. Further, systematic appraisal of the large volume of qualitative feedback generated by massive online e-learning courses manually is time consuming. This study aimed to evaluate the impact of an e-learning course targeting medical students collaborating in an international cohort study, with semi-automated analysis of feedback using text mining and machine learning methods.

Method: This study was based on a multi-centre cohort study exploring gastrointestinal recovery following elective colorectal surgery. Collaborators were invited to complete a series of e-learning modules on key aspects of the study and complete a feedback questionnaire on the modules. Quantitative data were analysed using simple descriptive statistics. Qualitative data were analysed using text mining with most frequent words, sentiment analysis with the AFINN-111 and syuzhet lexicons and topic modelling using the Latent Dirichlet Allocation (LDA).

Results: One thousand six hundred and eleventh collaborators from 24 countries completed the e-learning course; 1396 (86.7%) were medical students; 1067 (66.2%) entered feedback. 1031 (96.6%) rated the quality of the course a 4/5 or higher (mean 4.56; SD 0.58). The mean sentiment score using the AFINN was + 1.54/5 (5: most positive; SD 1.19) and + 0.287/1 (1: most positive; SD 0.390) using syuzhet. LDA generated topics consolidated into the themes: (1) ease of use, (2) conciseness and (3) interactivity.

Conclusions: E-learning can have high user satisfaction for training investigators of clinical studies and medical students. Natural language processing may be beneficial in analysis of large scale educational courses.

Keywords: Methods, Research design, Machine learning, Education, Computer-assisted instruction

Background

E-learning is ubiquitous worldwide in both undergraduate [1, 2] and postgraduate medical education [3]. Diverse topics ranging from anatomy [4] to evidence based medicine [5] and clinical research training [6] may be delivered by e-learning. These courses may be delivered by single institutions to discrete cohorts or

asynchronously to any number of students worldwide as in Massively Open Online Courses (MOOCs [7]).

These courses have been shown to have similar student satisfaction to traditional, face-to-face instruction [5, 8–11], whilst being significantly cheaper to run [12, 13].

Further, in the context of the ongoing COVID-19 pandemic, with nationwide lockdowns and social distancing measures precluding face-to-face instruction in many cases, e-learning has become essential. Many medical

Correspondence: a.borakati@doctors.org.uk

University Department of Surgery, Royal Free Hospital, Pond Street, London NW3 2QG, UK



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

educational programmes have shifted to e-learning due to the pandemic [14, 15].

Evaluating large volumes of qualitative data that these large online courses generate is a further challenge. Natural language processing (NLP) and machine learning aggregation methods are common in commercial contexts on the internet.

The key methods in NLP are sentiment analysis and topic modelling. The former involves assigning each word in a body of text a numeric value, typically representing positivity or negativity based on pre-existing dictionaries (known as lexicons) which have already assigned every word a value for positivity or negativity. The values are averaged across the whole body of text to give an overall value of sentiment for the text. Topic modelling involves using machine learning algorithms to automatically group related words into relevant topics or themes.

These methods have been used in educational studies generally to gather overall sentiment [16–18] of bodies of free text and to generate overarching topics or themes from this [19]. However, these methods have yet to be utilised in medical education contexts [20, 21].

In this article, the experience of deploying a mandatory e-learning course for investigators (the majority of whom were medical students in clinical training) in an international multicentre collaborative cohort study, the IMAGINE study [22] is reported. This study aimed to evaluate the experience of course participants and identify key themes for future development of e-learning courses using NLP and machine learning methods and a guide is provided for other researchers to implement this in their own work.

Materials and methods

Research platform

The Ileus Management International (IMAGINE) study was a multinational collaborative study assessing gastrointestinal recovery following elective colorectal surgery.

Table 1 Summary of e-learning modules; modules are accessible at: <https://starsurg.org/imagine-e-learning/>

Module Name	Learning Objectives
IMAGINE: Included Patients and How to Find Them	<ul style="list-style-type: none"> ○ Know the inclusion and exclusion criteria for the IMAGINE project ○ Be able to apply them to patients for the IMAGINE project, correctly identifying which patients need including and excluding
Gastrointestinal Function	<ul style="list-style-type: none"> ○ Have an understanding of the GI-2 measure for gastrointestinal recovery ○ Know the components assessed for gastrointestinal function ○ Be able to find the relevant data for gastrointestinal function in IMAGINE
The Clavien-Dindo Classification	<ul style="list-style-type: none"> ○ Understand the Clavien-Dindo classification and its importance ○ Know the different grades of the Clavien-Dindo classification ○ Find the relevant information you need in the clinical setting to assign a Clavien-Dindo grade ○ Apply the Clavien-Dindo classification to patients post-operatively, giving them an accurate score
REDCap and Data Safety	<ul style="list-style-type: none"> ○ Be able to keep patient data safe and secure at your hospital site ○ Know how to login to the REDCap server for data entry to the IMAGINE project ○ Know how to enter data for the IMAGINE project

It took place between January and March 2018 at 424 hospitals across 24 countries in Europe and Australasia. Collaborators must have been a medical student in the clinical years of their course or a qualified doctor to participate [22, 23].

Module development

A series of four e-learning modules were developed for investigators in the study, covering four key phases of protocol implementation: (1) Study eligibility criteria; (2) Outcome assessment; (3) Data collection procedures; (4) Data security (Table 1). Content was agreed by the study steering group to ensure consistency with the final study protocol. Modules were developed using the H5P software (H5P, Joubel AS, Tromsø, Norway) on Wordpress 4.7.4 (Wordpress, Automattic Inc., San Francisco, CA, USA). The e-learning modules included multimedia, interactive content and multiple-choice questions based on clinical vignettes to assess learning. Learning objectives were written based on Bloom's taxonomy. Modules were designed to take less than 30 min to complete.

Modules were publicly accessible from a web browser and could be translated to the users' language using an embedded Google Translate widget (Google Translate, Alphabet Inc., Mountain View, CA, USA) from English [24]. Translated modules were reviewed by native speakers in the respective languages for accuracy; if there were significant inaccuracies, users were advised to complete the modules in English. A basic level of English language proficiency was required for participation in the study and all translations could be reverted into the original English by hovering over selected text to ensure consistency. A disclaimer was given to check the native English before completing the module.

Participant evaluation

After completion of the modules, investigators were invited to complete an anonymous, voluntary online feedback questionnaire (Google Forms, Alphabet Inc.,

Mountain View, CA, USA; Table 2, supplemental file 2 and [24]). This was a closed survey accessible only to course participants and was available at the end of course. Feedback was only allowed to be entered in English. Participation was voluntary with no incentivisation. Respondents were only permitted to complete one entry and this was enforced by evaluating the unique cookie for each user. Users were able to submit feedback for analysis at any time following completion of the course between October 2017 to June 2018. Revision of responses was not allowed.

The questionnaire contained both quantitative (Likert scale) and qualitative (free text) where investigators could comment on the quality and utility of the course and its individual modules. All items were displayed on one single page.

Feedback was periodically reviewed during the time period that the e-learning course was live and minor corrections (e.g. grammar, spelling) were made as appropriate in response. Satisfactory completion of the modules, with a 75% pass mark, was mandatory for participation in the study. 75% was chosen as a pragmatic threshold, which allowed a reasonable degree of understanding, whilst not being too arduous to complete. Participants could repeat each question as many times as they required to pass.

Analysis

Data was analysed using Microsoft R Open 3.5.3 for Windows (Microsoft Corporation, Redmond, WA, USA). Analysis scripts and a guide on implementing the methods discussed here is available in supplementary file 1 and online [25].

Distributions of continuous data was assessed by visual inspection of histograms, those following a normal distribution were presented using means with standard deviations or percentages. Non-normally distributed data were summarised as medians and interquartile range.

Qualitative analysis of the free text responses was conducted according to a content analysis framework, using NLP and machine learning techniques as described below.

Free text data was coded using a text mining method with the ‘tm’ R package [26]. Responses were merged into a single body of text (corpus) for each question and each word ranked by frequency. Common English ‘stop words,’ such as ‘the’ and ‘a’ were then removed using the ‘stopwords’ function in tm. The resulting corpus left only key adjectives for analysis of the responses. These are presented below and divided into positives and negatives based on the question asked in the questionnaire.

An unsupervised machine learning approach (Latent Dirichlet Allocation (LDA) algorithm with Gibbs sampling [27, 28]) was used to generate overarching topics in the free text. The number of topics to be generated was determined by calculating the metrics for each number of topics using four different formulae [29–32] and then finding the optimum number of topics by combined graphical analysis of these functions as described by Nikita [33]. Further evaluation of the topics generated by this algorithm is described in supplementary file 1 with an explanation of the metrics.

Sentiment analysis was conducted with all free text merged (without stop words) into one corpus. Each word was assigned a sentiment score using both the AFINN-111 [34] and syuzhet [35] lexicons, two commonly used general purpose lexicons which assign each word in the dictionary a rating of positivity or negativity. The mean score in each lexicon was calculated for the whole corpus to give a quantitative measure of the overall positivity or negativity of the free text feedback. Valence shifters in sentences e.g. ‘this was not good’ which may have the sentiments incorrectly calculated by the lexicons were accounted for as described in the sentimentr package [36].

Overall consolidated thematic analysis of the free text including recommendations for future international

Table 2 Feedback questionnaire items; *indicates mandatory question; questionnaire accessible at: <https://starsurg.org/imagine-e-learning/> (no longer taking responses)

Question	Response options
(1) How would you rate the e-learning overall?*	Very bad 1 2 3 4 5 Very good
(2) What was good about the e-learning overall?	Free text
(3) What could be improved about the e-learning overall?	Free text
(4) Any other comments about the e-learning overall?	Free text
(5) Add any specific comments about module 1 (Inclusion/Exclusion criteria) here:	Free text
(6) Add any specific comments about module 2 (Gastrointestinal Function) here:	Free text
(7) Add any specific comments about module 3 (Clavien-Dindo Classification) here:	Free text
(8) Add any specific comments about module 4 (REDCap and Data Protection) here:	Free text

research projects was then conducted by reviewing the results of the three analyses above, this was performed by one author (AB). The workflow for the qualitative analysis is shown in Fig. 1.

Ethics & Governance

Data was stored in accordance with European Union General Data Protection Regulations and users were informed that their anonymous feedback may be used in research and improvement of future courses as per the privacy policy as a condition of participation [37]. As such implicit informed consent was obtained from all subjects. No formal ethics committee approval was

sought for this study as only anonymised feedback data was collected.

This article is reported according to the SRQR guidelines for qualitative research [38] and the CHERRIES guidelines for online surveys, [39] endorsed by the EQUATOR network [40].

Results

Overall, 1611 collaborators attempted the e-learning course, of whom 1067 (66.2%) entered feedback. The demographics of the participants are summarised in Table 3: the majority (1396; 86.7%) were medical students; 200 (12.4%) were junior doctors, and 15 (0.9%) were consultant/attending surgeons.

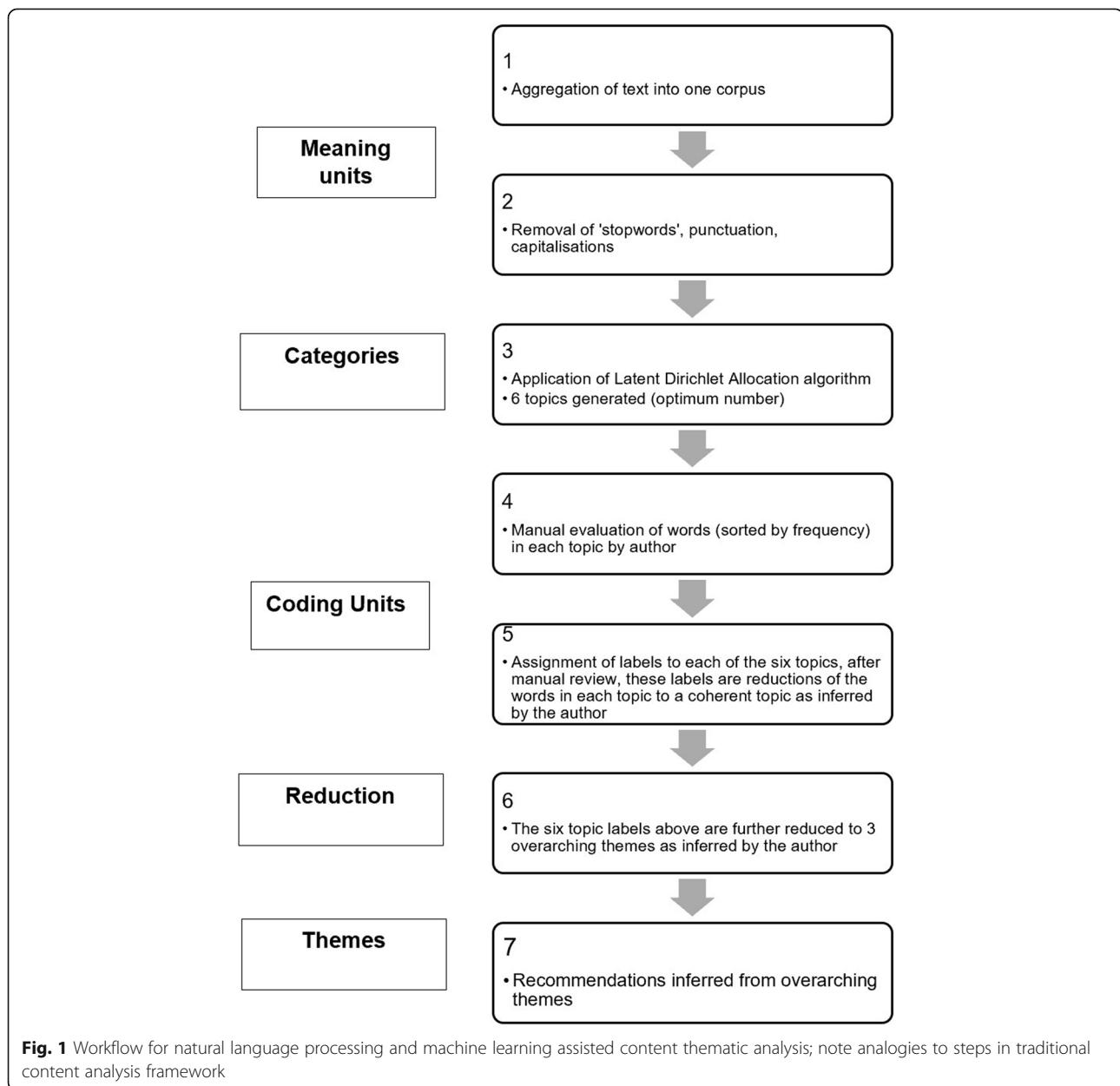


Table 3 Demographics of e-learning participants

Total number completing course (n)	1611
Medical students (%)	1396 (86.7)
Junior doctors (%)	200 (12.4)
Consultants (%)	15 (0.9)
Country	
United Kingdom (%)	918 (57.0)
Australia/ New Zealand (%)	119 (7.4)
Spain (%)	106 (6.6)
Italy (%)	64 (4.0)
Republic of Ireland (%)	62 (3.8)
Other (%)	342 (21.2)
Number completing feedback (%)	1067 (66.2)

Quantitative results

The mean of responses to the question ‘How would you rate the e-learning overall?’ was 4.56 out of 5 (Standard deviation (SD) 0.58). 637 (59.7%) respondents rated the course 5/5 (Fig. 2).

Qualitative feedback

The feedback entered for the question ‘What was good about the e-learning overall?’ is shown Table 4. The three most common words entered were ‘easy,’ ‘clear’ and ‘concise.’ ‘Easy,’ was typically followed with words to the effect of ‘easy to follow.’

Feedback entered for the question ‘What could be improved about the e-learning overall?’ is shown in Table 5. The most common words were ‘nothing,’

‘questions’ and ‘nil,’ with ‘questions’ typically being preceded by ‘more.’

The 8 most common words entered for the section ‘Any other comments about the e-learning overall’ are shown in Table 6. The three most common were ‘good,’ ‘none’ and ‘nil.’ All other words had less than 5 entries.

Topic modelling

The optimum number of topics for LDA was calculated to be 6. The most common relevant words in each topic are shown in Table 7 with the suggested themes: ‘ease,’ ‘assessment,’ ‘comprehension,’ ‘interactivity,’ ‘clarity’ and ‘summary.’

Sentiment analysis

The mean sentiment score using the AFINN lexicon was + 1.54/5 (SD 1.19) and + 0.287/1 (SD 0.390) using the syuzhet lexicon for all free text feedback. These both indicated overall positive feedback which is congruent with the most common words in the positive feedback and negative feedback questions (Tables 4 and 5) which many reported had nothing negative to feedback.

Thematic analysis

The overarching, consolidated themes identified were: (1) the ease of use of the modules, (2) succinctness and simplicity to understand the content, (3) interactivity with use of assessment and clinical vignettes.

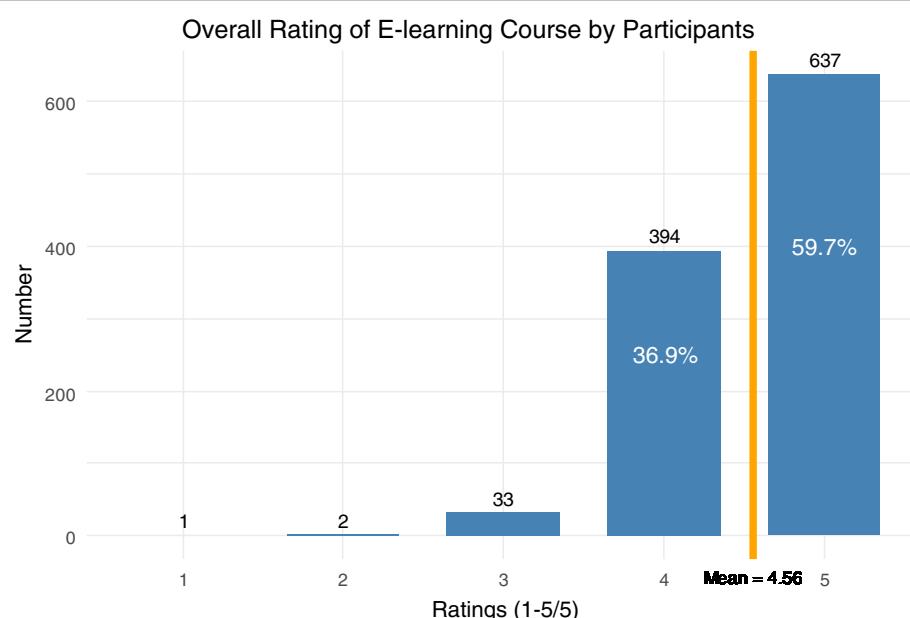


Fig. 2 Responses to question ‘How would you rate the e-learning overall?’, 1 = ‘Very Bad,’ 5 = ‘Very Good,’ n = 1067

Table 4 Frequency of top 20 words entered in response to question 'What was good about the e-learning overall?', n = 630 (59.0%)

Word	Frequency
Easy	125
Clear	93
Concise	79
Good	65
Information	58
Simple	55
Questions	45
Follow	40
Informative	39
Understand	37
Use	36
Interactive	30
Quick	28
Short	28
Well	23
Useful	20
Knowledge	19
Point	19
Cases	18

Table 5 Frequency of top 20 words entered in response to question 'What could be improved about the e-learning overall?', n = 426 (39.9%)

Word	Frequency
Nothing	68
Questions	37
Nil	17
Module	17
Redcap	15
Data	14
Videos	13
Information	12
Cases	11
Better	11
Examples	11
Can	11
Video	10
Spelling	10
Maybe	10
Good	10
None	8
Certificate	8
Test	7
Make	7

Reflections for application to future research

(1) Ease of use of the modules.

The modules were noted to be easy to use, with clear instructions on navigation and successful completion. Log in and registration processes were also mentioned to be simple with no email confirmation necessary.

The software used (H5P) further enabled straightforward navigation, accessibility and speed. Consequently, the modules generated were HTML5 packages, which allowed significantly better performance and mobile accessibility unlike many e-learning courses which depend on older Flash technologies. Navigation of the overall course was also fast with one single page for the whole course, reducing user frustration. In addition, course participants appreciated the ability to complete the course flexibly and return with their progress saved.

Recommendation: design of e-learning should consider the technologies used and prioritise user friendly and responsive software

(2) Succinctness and simplicity to understand the content.

The modules each had clear learning objectives based on Bloom's taxonomy [41]; text in each module was limited to small bullet points and relevant words highlighted in bold or in coloured boxes. Content was also varied with

the use of images, diagrams and videos to convey information.

The overall length of the course was short and appreciated by participants. Although the time taken to complete it was not tracked, many responses found that it took less than 30 min to complete. The ability to come back and complete the course in discrete sittings further enhanced this benefit.

Table 6 Frequency of top 8 words entered in response to question 'Any other comments about the e-learning overall', n = 243 (22.8%)

Word	Frequency
Good	32
None	16
Nil	12
Easy	8
Well	7
Great	6
Overall	5

Table 7 Selection of most common, relevant terms in topics generated from Latent Dirichlet Allocation topic modelling of combined free text feedback

Topics	1	2	3	4	5	6
Inferred Themes						
Ease	Assessment	Comprehension	Interactivity	Clarity	Summary	
Easy	Questions	Useful	Informative	Concise	Good	
Knowledge	Tests	Simple	Short	Information	Clear	
Information	Certificate	Examples	Interactive	Point	Quick	
Study	Test	Explained	Cases	Clear	Summary	
Clear	Quiz	Understand	Clinical	Succinct	Relevant	

Recommendation: concise modules with short, to the point sentences conveying key points are needed; the overall length of the course should be minimised
(3) Interactivity with use of assessment and clinical vignettes.

The interactive elements in the modules were positively received by participants. A range of elements were employed including true/ false questions, matching and multiple choice questions. These were interspersed throughout the modules but also at the end as a formal assessment.

The assessment questions were based around clinical vignettes that were likely to be encountered in the study as well as clinical practice. Collaborators found these helpful and practical for the study.

Recommendation: ensure modules contain interactive elements throughout; ensure content is relevant to practice, clinical vignettes are recommended

Discussion

This study showed that a massive online investigator training resource using e-learning was a successful, positively received modality for training collaborators, in particular, medical students, to deliver the protocol for this international study. The majority of these investigators were medical students and this shows the applicability of such courses to medical education in general.

This study was evaluated using online feedback questionnaires and analysed this large volume of data using natural language processing and machine learning techniques. To our knowledge, this is the only evaluation of an educational course in the medical field or in clinical studies to use such methods. This study demonstrates these methods as a successful proof of concept for evaluation of medical education courses, in particular Massively Open Online Courses (MOOCs).

In the context of the ongoing COVID-19 pandemic, where many universities are closed and plans are in place to re-start learning activities in virtual formats [42],

[43], these courses are of increasing importance for the foreseeable future as will virtual methods of evaluation.

Previous literature has established e-learning as having equal efficacy and satisfaction [1, 5, 8–10, 44] to face to face methods, at lower cost [45] to both under-[46] and post-graduates in medicine [11]. These courses have covered basic science topics such as anatomy [47], but also evidence-based medicine [11], clinical skills [48, 49] and simulation [50]. No literature is available evaluating e-learning for investigators in clinical studies.

Quality assurance of research studies is typically achieved by site initiation visits, study of detailed protocols and accreditation of investigators. As rapid, one-off events, it may be difficult for investigators to retain relevant information and apply protocols appropriately based on these methods. E-learning may provide a more accessible way of disseminating this information, which may be digested at a researcher's own pace and referred to on demand. These assurance visits are typically only performed for prospective interventional studies, here, the value for observational studies is shown as well.

Educational evaluation by machine learning methods

Machine learning methods have been implemented in other areas of education research [51] and are used commercially in search engines and social media [20, 52]. Data mining has been used in educational studies generally (not medical education) to gather overall sentiment [16–18] and topic modelling [19] as done in the present study, but also for 'student modelling' whereby students' predicted preferences for teaching and course outcomes are modelled [53–55]. In non-medical MOOCs, recent literature has emerged on the use of topic modelling and sentiment analyses on monitoring feedback and discussions [56–58]. Marking of essays and other free text coursework has also been demonstrated using similar machine learning techniques as in this study [59].

While there has been no or few similar approaches to analyse feedback in medical education, similar principles have been employed in the nascent field of virtual reality

simulation in medicine to analyse competence in procedures [60]. Hajshirmohammadi showed their fuzzy model could accurately classify experienced and novice surgeons in knot tying tasks on a virtual reality laparoscopic surgical simulator [61]. Megali [62] and Loukas [63] have further used hidden Markov models and multivariate autoregressive models for the same purpose with increasing success. These methods may pave way for more objective assessment of technical skills in clinical practice.

Strengths and limitations

Prior studies of e-learning in medicine have thus far have been on a small scale, with only a minority having greater than 100 participants; with inclusion of over 1000 responses, this study has been able to apply machine learning techniques to this larger natural language dataset. Through these machine learning algorithms it is demonstrated a method for reproducibility and objectivity in the analysis of free text data [64, 65]. This study is limited however, in not collecting data on demographic characteristics of feedback respondents. It is thus not possible to determine if respondents are representative of all collaborators who took part in the study, however, since a large majority responded it is unlikely sample was excessively skewed. Further, it is not possible to evaluate the effects of demographic factors such as gender, stage of training and country on feedback, which might otherwise facilitate iterative and targeted improvements in the resource content. Although collaborators were required to have a basic level of English proficiency to complete the e-learning course and free text feedback was mandated to be in English, it is possible that non-native speakers may have been dissuaded from entering feedback or refrain from entering as rich feedback as they might in their native language. This effect is also minimised somewhat as the majority of course participants were from English speaking countries.

Conventional qualitative methods such as interviews and focus groups may have yielded a greater depth of information and may have facilitated better exploration of ideas than electronic feedback, face to face visits may also be effective in building relationships, fostering morale and answering specific questions more effectively than e-learning. However, in the present context, these would be impractical due to the large scale of the study, current social distancing guidelines and the geographic distribution of participants. Future directions of research should evaluate e-learning as part of training for randomised controlled trials alongside more traditional methods. Further validation of natural language

processing and machine learning approaches to free-text data in medical education are needed.

Conclusion

E-learning can be a successful method for training participants of large-scale clinical studies and medical students, with high user satisfaction. Natural language processing approaches may be beneficial in evaluating medical education feedback as well as large scale educational programmes, where other methods are impractical and less reproducible.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12909-021-02609-8>.

Additional file 1.

Additional file 2.

Acknowledgements

I would like to thank James Glasbey and Stephen Chapman for critical review of the manuscript.

I would also like to thank the EuroSurg and STARSurg Collaboratives for delivering the study and disseminating the e-learning course.

The following contributed to content in the e-learning modules: Aditya Borakati, James Glasbey, Ross Goodson, Thomas Drake, Samuel Lee, Stephen Chapman.

Author's contributions

AB developed the modules, conceived the study, collected feedback, analysed the data and drafted the manuscript. The author(s) read and approved the final manuscript.

Funding

Funding for hosting of the e-learning course and associated website was made available by the British Journal of Surgery Society.

Availability of data and materials

The e-learning course and feedback questionnaire are available publicly at <https://starsurg.org/imagine-e-learning/>. The datasets used and/or analysed during the current study available from the corresponding author on reasonable request. Data analysis scripts and a guide on implementing the methods employed are available in supplementary file 1.

Declarations

Ethics approval and consent to participate

Not applicable, non-clinical subjects, participants agreed to Privacy Policy statement for participation, which stated anonymised data may be used for research and data used as per terms of EU GDPR.

Consent for publication

Not applicable as above.

Competing interests

The author declares no conflicts of interest.

Received: 25 November 2020 Accepted: 16 February 2021

Published online: 25 March 2021

References

1. Zehry K, Halder N, Theodosiou L. E-Learning in medical education in the United Kingdom. Procedia Soc Behav Sci. 2011;15:3163–7 10/bk6pwt.
2. Frehywot S, Vovides Y, Talib Z, Mikhail N, Ross H, Wohltjen H, Bedada S, Korhumel K, Koumare AK, Scott J. E-learning in medical education in

- resource constrained low- and middle-income countries. *Hum Resour Health.* 2013;11(1):4. <https://doi.org/10.1186/1478-4491-11-4>.
3. Wittich CM, Agrawal A, Cook DA, Halvorsen AJ, Mandrekar JN, Chaudhry S, Dupras DM, Oxentenko AS, Beckman TJ. E-learning in graduate medical education: survey of residency program directors. *BMC Med Educ.* 2017; 17(1):114. <https://doi.org/10.1186/s12909-017-0953-9>.
 4. Trelease RB. Essential E-learning and M-learning methods for teaching anatomy. In: Chan LK, Pawlina W, editors. *Teaching anatomy: a practical guide.* Cham: Springer International Publishing; 2020. p. 313–24. https://doi.org/10.1007/978-3-030-43283-6_32.
 5. Kulier R, Coppus SF, Zamora J, Hadley J, Malick S, Das K, et al. The effectiveness of a clinically integrated e-learning course in evidence-based medicine: A cluster randomised controlled trial. *BMC Med Educ.* 2009;9:21 10/b93m8f.
 6. Knatterud GL, Rockhold FW, George SL, Barton FB, Davis CE, Fairweather WR, Honohan T, Mowery R, O'Neill R. Guidelines for quality assurance in multicenter trials: a position paper. *Control Clin Trials.* 1998;19(5):477–93. [https://doi.org/10.1016/S0197-2456\(98\)00033-6](https://doi.org/10.1016/S0197-2456(98)00033-6).
 7. Doherty I, Sharma N, Harbutt D. Contemporary and future eLearning trends in medical education. *Med Teach.* 2015;37(1):1–3. <https://doi.org/10.3109/0142159X.2014.947925>.
 8. Kulier R, Hadley J, Weinbrenner S, Meyerrose B, Decsi T, Horvath AR, et al. Harmonising Evidence-based medicine teaching: a study of the outcomes of e-learning in five European countries. *BMC Med Educ.* 2008;8:27 10/b8xbzz.
 9. Kulier R, Gürmezoglu AM, Zamora J, Plana MN, Carroli G, Cecatti JG, et al. Effectiveness of a Clinically Integrated e-Learning Course in Evidence-Based Medicine for Reproductive Health Training: A Randomized Trial. *JAMA.* 2012; 308:2218 10/f4fpp5.
 10. Cahill D, Cook J, Sithers A, Edwards J, Jenkins J, et al. *Med Teach.* 2002;24: 425–8 10/ff4vz4.
 11. Hadley J, Kulier R, Zamora J, Coppus SF, Weinbrenner S, Meyerrose B, et al. Effectiveness of an e-learning course in evidence-based medicine for foundation (internship) training. *J R Soc Med.* 2010;103:288–94 10/c6npw9.
 12. Walsh K, Rutherford A, Richardson J, Moore P. NICE medical education modules: an analysis of costeffectiveness. *Educ Prim Care Off Publ Assoc Course Organ Natl Assoc GP Tutors World Organ Fam Dr.* 2010;21:396–8.
 13. Maloney S, Nicklen P, Rivers G, Foo J, Ooi YY, Reeves S, Walsh K, Ilic D. A cost-effectiveness analysis of blended versus face-to-face delivery of evidence-based medicine to medical students. *J Med Internet Res.* 2015; 17(7):e182. <https://doi.org/10.2196/jmir.4346>.
 14. Rose S. Medical student education in the time of COVID-19. *JAMA.* 2020; 323(21):2131–2. <https://doi.org/10.1001/jama.2020.5227>.
 15. Hilburg R, Patel N, Ambruso S, Biewald MA, Farouk SS. Medical education during the coronavirus Disease-2019 pandemic: learning from a distance. *Adv Chronic Kidney Dis.* 2020;27(5):412–7. <https://doi.org/10.1053/j.ackd.2020.05017>.
 16. Merceron A, Yacef K. Educational data mining: a case study. In: Proceedings of the 2005 conference on artificial intelligence in education: supporting learning through intelligent and socially informed technology. Amsterdam: IOS Press; 2005. p. 467–74.
 17. Merceron A, Yacef K. Tada-ed for educational data mining. *Interact Multimed Electron J Comput-Enhanc Learn.* 2005;7:267–87.
 18. Ranjan J, Malik K. Effective educational process: a data-mining approach. *VINE.* 2007;37:502–15 10/dwqx2.
 19. Colace F, De Santo M, Greco L. Safe: a sentiment analysis framework for e-learning. *Int J Emerg Technol Learn IJET.* 2014;9:37 10/ggfhz2.
 20. Boyan J, Freitag D, Joachims T. A machine learning architecture for optimizing web search engines. *Proc AAAI Workshop Internet-Based Inf Syst.* 1996;8.
 21. Hu X, Liu H. Text analytics in social media. In: Aggarwal CC, Zhai C, editors. *Mining text data.* Boston: Springer US; 2012. p. 385–414. https://doi.org/10.1007/978-1-4614-3223-4_12.
 22. Chapman SJ, Collaborative ES. Ileus management international (IMAGINE): protocol for a multicentre, observational study of ileus after colorectal surgery. *Colorectal Dis Off J Assoc Coloproctology G B Irel.* 2018;20:017–25.
 23. Jang H, Kim KJ. Use of online clinical videos for clinical skills training for medical students: benefits and challenges. *BMC Med Educ.* 2014;14:56. <https://doi.org/10.1186/1472-6920-14-56>.
 24. IMAGINE E-Learning – STARSurg. <https://starsurg.org/imagine-e-learning/>. Accessed 17 Nov 2019.
 25. Borakati A. Supplementary file 1 - analysis and guide to implement methods. 2019. <https://aborakati.github.io/E-learning-Analysis/>. Accessed 18 Sep 2020.
 26. Feinerer I, Hornik K, Meyer D. Text mining infrastructure in R. *J Stat Softw.* 2008;25:1–54 <http://www.jstatsoft.org/v25/i05>.
 27. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. *J Mach Learn Res.* 2003;3:993–1022 <https://jmlr.csail.mit.edu/papers/v3/blei03a.html>. Accessed 18 Sep 2020.
 28. Grün B, Hornik K. Topicmodels: an R package for fitting topic models. *J Stat Softw.* 2011;40:1–30.
 29. Griffiths TL, Steyvers M. Finding scientific topics. *Proc Natl Acad Sci.* 2004; 101(Supplement 1):5228–35 10/adf.
 30. Cao J, Xia T, Li J, Zhang Y, Tang S. A density-based method for adaptive LDA model selection. *Neurocomputing.* 2009;72:1775–81 10/dvh9dh.
 31. Arun R, Suresh V, Veni Madhavan CE, Narasimha Murthy MN. On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In: *Advances in Knowledge Discovery and Data Mining.* Berlin Heidelberg: Springer; 2010. p. 391–402. https://doi.org/10.1007/978-3-642-13657-3_43.
 32. Devaeud R, SanJuan E, Bellot P. Accurate and effective latent concept modeling for ad hoc information retrieval. *Doc Numér.* 2014;17:61–84 10/gbbjh.
 33. Nikita M. Imdatuning: Tuning of the Latent Dirichlet Allocation Models Parameters. 2016. <https://CRAN.R-project.org/package=Idatuning>.
 34. Nielsen FÅ. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *ArXiv11032903 Cs.* 2011. <http://arxiv.org/abs/1103.2903>. Accessed 27 Nov 2019.
 35. Jockers ML. Syuzhet: extract sentiment and plot arcs from text. 2015. <https://github.com/mjockers/syuzhet>.
 36. Rinker TW. Sentiment. Buffalo: Calculate Text Polarity Sentiment; 2019. <http://github.com/trinker/sentimentr>
 37. Borakati A. Privacy Policy – STARSurg. 2018. <https://starsurg.org/privacy-policy/>. Accessed 19 Dec 2019.
 38. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for Reporting Qualitative Research: A Synthesis of Recommendations. *Acad Med.* 2014;89:1245–10/f6hg5t.
 39. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERES). *J Med Internet Res.* 2004; 6(3):e34.
 40. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) | The EQUATOR Network. <https://www.equator-network.org/reporting-guidelines/improving-the-quality-of-web-surveys-the-checklist-for-reporting-results-of-internet-e-surveys-cherries/>. Accessed 17 Dec 2018.
 41. Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, et al. A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives, complete edition. 2Rev Ed edition. New York: Pearson; 2000.
 42. Murphy MPA. COVID-19 and emergency eLearning: consequences of the securitization of higher education for post-pandemic pedagogy. *Contemp Secur Policy.* 2020;41(3):492–505. <https://doi.org/10.1080/13523260.2020.1761749>.
 43. Almarzooq ZI, Lopes M, Kochar A. Virtual learning during the COVID-19 pandemic: a disruptive Technology in Graduate Medical Education. *J Am Coll Cardiol.* 2020;75(20):2635–8. <https://doi.org/10.1016/j.jacc.2020.04.015>.
 44. Gagnon M-P, Legare F, Labrecque M, Fremont P, Cauchon M, Desmarais M. Perceived barriers to completing an e-learning program on evidence-based medicine. *J Innov Health Inform.* 2007;15:83–91 10/ggfhppt.
 45. Gibbons AS, Fairweather PG. Computer-based instruction: design and development: Educational Technology; 1998.
 46. Khogali SEO, Davies DA, Donnan PT, Gray A, Harden RM, McDonald J, et al. Integration of e-learning resources into a medical school curriculum. *Med Teach.* 2011;33:311–8 10/cx7dqr.
 47. Svirko E, Mellanby DJ. Attitudes to e-learning, learning style and achievement in learning neuroanatomy by medical students. *Med Teach.* 2008;30:e219–27 10/dmsfbz.
 48. Jang HW, Kim K-J. Use of online clinical videos for clinical skills training for medical students: benefits and challenges. *BMC Med Educ.* 2014;14:56 10/f5xp7r.
 49. Bloomfield JG, Jones A. Using e-learning to support clinical skills acquisition: Exploring the experiences and perceptions of graduate first-year pre-

- registration nursing students — A mixed method study. *Nurse Educ Today*. 2013;33:1605–11 10/f5hb5.
50. Smolle J, Praise G, Smolle-Jüttner F-M. Emergency treatment of chest trauma — an e-learning simulation model for undergraduate medical students. *Eur J Cardiothorac Surg*. 2007;32:644–7 10/bts5qp.
 51. Al-Shammari I, Aldhafiri M, Al-Shammari Z. A meta-analysis of educational data mining on improvements in learning outcomes. 2013. <https://www.ingentaconnect.com/content/prin/csj/2013/00000047/00000002/art00008>. Accessed 15 Dec 2019.
 52. Hu X, Liu H. Text analytics in social media. In: Mining Text Data; 2012.
 53. Hung J, Zhang K. Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching. *MERLOT J Online Learn Teach*. 2008; https://scholarworks.boisestate.edu/edtech_facpubs/4.
 54. García E, Romero C, Ventura S, Gea M, de Castro C. Collaborative data mining tool for education. In: International Working Group on Educational Data Mining; 2009. <https://eric.ed.gov/?id=ED539090>. Accessed 15 Dec 2019.
 55. Kotsiantis SB. Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. *Artif Intell Rev*. 2012;37:331–44 10/fmbng4.
 56. Chen Z. In: Sun X, Pan Z, Bertino E, editors. Automatic self-feedback for the studying effect of MOOC based on support vector machine. Cham: Springer International Publishing; 2019. https://doi.org/10.1007/978-3-030-24265-7_27.
 57. Liu Z, Zhang W, Sun J, Cheng HNH, Peng X, Liu S. Emotion and Associated Topic Detection for Course Comments in a MOOC Platform. In: 2016 International Conference on Educational Innovation through Technology (EITT); 2016. p. 15–9.
 58. Hew KF, Hu X, Qiao C, Tang Y. What predicts student satisfaction with MOOCs: a gradient boosting trees supervised machine learning and sentiment analysis approach. *Comput Educ*. 2020;145:103724. <https://doi.org/10.1016/j.comedu.2019.103724>.
 59. Altoe F, Joyner D. Annotation-free automatic examination essay feedback generation. In: 2019 IEEE learning with MOOCs (LWMOOCs). Milwaukee: IEEE; 2019. p. 110–5.
 60. Winkler-Schwartz A, Bissonnette V, Mirchi N, Ponnudurai N, Yilmaz R, Ledwos N, et al. Artificial Intelligence in Medical Education: Best Practices Using Machine Learning to Assess Surgical Expertise in Virtual Reality Simulation. *J Surg Educ*. 2019;76:1681–90 10/ggm3v7.
 61. Hajshirmohammadi I, Payandeh S. Fuzzy Set Theory for Performance Evaluation in a Surgical Simulator; 2007. <https://doi.org/10.1162/pres.16.6.603>.
 62. Megali G, Sinigaglia S, Tonet O, Dario P. Modelling and evaluation of surgical performance using hidden Markov models. *IEEE Trans Biomed Eng*. 2006;53(10):1911–9. <https://doi.org/10.1109/TBME.2006.881784>.
 63. Loukas C, Georgiou E. Multivariate autoregressive modeling of hand kinematics for laparoscopic skills assessment of surgical trainees. *IEEE Trans Biomed Eng*. 2011;58(11):3289–97. <https://doi.org/10.1109/TBME.2011.2167324>.
 64. Eickhoff M, Wienke R. Understanding topic models in context: a mixed-methods approach to the meaningful analysis of large document collections, vol. 10; 2018.
 65. Piepenbrink A, Gaur AS. Topic models as a novel approach to identify themes in content analysis. *Acad Manag Proc*. 2017;2017:11335 10/ggndf2.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

