DETERMINING THE EFFECTIVENESS OF A MASSIVE OPEN ONLINE COURSE IN DATA SCIENCE FOR HEALTH

Abrar Alturkistani
Imperial College London
Global eHealth Unit, Department of Public Health and Primary Care, School of Public Health, London, United Kingdom

Josip Car
Imperial College London
Global eHealth Unit, Department of Public Health and Primary Care, School of Public Health, London, United Kingdom

Azeem Majeed
Imperial College London
Department of Public Health and Primary Care, School of Public Health, London, United Kingdom

David Brindley
University of Oxford
Department of Pediatrics, Medical Sciences Division,, Oxford, United Kingdom

Glenn Wells
Oxford Academic Health Science Centre
Oxford, United Kingdom

Edward Meinert
University of Oxford
Department of Pediatrics, Medical Sciences Division, Oxford, United Kingdom

ABSTRACT
Massive Open Online Courses (MOOCs) are widely used to deliver specialized education and training in different fields. Determining the effectiveness of these courses is an integral part of delivering comprehensive, high-quality learning. This study is an evaluation of a MOOC offered by Imperial College London in collaboration with Health iQ called, Data Science Essentials: Real World Evidence. The paper analyzes the reported learning outcomes, attitudes and behaviours of students after completing the MOOC. The study used mixed-methods, drawing from a Kirkpatrick evaluation-using data from semi-structured interviews transcribed and analyzed through Braun and Clark's method for thematic coding. 191 learners joined the MOOC. Two participants who completed at least 75% of the course were interviewed for the course evaluation. The findings of the analysis suggest that the course attracted target learners and learners found its application and engagement methods effective. Learners found the training provided by the MOOC to be helpful and with the potential to be applied in their work environment in the future and identified some work-related barriers that prevent knowledge application. Networking during and post-MOOC was identified as an area that needs improvement and development in the future. Findings derived from this evaluation support the fact that generally, MOOCs can improve learning and knowledge attainment in practical skills-based knowledge. One implication of this study is to inform factors that engage learners in the design and implementation of MOOC. The findings have shown that factors that affect the learners' engagement are the availability of lecture videos, self-assessment tools and high networking and communication between learners. In terms of knowledge application, support and availability of the right resources are essential because learners are not able to apply learning in their workplace if the workplace lacked the right resources and support. Developers of MOOCs for continuing professional development should take into consideration work-related barriers when designing their MOOCs.

KEYWORDS
Massive Open Online Course (MOOC), e-learning, Qualitative analysis, Continuing professional development (CPD)

1. INTRODUCTION

Evaluating Massive Open Online Courses (MOOCs) can help appraise their effectiveness and improve utilization (Chapman et al., 2016). There is a need for more evidence analyzing the impact of MOOCs on learners' knowledge, skills and attitudes (Khalil, 2014). Although some studies have found that MOOCs have the potential to foster student autonomy and create learning communities conducive to the learning process (Goldie, 2016), other research suggests significant issues in MOOC efficacy particularly due to factors including 1) Dropout rates; on average less than 10% of learners who signup actually make it to course completion (Khalil, 2014), 2) Social connections between learners not being a universal occurrence, raising questions about openness and diversity, 3) The need for the "social presence" of course facilitators to not only stimulate but maintain active learner participation (Goldie, 2016). More evidence is required to better understand how MOOCs can be used to address these factors to encourage higher rates of engagement.

The MOOC: Data Science Essentials: Real World Evidence (RWE) was offered by Imperial College and Health IQ through the online learning platform: GOMO and was designed by Imperial College. The MOOC lasted for five weeks and was offered twice. The 4-weeks course was available for free to all learners and the fifth week of the MOOC was exclusively available to participants who signed up for certification. To investigate the success of the MOOC in reaching its aim, this evaluation was conducted to better understand what impact the course had on further use of the skills taught in the course.

- Primary research question:
    - How has the course impacted the learners' knowledge, skills and attitudes on the use of data science in healthcare?
- Primary research question:
    - What evidence is there that the intended target audience was reached?
    - What evidence is there that the MOOC has made a difference to participants in their work or studies?
    - What evidence is there of participant networks for data science in healthcare being adopted during the MOOC?
    - What evidence is there that the MOOC format and materials engaged participants?
    - What evidence is there of participant networks for data science in healthcare being sustained post MOOC?

1.1 About the MOOC

There is an increased demand for increasing healthcare professionals' training and skills in data science and use of Information and Communication Technologies (ICT) (Gallagher, 2015). MOOCs are being used to teach professionals in the healthcare field new skills (Hossain et al., 2015) and can be used for continued professional development for healthcare professionals. Despite data analysis opportunities in healthcare, which can improve its effectiveness and efficiency, this is an area that requires continuous skill development because of the rapid changes of methods and data sources available. Using population-level big data, collected through the various activities that occur in a healthcare system can make it possible to create models that can predict disease, enable better preventive measures and create more personalised care for patients (Raghupathi and Raghupathi, 2014). This MOOC aimed to introduce students the impact data science can have on medicine and inspire the application of these methods across various undergraduate curriculum disciplines, NHS commissioning support organisations, healthcare regulation organisations and life sciences industries (Imperial College London, 2017). The MOOC was offered twice, in August and in October, with each MOOC lasting for five weeks. A total of 191 learners have joined both MOOCs, 135 of them from the August cohort and 56 of them from the October cohort. 11 learning outcomes were formulated to meet the aims of the MOOC (Table 1).

Table 1. MOOC learning outcomes

Order	Learning Outcome
1	Acquire knowledge in the fundamentals of Real World Evidence (RWE) to include definition and background, current RWE trends and themes, benefits and limitations of RWE, and its place today in organisations dealing in patient care/data.
2	Acquire knowledge of information governance requirements and policy with regard to patient data as well as knowledge of key datasets that RWE can exploit across primary and secondary care (HES/CPRD).
3	Understand the difference between what Real World is and what is not.
4	Understand the essential theory of using RWE with data science, and key differences between using RWE with and without data science.
5	Understand the different data investigation tasks and the most appropriate algorithms for selecting/addressing them.
6	Identify and apply appropriate data analytic techniques to a problem using an RWE framework (decision tree) further to practical group sessions thereby demonstrating an understanding of knowledge gained.
7	Carry out exploratory analysis of Real World Data (RWD) (structured data).
8	Evaluate RWD, models or algorithms for accuracy in order to make an informed decision with regard to their use.
9	Conceptualise a [data mining solution] to a practical problem through teamwork and collaboration.
10	Critique the results of a [data mining] exercise and the pitfalls of analysing RWD.
11	Develop hypotheses based on the analysis of the results obtained and test them.

2. METHODS

The Kirkpatrick Model was used to evaluate the effectiveness of training to impact professional practice (Kirkpatrick and Kirkpatrick, 2006). The model assesses training through four levels. Level 1 Reaction; assesses participants' response to the training. Level 2 Learning; assesses participants learning from training. Level 3 Behaviour; assesses participants' use of training in their job and Level 4 Results; evaluates the impact of training on the organization. The reason that the Kirkpatrick model was selected as an evaluation model was due to its suitability for supporting professional development training and its approach to measuring behaviour following a three to six month time interval post training. As sufficient time had passed to investigate these outcomes, this model was well suited. To address the four levels of the Kirkpatrick evaluation method, data were collected from participants who joined and participated in the MOOC. All participants who were registered in the MOOC were recruited to be interviewed for the evaluation. 7 participants volunteered to be interviewed, 5 who subsequently declined or did not respond to interview invitations. The remaining two participants participated in the one-on-one semi-structured interviews. The participants interviewed for this study were employed adults, one male and one female. Only one participant

was a part-time postgraduate student studying a healthcare related topic at the time of the interview. Both participants worked in the healthcare field and used data science in their work. One participant had a MOOC completion rate of 100% while the other had a completion rate of 75%. Data from the two participants were evaluated using thematic analysis and Kirkpatrick evaluation methods. Interviews were conducted through conference calls. Questions about each level of the Kirkpatrick evaluation were incorporated into the interview questions. Kirkpatrick evaluation was completed for each interview data separately, and results from both analyses were concluded in a single report and summarized in this evaluation.

2.1 Data collection

Data for the evaluation was collected through semi-structured interviews. The interview questions included questions about learners' background, reasons for joining the MOOC, their use of the information in their workplace, participant's interaction with other learners and what participants liked or disliked about the MOOC.

2.2 Data analysis

Interview recordings were transcribed and anonymized. Thematic analysis was performed using Braun and Clarke's framework for thematic data analysis. Thematic analysis method allowed for aggregation of responses and provided the ability to perform an in-depth investigation of learner perspectives on the question categories and research questions. This analysis was performed through 6 steps: familiarisation with data, generation of initial codes, searching for themes, reviewing themes, defining and naming themes, and producing a report (Clarke and Braun, 2013). After defining and naming the themes, a thematic map was formulated to review themes and show the relationships between them (Daley, 2004). The Kirkpatrick evaluation followed thematic analysis to enhance analysis reliability by using more than one method for data analysis (Patton, 1999).

3. RESULTS

3.1 Thematic analysis results

Thematic analysis of the interview data gave rise to three main themes 'learner background', 'MOOC learning' and 'MOOC' features (Figure 1). Themes were formulated inductively using interview excerpts. In this section, in-depth results for each theme are discussed using supporting interview quotes.

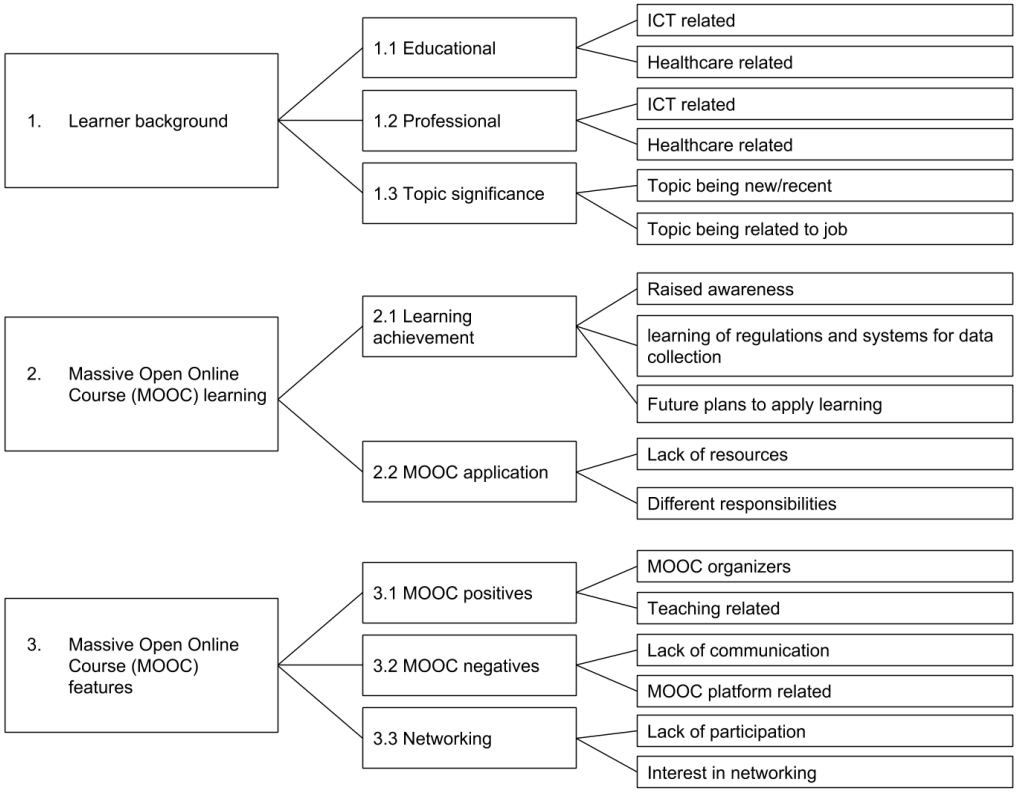


Figure 1. Figure showing themes, sub-themes and codes that resulted from the thematic analysis of the interview data.

3.1.1 Theme 1: Learner background

The learner background theme (Figure 1) represented participants' educational and professional experience showing that both participants worked in the healthcare sector and were involved with data analysis. The codes ICT related and healthcare related (Figure 1) represented learners educational and professional backgrounds that were relevant to the topic of the MOOC: data science skills for healthcare. Both participants were exposed to data analysis through their educational and professional backgrounds. For example, one participant was studying Masters in Economic Evaluation in Healthcare and the other had completed Masters in Biostatistics. Moreover, both participants were involved in healthcare-related jobs. The topic significance subtheme was created to represent participants' interest in the topic of the MOOC. It was made clear that participants joined the MOOC for its relevance to their jobs, and for improving job-related skills: "The reason I joined this course is because I anticipated that having, being equipped with this knowledge put me in a better position within my job" (Participant 1).

3.1.2 Theme 2: MOOC learning

The MOOC learning theme (Figure 1) represented participants gained learning and application of the learning in their workplace. The raised awareness subtheme was concluded from participants' comments about having better awareness after the MOOC: "it definitely made me more conscious…" (Participant 1) and "There are many many sources of datasets I didn't know they existed" (Participant 2). In terms of

knowledge application, both participants strongly believed that they would be using the knowledge in the future and both mentioned work-related barriers that prevented them from applying learning from the MOOC. For example, one participant mentioned the lack of data sources as a barrier for using the skills learned through the MOOC: “currently in my current role, we, unfortunately, don’t have data source, but we are planning to discover and to develop some, but I should be, I am assigned to do this, but we did not start yet” (Participant 2). Learning of regulations and systems for data collection subtheme was created because these were key topics delivered through the course, and both participants have demonstrated that they have learned them thoroughly from the MOOC. One participant has explained this by emphasizing the importance of disseminating the data properly to be as representative of the original data as possible: “is important to maintain the, well to improve the data integrity as much as possible during the data collection, because data collection is such a laborious process and there is a high chance that if you don't implement the right systems, that you’re going to get messing the data because people who are collecting the data, don’t know what you need or they are not properly trained, so I do feel like it did help me” (Participant 1). The other participant has emphasized the importance of the systems explained through the course for data collection and analysis: “actually and just it puts the whole process into perspective, in a system, like now I know that there is a system existing for payroll data and pharmaceutical academic collaboration.” (Participant 2). Despite not being able to apply learning from the MOOC in their workplace, both participants were confident the learning will be put to use in the future. For instance, one participant mentioned that the resources they learned about in the course will be of great use in the future: “I'm sure I will get back to them one day” (Participant 2). Also, the same participant has added about the regulations taught in the MOOC: “...I believe, they will give, they are a very good example of the existing regulations, and also different resources and sources of datasets, I believe this will be very helpful” (Participant 2).

3.1.3 Theme 3: MOOC features

The MOOC features theme (Figure 1) represented the positives and negatives participants mentioned about the MOOC. In terms of positives, participants liked the videos and assessments provided throughout the course: “The videos were the most engaging. I like both. I like the videos and the articles, but the videos were more engaging for me. They are easier to follow maybe” (Participant 2) “I like the questions throughout the lesson because it does test you, whether you’re actually concentrating or flicking through the MOOC. Yeah, I definitely appreciate that part. yeah, because you didn’t have to pay for that even if, you wanted to do it, it was just like a final assessment that you didn’t have to pay for. yeah, it just tested to see whether you were concentrating” (Participant 1). In terms of MOOC negatives, lack of communication and inactivity in the course’s social media page was seen as a negative, because both participants were looking to communicate with other learners. Other negatives mentioned were about MOOC platform features such as pausing videos, or downloading videos for offline viewing, both features that were not available, and seen as a shortfall by participants. Networking is an important part of most MOOCs (Liyanagunawardena, Adams and Williams, 2013). Networking was an important part of this MOOC as well, for a social media page was created to increase socializing and networking among participants. Course coordinators posted questions on the social media page to encourage discussions and joining of networks of health science between learners. However, through the evaluation interviews, it was discovered that there was a lack of participation and networking through the social media page of the course. Participants attributed this to the social media page being inactive and lacking participation from other learners: “ when I first started the course, I think it was like the first two weeks, so I looked at the hashtag, which encourages conversation on Twitter, but I did notice that there wasn’t that much going on, I guess because there were so few people actually speaking or having a conversation about those topics, that I ended up not going forward with joining in the conversation to say and just based on looking at the weekly hashtag, well not the weekly hashtag, the hashtag in general,” (Participant 1). “Unfortunately not, I tried to follow at the beginning the hashtag of the course on Twitter, but I didn’t find it very active, so I didn’t follow up after the first week. I just viewed what are the topics of discussion, but they were not very active. So I didn’t initiate any conversations.” (Participant 2). Despite not being able to actively network through the MOOC, both participants have demonstrated that they would have preferred an increased networking opportunity, which was represented in the code: interest in networking.

3.2 Kirkpatrick evaluation results:

The evaluation levels: reaction, learning, behaviour and results were all analyzed using the data from the semi-structured interviews.

3.2.1 Level 1 Evaluation – Reaction

This level tests participants' perception of the course and answers questions such as; did participants enjoy the course, did they find it useful, and what materials did they find most engaging? Participants' reaction to the course was generally positive. There was a consensus on the course being unique, for offering learning in a brand new topic; RWE, and for being offered by Imperial College; a renowned institute according to participants. Participants found the course videos, assessment to be the most engaging, and appreciate that the course content was up-to-date with the latest research, Participants reacted negatively to the course platform's technical issues and for the lack of communication and networking during and after the course.

3.2.2 Level 2 Evaluation – Learning

This level tests participants' gained learning from the course, which can be in the form of "advancement" in skills, knowledge or attitude (Ayub, Wei and Yue, 2017). Participants demonstrated that they have gained learning from the course by discussing the key topics learned and explaining what they understood. For instance, they talked about information governance, Real World Data, data sources and frameworks for data analysis. These discussions demonstrated that participants have gained knowledge from the course evidenced by their ability to talk about and discuss the course content and topics.

3.2.3 Level 3 Evaluation – Behavior

In terms of behaviour, both participants believed that the course offered knowledge in practical skills they can use in real life. However, both explained that this was not possible yet due to lack of resources or support in their workplace.

3.2.4 Level 4 Evaluation – Results

The course description of the MOOC mentions that the aim of the course is to help students "develop new methods for data analysis" to "inform decision-making in healthcare" (Imperial College London, 2017). Therefore, the most accurate evaluation of the results would have been to evaluate whether the course has affected decision-making in participants' workplace. However, due to the short time period between the course end date and the evaluation interviews, it was not possible to report such results. Nevertheless, overall reaction to the course indicates that these results are likely to be seen in the future.

4. CONCLUSION

This study adds to the current literature on MOOCs developed for counting professional development. The findings acknowledge that in general, MOOCs can improve learning and knowledge attainment in practical skills-based knowledge. One of the implications of this study is to inform factors that engage learners in the design and implementation of MOOC. The findings have shown that factors that affect the learners' engagement are availability of lecture videos, self-assessment tools and high networking and communication between learners. In terms of knowledge application, support and availability of the right resources in the workplace are essential because learners are not able to apply learning in their workplace if lacking the right resources and support. Developers of MOOCs for continuing professional development should take into consideration work-related barriers when designing their MOOCs.

Participants reported increased learning and being introduced to new topics and resources as a result of joining the MOOC. Previous evaluations of MOOCs have reported very positive results in terms of learning. One study comparing learning outcomes from a MOOC to a traditional university classroom, reported better learning outcomes among MOOC students (Colvin et al., 2014). Positive learning outcomes as a result of joining a MOOC was also reported for teaching practical skills to healthcare professionals as a

randomized trial found that a MOOC was sufficient to teach and train physical therapists about spinal cord injuries (Hossain et al., 2015).

Three features were identified as essential features that participants liked in the MOOC; availability of lecture videos, self-assessments and increased networking and communication between participants. While the former two characteristics were available in the current MOOC, the last characteristic was a feature participants criticized for not being delivered effectively. Networking as part of a MOOC is a very important feature as it can increase the number of students joining the MOOC (Liyanagunawardena, Adams and Williams, 2013), and can increase learner satisfaction (Hossain et al., 2015). Therefore, increased effort is needed to increase networking opportunities for learners in the MOOC and to encourage more participation in discussion posts, and making sure that participants return to the posts and continue to actively join discussions.

In terms of applying skills in the workplace and contributing to the continued professional development, this evaluation indicated that participants were not able to take skills from the MOOC and apply them to daily life. This may be due to the topic of the MOOC (RWE) being relatively new, and the data analysis skills taught in the MOOC requiring a complicated set of resources and support to be applied in the workplace. In fact, analyzing RWE requires the availability of multiple sources of data, competent patient protection policies, organizational support and a set of resource (Hubbard and Paradis, 2015), meaning that even if the learning was effective, the lack of these resources prevents students from applying their learning.

4.1 Strengths and limitations

The strength of this evaluation is that it used qualitative data to evaluate learners' reaction, learning and skills gained from the MOOC. Learning from an online-course is most valuable not only when it is offered through the highest quality and latest technologies, but when the learning from the course can affect learners' day-to-day activities in a positive way (Romiszowski, 2003). For these reasons, this evaluation focused the most on evaluating participants' learning and how much of the learning they were able to or will be able to apply in their professional activities. The limitations of the study include lack of data sources in measures such as pre-course survey and post-course surveys and relying mostly on participants' self-reported data to complete the evaluation, which may be at risk of recall bias. Finally, a limitation in our use of the Kirkpatrick evaluation model was that it is intended to be applied 6 months after training, whereas in our evaluation we have used it four months after the course. However, findings from this evaluation could help future MOOC evaluations in determining which factors to study to evaluate the effectiveness of the MOOC and could help researchers consider factors other than learners' knowledge to understand how we can help improve the applicability of the learning from the MOOC in real life.

ACKNOWLEDGEMENT

This work was funded by the Higher Education Funding Council for England.

REFERENCES

Ayub, E., Wei, G.W. and Yue, W.S., 2017. Exploring Factors Affecting Learners' Acceptance of MOOCs Based on Kirkpatrick's Model. In: *Proceedings of the 8th International Conference on E-Education, E-Business, E-Management and E-Learning*, IC4E '17. [online] New York, NY, USA: ACM, pp.34–39. Available at: <http://doi.acm.org/10.1145/3026480.3026490> [Accessed 12 Jan. 2018].

Chapman, S.A., Goodman, S., Jawitz, J. and Deacon, A., 2016. A strategy for monitoring and evaluating massive open online courses. *Evaluation and Program Planning*, 57, pp.55–63.

Clarke, V. and Braun, V., 2013. Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The psychologist*, 26(2), pp.120–123.

Colvin, K.F., Champaign, J., Liu, A., Zhou, Q., Fredericks, C. and Pritchard, D.E., 2014. Learning in an introductory physics MOOC: All cohorts learn equally, including an on-campus class. *The International Review of Research in Open and Distributed Learning*, [online] 15(4). Available at: <http://www.irrodl.org/index.php/irrodl/article/view/1902> [Accessed 5 Feb. 2018].

Daley, B.J., 2004. Using Concept Maps In Qualitative Research. [online] Int. Conference on Concept Mapping. Pamplona, Spain. Available at: <http://cmc.ihmc.us/papers/cmc2004-060.pdf>.

Gallagher, D., 2015. The Talent Dividend: Interactive Infographic. *MIT Sloan Management Review*. Available at: <https://sloanreview.mit.edu/article/the-talent-dividend-interactive-infographic/> [Accessed 5 Feb. 2018].

Goldie, J.G.S., 2016. Connectivism: A knowledge learning theory for the digital age? *Medical Teacher*, 38(10), pp.1064–1069.

Hossain, M.S., Shofiqul Islam, M., Glinsky, J.V., Lowe, R., Lowe, T. and Harvey, L.A., 2015. A massive open online course (MOOC) can be used to teach physiotherapy students about spinal cord injuries: a randomised trial. *Journal of Physiotherapy*, 61(1), pp.21–27.

Hubbard, T.E. and Paradis, R., 2015. *Real World Evidence: A New Era for Health Care Innovation*. [Issue Breif] The Network for Excellence in Health Innovation. Available at: <https://www.nehi.net/writable/publication_files/file/rwe_issue_brief_final.pdf> [Accessed 5 Feb. 2018].

Imperial College London. 2017. Data Science Essentials: Real World Evidence. Available at: <http://www.imperial.ac.uk/admin-services/continuing-professional-development/short-courses/medicine/public-health/data-science/ >[Accessed 5 Feb. 2018].

Khalil, H., 2014. MOOCs Completion Rates and Possible Methods to Improve Retention - A Literature Review. [online] EdMedia: World Conference on Educational Media and Technology. Association for the Advancement of Computing in Education (AACE), pp.1305–1313. Available at: <https://www.learntechlib.org/p/147656/> [Accessed 19 Mar. 2018].

Kirkpatrick, D. and Kirkpatrick, J., 2006. *Evaluating Training Programs: The Four Levels*. Berrett-Koehler Publishers.

Liyanagunawardena, T.R., Adams, A.A. and Williams, S.A., 2013. MOOCs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distributed Learning*, 14(3), pp.202–227.

Patton, M.Q., 1999. Enhancing the quality and credibility of qualitative analysis. *Health services research*, 34(5 Pt 2), p.1189.

Raghupathi, W. and Raghupathi, V., 2014. Big data analytics in healthcare: promise and potential. *Health information science and systems*, 2(1), p.3.

Romiszowski, A., 2003. The future of e-learning as an educational innovation: Factors influencing project success and failure. *Brazilian Review of Open and Distance Education-Teorias Aspectos Teóricos e Filosóficos*.