Mapping the MOOC Research Landscape: Insights from Empirical Studies

https://doi.org/10.3991/ijet.v17i14.28721

Eamon Costello[1](✉), Tiziana Soverino[2], Richard Bolger[1]
[1] Dublin City University, Dublin, Ireland
[2] Crumlin College of Further Education, Dublin, Ireland
eamon.costello@dcu.ie

Abstract—Several reviews have been conducted of empirical studies of MOOC learners and teachers. The scope and foci of such reviews has varied, as has the reporting of the details of how they were conducted. This study analysed 1,435 published articles, determining 922 to be empirical studies. We analysed the full text of 826 of these articles to which the research team had access using the scientometric tool SciVal, manual researcher evaluation and topic modelling to determine: the impact as measured by citations; geographic and institutional publishing patterns; and the themes and types of MOOC research. We found that MOOC research is mostly clustered in the discipline of computer science. Learner persistence and self-regulated learning continue to be a focus of study and most impactful finding respectively as studies of previous periods have found. Research is carried out worldwide, with the most influential studies and researchers clustered in particular institutions and countries. Implications of this study are that MOOC research is clustered in certain ways which may give rise to particular biases, that researchers should consider more interdisciplinary approaches in their research and greater awareness and use of open science principles and practices in their work.

Keywords—MOOCs, empirical research, reviews, scientometrics, online learning

1 Introduction

The "Massive" in MOOC generally refers to the number of learners in a course. MOOCs collectively however are also a massive phenomenon. Following the "Year of the MOOC" in 2012 there was a dramatic rise in scholarly publications on the topic up to 2016. The number of research publications on MOOCs continued to rise from 2016 to 2018 thought the rate of increase has slowed. Figure 1 illustrates this, via a chart showing the number of research articles, indexed by Scopus, in which the term "MOOC" appears in the title, abstract, or article keywords.

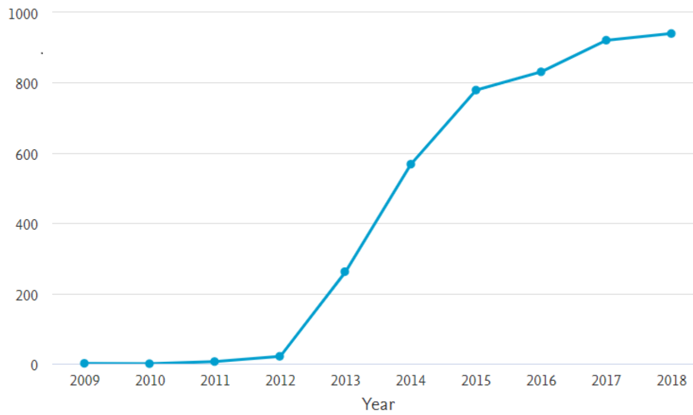

Fig. 1. MOOC publications indexed in Scopus 2009 - 2018

2 MOOC meta-studies and research reviews

In order to make sense of this literature, several reviews have been made [1]. Here we give an overview of these. These meta-studies cover various periods, have various inclusion criteria, different search sources and strategies; and have different foci in their analyses. Some reviews do not exclude empirical studies from their analysis such as Liyanagunawardena, Adams & Williams [2] who focused on studies that "explore the concept of a MOOC or the implications for higher education, report on experiments with MOOCs, or compare MOOCs with other educational approaches". Liyanagunawardena, Adams & Williams [2] covered the period 2008 - 2012 (48 months of data) analysing 45 studies. This is a fraction of the research that is currently being published on MOOCs, even if we confine ourselves to empirical studies, as we will later show. Another review of the early MOOC research [3] considered peer reviewed journal articles published between 2012 to 2013, analysing 25 studies after excluding thought/opinion pieces and editorials. They reported a chronological change from Connectivist cMOOC research focused on Engagement and Creativity, to a xMOOC phase focused on Learning Analytics, Assessment, and Critical Discourse.

A subsequent review analysed literature published between January 2013 and January 2015 (24 months of data) [4]. This review differentiated itself from previous studies in including only empirical research. It did, however, admit studies using both primary and *secondary* data. A subsequent study [5] reported 228 empirical papers that were published in 2013 and the nine first months of 2014 (21 months of data). This review reported very little information on the methodology followed, such as: the process for determining empirical papers, whether abstracts or full texts were analysed or how the articles selected for inclusion were analysed.

A study published in 2018 [6] Identified only studies that addressed a problem related to predicting either learning or persistence in MOOCs. The authors excluded further or adult education. They also excluded qualitative studies, which is arguably a somewhat narrow conception of empirical research. Their study reported searching several databases, and other sources for studies published between 2012 and 2015 (36 months). They analysed a total of 38 studies, after filtering out irrelevant studies due to strict inclusion criteria and did consider studies based on secondary data. In order to determine empirical studies they reported reading 1,004 abstracts. They then reported reading the full texts of an unspecified further number of articles in cases where the abstract contained insufficient information to make a determination as to whether a study was empirical or not. They found most studies to be exploratory in nature and that researchers did not take enough consideration of existing educational frameworks that might have allowed for more salient interpretations of findings.

[7] conducted a review of MOOC literature between 2008 and 2015 (presumably 96 months). It did not distinguish between empirical and non-empirical studies and included only peer reviewed journal articles. After applying exclusion criteria, 362 articles were selected and analysis was conducted on abstracts and titles of articles i.e. not the full article text. Unlike other studies this review used automated content analysis (via the Leximancer software). They reported that studies they analysed were focused on: opportunities and challenges posed by MOOCs for universities; different MOOC platforms; learners and content; and instructional design and quality of learning.

Finally a review [8] analysed studies published between October 2014 and November 2016 (25 months) in Scopus and selected journals. Empirical studies were determined by researchers reading abstracts and, where necessary, full texts, resulting in 146 articles selected for analysis. This review was later augmented with a further seven months of data to July 2017 comprising 51 articles [9]. These studies analysed topics and research approaches of MOOC research. Similar to other reviews, they found the MOOC research has been focused on learner motivation, retention and instructional design.

2.1 Specialisation of MOOC research reviews

It is worth noting that as the MOOC phenomenon has matured, more reviews have been conducted on specialised areas. Such reviews include the topics of: Vocational Education and Training MOOCs[10], Gamification in MOOCS [11], MOOCs for social mobility [12], MOOCs and Twitter [13], predictive models for MOOCs [14], accessibility of MOOCs [15], MOOCs in medical education [16], questionnaires in MOOCs [17], sensory learning [18] and self-regulated learning in MOOCs [19].

Following from the review above of reviews of MOOC studies we have seen that early reviews considered relatively small numbers of studies in their analyses. We have also noted varying degrees of details reported on how the studies were undertaken. This is against the background of increasing acknowledgment of the importance of more open and reproducible approaches to research [20], including in the area of digital learning [21].

We undertook to conduct the current review to address some notable gaps: There is a lack of up to date reviews of recent research on MOOCs [22], in particular there is a dearth of reviews that distinguish empirical studies based on primary data (human participants), and many studies as we have seen are unclear on the full details of their methodology followed. The main research question this study sought to address was to determine what impact empirical MOOC research is having and in which topic areas.

- What MOOC empirical MOOC research is having the most impact as measured by citations?
- What publication outlets are publishing these studies and from which countries and institutions?
- What is the overall structure of themes and topics of research into MOOCs?

3 Method

The present study used the Scopus database as a data source to search the MOOC literature. An advantage of using Scopus over many other data sources is that it provides reliable article level metadata, particularly on article citation [23]. Moreover, the journals and articles it indexes must have met various criteria signifying research quality, including an articulated peer review process, editorial board composition, policy on ethics, plagiarism detection and so on [24] Some search strategies combine multiple data sources, which is often required for a classic systematic literature review methodology. This has the advantage of broadening a search, however it has a downside in that such studies cannot consider citations, as not all of the indices measure citations, and each does so in a different way. Moreover, the majority of the major sources of MOOC related research are indexed in Scopus such as relevant ACM proceedings (Learning @ Scale, Learning Analytics and Knowledge), IEEE proceedings (Learning with MOOCS Conference), Springer Lecture Notes in Computer Science (European MOOCs Stakeholder Summit Proceedings; European Conference on E-learning). Scopus also has a very wide coverage of the relevant journals in the field.

To start with, we conducted a search of Scopus for articles which had the string "MOOC" in either the title, abstract or metadata keywords.

The below inclusion criteria were used to select papers:

a) The papers had to be written in English
b) The papers had to be published in journals or peer-reviewed conference proceedings
c) They had to be published between January 2016 and December 2017 inclusive
d) They had to be electronically available in Scopus

After developing and running this search, a total of 1,435 papers were returned. Following **Error! Bookmark not defined.**, we sought to determine a sample size suitable to conduct inter-rater reliability of two evaluators for a subsample of the articles. We used the IRR package of the statistical software platform R invoking the N.cohen.kappa function with power specified as 95 and alpha as 0.05. This gave a value of 186. Accordingly, two evaluators, Researchers 1 and 3, read the abstracts of 186 articles in the dataset, and independently recorded whether they believed each one related to an empirical study involving human participants. The results were recorded in two spreadsheets. These were then merged and the evaluations of both researchers were compared via Cohen's Kappa and a value of 0.89 was found, indicating a high degree of inter-rater reliability. At this point the researchers discussed the discordant items and reached a consensus on them. Then one researcher, Researcher 3, proceeded to read and classify the remaining 1,249 abstracts.

A close reading of the abstracts of these 1,249 papers revealed that 922 reported on empirical studies involving human participants. At this stage, some duplicates and articles focusing on different topics, that were mistakenly included in the initial search results, were also discovered and eliminated. We thus determined that 922 papers (64.29%) were based on research using data from human participants. The remaining 478 papers (33.33%) were either found to be thought/opinion pieces, position papers, literature reviews or did not contain enough information to make a definite determination from the abstract.

An attempt was made to retrieve the full available texts of these 922 papers from sources available to the research team. 469 were automatically retrieved from Scopus via the batch download feature, which allows retrieval of 50 articles at a time. A manual search was then made for copies of the remaining articles that were legally available to the research team and 827 articles in total were collected. This gave us a dataset of the abstracts and metadata for 922 papers from the period and a corpus of the full text of the articles for 827 of those studies.

To get an overview of the literature corpus as a whole, we undertook topic modeling on the 827 articles for which we had full text copies. Latent Dirichlet Allocation (LDA) is a topic modeling approach that can be used to automatically infer topics in text through word co-occurrence probabilities [25]. Following the recommendations of [26], who conducted an empirical analysis of 420 topic modeling projects to find optimal topic numbers, combined with our own experimentation on the data with various topic numbers, we determined a topic number of 30. We then derived 30 topics from the corpus. These topics were then subjected to manual qualitative analysis by content experts. Firstly, we sought to exclude those that were not coherent topics such as generated from references in the paper or common but meaningless text. Then we sought to classify the topics according to subject of study and focus of the research.

We next exported the dataset of the metadata of 922 papers from Scopus into the SciVal reporting tool, which is a research performance analysis tool [27]. 920 records were successfully returned. This allowed us to generate reports on bibliometric trends in the data, such as the most prominent countries involved in English-language MOOC research, the most prominent institutions and the most highly cited research.

4 Results

Reports were generated from SciVal which first gave an overview of the literature as a whole. Field weighted citation impact refers to a publication's citation count compared to other articles in the same timeframe and field [28]. Publications in the dataset that are outperforming other publications in their field have field weighted citation impacts greater than 1. The average field weighted citation impact of the publications analysed was 1.61, suggesting empirical research on MOOCs is on average more impactful than comparable research in a given field of study. We drilled down further into these fields to get a better idea of what they are. Figure 2 shows the field weighted citation impact of 24 fields (subject areas) within the dataset and the number of papers from the subject area in the dataset (scholarly output). The subject area that papers were most often categorised in were Computer Science, with 682 papers and Social Sciences, with 459 papers (papers can be categorised in more than one category).

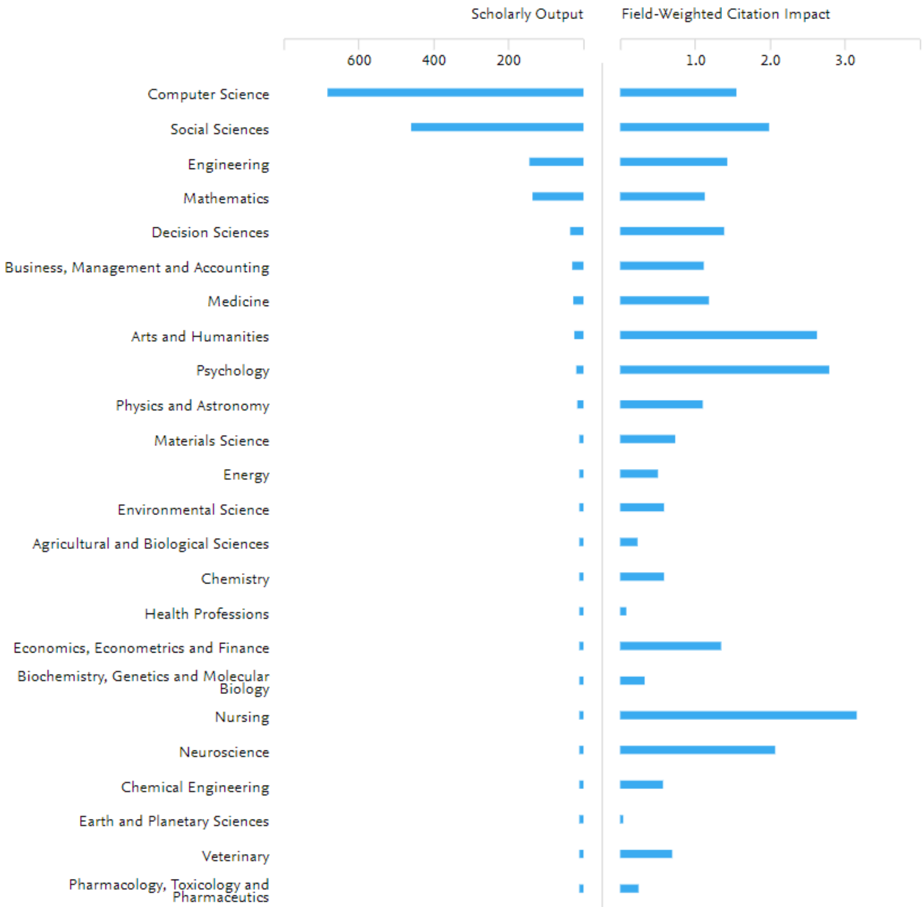

Fig. 2. Field weighted citation impact and scholarly output per category of 920 MOOC publications

Next, we examined publications by publication source. Of the top 10 publication sources, three were journals (*International Review of Research in Open and Distributed Learning, Computers and Education* and *Journal of Computing in Higher Education*) and seven were conference proceedings. As can be seen in Table 1 , the three journals are highly ranked, as evidenced by having high Scopus CiteScores.

Table 1. Empirical studies on MOOCs by Publication Source 2016 - 2018

Scopus Source	Publications	Citations	Authors	CiteScore
Lecture Notes in Computer Science	98	123	312	0.90
CEUR Workshop Proceedings	41	13	134	0.31
2016 Proceedings of the 3rd 2016 ACM Conference on Learning at Scale	34	158	104	0.00
2017 Proceedings of the 4th (2017) ACM Conference on Learning at Scale	27	61	88	0.00
International Review of Research in Open and Distance Learning	22	80	55	2.73
Proceedings • Frontiers in Education Conference, FIE	15	16	48	0.45
ACM International Conference Proceeding Series	15	131	58	0.56
Computers and Education	13	295	41	588
Communications in Computer and Information Science	12	16	39	0.39
Journal of Computing in Higher Education	10	70	36	2.44

Table 2 shows the distribution by country which was determined by lead author affiliation. It can be seen that most research is published by authors from the United States, China and countries in Western Europe.

Table 2. Top 20 countries publishing empirical MOOC research studies 2016 – 2018

Country	Publications	% of total
United States	116	12.6%
China	63	6.8%
United Kingdom	38	4.1%
Spain	34	3.7%
France	32	3.5%
Australia	25	2.7%
India	22	2.4%
Germany	22	2.4%
Taiwan	20	2.2%
Italy	13	1.4%
Canada	11	1.2%
Netherlands	10	1.1%
Japan	8	0.9%
Brazil	8	0.9%
South Korea	7	0.8%
Sweden	7	0.8%
Russian Federation	7	0.8%
Norway	7	0.8%
Greece	6	0.7%

We further broke this down by institution and found that there were 583 institutions involved in publishing MOOC research. The top 20 institutions by scholarly output are listed in Table 3.

Table 3. Top 20 institutions publishing empirical MOOC research by publication count

Institution	Country	Output	Authors
Purdue University	United States	21	25
Massachusetts Institute of Technology	United States	21	28
Carnegie Mellon University	United States	20	34
Universidad Carlos III de Madrid	Spain	19	25
Tsinghua University	China	18	52
Delft University of Technology	Netherlands	17	17
Technical University of Madrid	Spain	17	29
Harvard University	United States	16	26
University of Potsdam	Germany	16	20
Open University of the Netherlands	Netherlands	15	17
Swiss Federal Institute of Technology Lausanne	Switzerland	14	21
Pontificia Universidad Catolica de Chile	Chile	13	14
Graz University of Technology	Austria	13	19
Peking University	China	13	31
National Distance Education University	Spain	13	30
Stanford University	United States	12	20
Georgia Institute of Technology	United States	12	22
Pennsylvania State University	United States	11	27
Open University Milton Keynes	UK	11	15
Universidad de Salamanca	Spain	11	11

To gain further insight into the impact of these studies, however, we determined the top 20 institutions by field weighted citation impact as per Table 4.

Table 4. Top 20 institutions by field weighted citation impact of publications

Institution	Country	Field-Weighted Citation Impact	Citations	Output
Brunel University	United Kingdom	21.39	75	1
University of Oxford	United Kingdom	16.48	88	1
Technion-Israel Institute of Technology	Israel	15.97	56	1
Tongji University	China	14.13	53	1
University of Macau	Macao	13.69	48	1
The California State University	United States	12.03	11	1
Future University in Egypt	Egypt	11.92	82	2
University of Houston	United States	10.55	38	1
Universidad de Cuenca	Ecuador	10.45	117	5
Texas Tech University	United States	10.05	63	1

Glasgow Caledonian University	United Kingdom	9.52	123	3
Stanford University	United States	8.57	247	12
Union Memorial Hospital	United States	8.27	24	1
University of Arizona	United States	7.63	19	2
Central University of Finance and Economics	China	6.92	14	1
University Politehnica of Bucharest	Romania	6.54	25	1
Goethe University Frankfurt	Germany	6.46	8	1
University of Western Australia	Australia	6.42	13	1
National University of Singapore	Singapore	6.24	50	3

Table 3 shows that a small number of publications account for the largest field weighted citation impact. This is explained by the publications in prestigious journals that have also attracted high numbers of citations. For example, the top five publications by citation are listed in Table 5.

Table 5. Five most cited publications

Publication	Citations	Field Weighted Citation Impact
Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *The Internet and Higher Education*, *29*, 40–48.	88	16.48
Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education*, *104*, 18–33.	77	31.13
Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting MOOC retention: A survey study. *Computers & Education*, *98*, 157–168.	75	21.39
Hew, K. F. (2016). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS. *British Journal of Educational Technology*, *47*(2), 320–341.	68	19.96
Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal prediction of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior*, *58*, 119–129.	63	10.05

As can be seen in Table 5 above the five most cited publications [29]–[33] are all in prestigious journals (Computers in Education, The Internet and Higher Education and Computers in Human Behavior) which contributes to their high field weighted citation impact scores. It was also noted that a publication such as [34] on MOOC assessment in the journal College Composition and Communication has a high field weighted citation impact score 12.03 despite having only 11 citations. This is accounted for by the fact that it is in the field of Literature and Literary Theory which attracts relatively few citations on average to papers. This confirms the finding as shown in Figure 2 that the MOOC literature is having a large citation impact on certain fields, relative to other publications in those fields.

Manual coding by expert evaluators, Researchers 1 and 2, of the 30 topics resulted in 11 coherent topics. Drawing on the analysis conducted by [7] these topics were then further classified by the evaluators into broad subject areas. Two main areas emerged:

Research Problem and Instructional Design (Indeed Instructional Design appeared as its own topic). Below we show how we thematically grouped the 11 topics.

- Dominant Research Problem
  - Completion
  - Predictive Algorithms
  - Motivation/engagement
- Instructional Design
  - Assessment
    - Multiple choice questions
    - Discussion forums
  - Content
    - Coding
    - The Association for Computing Machinery (ACM)/The Institute of Electrical and Electronics Engineers (IEEE)
  - MOOC platforms

Figure 3 shows an example of the topic model that we labelled "Discussion forums":

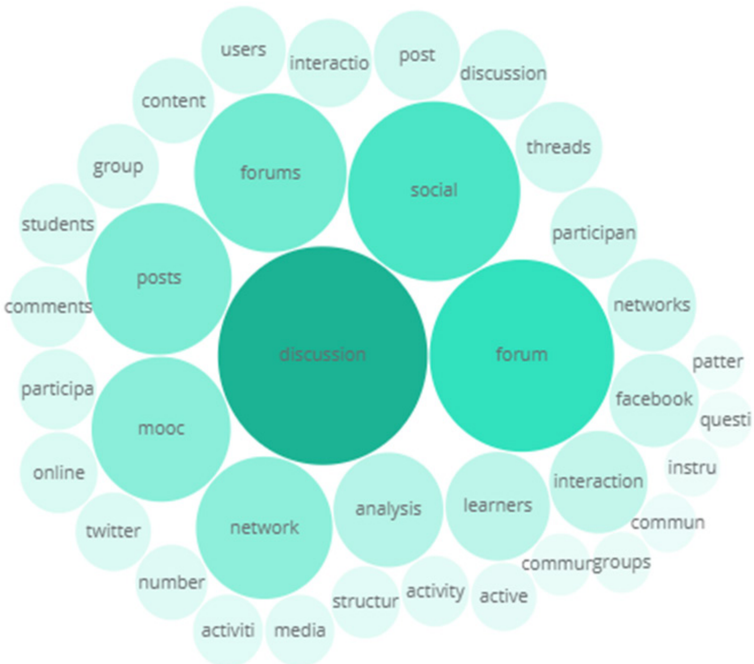

Fig. 3. "Discussion forums" topic

5 Discussion

This article set out to map topics and influence of empirical MOOC research through systematic analysis of literature. It considered 1,435 abstracts, which led to a review of 922 empirical studies. In our topic analyses of the literature corpus, we found completion, predictive algorithms, and Motivation/engagement to be the dominant foci of research questions. This is borne out by recent reviews (e.g. [1]), but it also ties with the most influential recent literature in this area as measured by citations. For instance, the top two most cited studies are on self-regulated learning [29][30]. Why some learners persist in their studies but not others have been a longstanding topic of educational research [35] and MOOCs have provided an opportunity to look at this problem at scale online. How students engage via self-regulated learning appears to be one of the most interesting factors to researchers.

One of the topics that emerged contained the words ACM, IEEE and coding and highlights how MOOC research is dominated by Computer Scientists. This has been borne out by previous research which has highlighted how computing studies are the dominant type of empirical MOOC research [7]**.** This may skew the focus of research on MOOCs. Researchers from different tribes or traditions may have very different philosophical worldviews [36]. The topics we found on multiple choice questions and discussion forums were also found with words such as learners, participants, students and groups (see Figure 3) i.e. with a subtopic of Learner. There is less mention of teachers and teaching, perhaps understandably as they are outnumbered by learners. However, commentators have posed the question: "Where are the educators?"[37]. The implication being, that the AI and machine learning employed in MOOCs is being seen as a way to automate teaching and perhaps replace teachers. This claim of the rise in emphasis on learning as opposed to teaching, which has been referred to as the "learnification" of education [38], is supported by our findings of the research on MOOCs. An interesting counterpoint is made however to this trend by Dillenbourg [39], who predicts that *more teachers* will be needed to help orchestrate the increasingly complex and powerful AI ensembles available in educational contexts.

A country such as Spain, as evidenced in Table 3, is the source of a high number of publications on MOOCs. However, when the impact of these publications is taken into account, we saw that the institutions responsible are more clustered in the United States. We see that the most impactful areas as measured by citations continue to be in retention and self-regulated learning. Self-regulated learning appears less prominently in the body of research overall (as determined by topic modelling of the corpus) in terms of its themes. Our findings from this analysis of the literature show that self-regulated learning continues to be one of the key research themes in MOOC research as literature reviews covering previous periods have reported. This may be one of the enduring contributions of MOOCs to the research literature of online learning as measured by citation-based impact [6][7][8][9].

As we highlighted at the outset, the meta-studies of MOOC research to date vary in their scope and methodology. There are some notable gaps in both what they cover and what they report in their methods. This study attempted to address this gap and build on the work of previous studies in this area through the generation of a published open dataset of empirical MOOC research (see https://doi.org/10.5281/zenodo.3736141). We explained in detail the exact steps we took and methods we followed in our review so that others can replicate this work or interrogate and use the data we have created.

6 Limitations and further work

For topic modelling, research indicates that corpora with small numbers of documents produce low-quality topics, and that as the number of documents reaches about 1,000, the topic coherence stabilizes [40]. The 922 articles in our study do not quite meet this ideal threshold of 1,000 documents. We plan to widen the timeframe of published studies and hence expand the dataset in future to encompass more documents in order to get a richer picture of the latent topics.

Our main data source Scopus does not cover all possible journals where MOOC research is published. However, a recent review [8] that used Scopus, combined with selected other journals, found Scopus to index the vast majority of MOOC related research and particularly the types of study we sought to determine in our research (such as impactful studies). This suggests there would only be minimal value in including non-Scopus indexed sources in our search. Moreover, to analyse citation data, a single indexing source should only be used i.e. citation counts cannot be aggregated from different sources, as citations are tracked and counted differently in each different indices e.g. Web of Science citations will be different to those from Scopus and from Google Scholar and so on. Scopus is a valuable source for this study as it indexes proceedings of the ACM and IEEE and all the major conferences on learning analytics and MOOCS. However, the findings of this research warn that there is a need to expand the research net beyond computer science. Otherwise we may get a skewed representation of what learning in MOOCs looks like. Moreover, we have shown that researchers from other home disciplines may gain a greater impact for their research, relative to their field, by publishing studies on MOOCs. This may act as an incentive for scholars from multiple fields to get involved and help produce the wide interdisciplinary type of research that is needed into how learning happens at scale.

There are many other research questions that the dataset we generated in this study could be used to answer. For example, there is an increasing concern over ethics in digital learning research, and in particular about the treatment of learners in studies that are conducted at scale such as with MOOCs [41], [42]. The papers in this dataset could be interrogated to determine the reported ethical considerations of researchers in order to ask questions such as: whether they had received approval from an Institutional Review Board/Ethics Committee for the research; whether learners had consented to the research and what other ethical principles or protocols were followed. It could be particularly interesting to see whether the treatment of the data participants has changed due to recent GDPR legislation. Using the large body of empirical studies in this dataset, researchers could examine the practices of researchers pre and post the introduction of GDPR. This is an active line of future research we are pursuing. Lastly, similar to several of the major studies published in this area that report on historical datasets [6], [7], we are aware that there are already a whole swathe of new articles published since this study was conducted. We hope that our data and approach however will help in this future research.

7 Conclusion

This paper has highlighted that there are gaps in the reviews to date of empirical research into MOOCs, both in the methodologies employed and the time periods covered. This study attempted to address this gap through the creation of a dataset on a significant body of published research. We then further used this dataset to highlight prominent topics in MOOC research, such as learning design and self-regulated learning and show the impact by weighted citation values that this research is having. We caution that this research is clustered in the domain of computer science, and we will need to widen the disciplinary net in future to ensure broader, more representative research. Our findings should help guide future research into how learning may best be conducted at scale.

8 Declarations

Funding: The Open Education Unit of the National Institute for Digital Learning in Dublin City University is gratefully acknowledged for funding support to make this publication open access.

Availability of data and materials: E. Costello, R. Bolger, and T. Soverino, "Dataset of Empirical MOOC Research 2016 – 2017". Zenodo, Apr. 01, 2020. doi: 10.5281/ZENODO.3736140.

9 References

[1] A. Bozkurt, 'Surfing on Three Waves of MOOCs: An Examination and Snapshot of Research in Massive Open Online Courses', Open Praxis, vol. 13, no. 3, p. 296, Dec. 2021. https://doi.org/10.5944/openpraxis.13.3.132

[2] T. R. Liyanagunawardena, A. A. Adams, and S. A. Williams, 'MOOCs: A systematic study of the published literature 2008-2012', The International Review of Research in Open and Distributed Learning, vol. 14, no. 3, p. 202, Jul. 2013. https://doi.org/10.19173/irrodl.v14i3.1455

[3] M. Ebben and J. S. Murphy, 'Unpacking MOOC scholarly discourse: a review of nascent MOOC scholarship', Learning, Media and Technology, vol. 39, no. 3, pp. 328–345, Jul. 2014. https://doi.org/10.1080/17439884.2013.878352

[4] G. Veletsianos and P. Shepherdson, 'A Systematic Analysis and Synthesis of the Empirical MOOC Literature Published in 2013–2015', The International Review of Research in Open and Distributed Learning, vol. 17, no. 2, Mar. 2016. https://doi.org/10.19173/irrodl.v17i2.2448

[5] A. Sangrà, M. González-Sanmamed, and T. Anderson, 'Meta-Analysis of the Research About Mooc During 2013-2014', Educación XX1, vol. 18, no. 2, 2015. https://doi.org/10.5944/educxx1.14808

[6] S. Joksimović et al., 'How Do We Model Learning at Scale? A Systematic Review of Research on MOOCs', Review of Educational Research, vol. 88, no. 1, pp. 43–86, Feb. 2018. https://doi.org/10.3102/0034654317740335
[7] O. Zawacki-Richter, A. Bozkurt, U. Alturki, and A. Aldraiweesh, 'What Research Says About MOOCs – An Explorative Content Analysis', 1, vol. 19, no. 1, Feb. 2018. https://doi.org/10.19173/irrodl.v19i1.3356
[8] M. Zhu, A. Sari, and M. M. Lee, 'A systematic review of research methods and topics of the empirical MOOC literature (2014–2016)', The Internet and Higher Education, vol. 37, pp. 31–39, Apr. 2018. https://doi.org/10.1016/j.iheduc.2018.01.002
[9] M. Zhu, A. Sari, and C. Bonk, 'A Systematic Review of MOOC Research Methods and Topics: Comparing 2014-2016 and 2016-2017', Jun. 2018, pp. 1673–1682. Accessed: May 25, 2019. [Online]. Available: https://www.learntechlib.org/primary/p/184395/
[10] R. M. Paton, A. E. Fluck, and J. D. Scanlan, 'Engagement and retention in VET MOOCs and online courses: A systematic review of literature from 2013 to 2017', Computers & Education, vol. 125, pp. 191–201, Oct. 2018. https://doi.org/10.1016/j.compedu.2018.06.013
[11] M. Khalil, J. Wong, B. de Koning, M. Ebner, and F. Paas, 'Gamification in MOOCs: A review of the state of the art', in 2018 IEEE Global Engineering Education Conference (EDUCON), Apr. 2018, pp. 1629–1638. https://doi.org/10.1109/EDUCON.2018.8363430
[12] K. van de Oudeweetering and O. Agirdag, 'MOOCS as Accelerators of Social Mobility? A Systematic Review', Journal of Educational Technology & Society, vol. 21, no. 1, pp. 1–11, 2018.
[13] E. Costello, M. Brown, M. N. G. Mhichíl, and J. Zhang, 'Big course small talk: twitter and MOOCs — a systematic review of research designs 2011–2017', Int J Educ Technol High Educ, vol. 15, no. 1, p. 44, Dec. 2018. https://doi.org/10.1186/s41239-018-0127-9
[14] P. M. Moreno-Marcos, C. Alario-Hoyos, P. J. Muñoz-Merino, and C. D. Kloos, 'Prediction in MOOCs: A review and future research directions', IEEE Transactions on Learning Technologies, pp. 1–1, 2018. https://doi.org/10.1109/TLT.2018.2856808
[15] S. Sanchez-Gordon and S. Luján-Mora, 'Research challenges in accessible MOOCs: a systematic literature review 2008–2016', Univ Access Inf Soc, vol. 17, no. 4, pp. 775–789, Nov. 2018. https://doi.org/10.1007/s10209-017-0531-2
[16] F. Zhao et al., 'The comparison of teaching efficiency between massive open online courses and traditional courses in medicine education: a systematic review and meta-analysis', Ann Transl Med, vol. 6, no. 23, Dec. 2018. https://doi.org/10.21037/atm.2018.11.32
[17] M. Lu, T. Cui, Z. Huang, H. Zhao, T. Li, and K. Wang, 'A Systematic Review of Questionnaire-Based Quantitative Research on MOOCs', IRRODL, vol. 22, no. 2, pp. 285–313, Jan. 2021. https://doi.org/10.19173/irrodl.v22i2.5208
[18] P. Limone, S. Pati, G. A. Toto, R. Di Fuccio, A. Baiano, and G. Lopriore, 'Literature Review on MOOCs on Sensory (Olfactory) Learning', Computers, vol. 11, no. 3, p. 32, Feb. 2022. https://doi.org/10.3390/computers11030032
[19] J. Wong, M. Baars, D. Davis, T. V. D. Zee, G.-J. Houben, and F. Paas, 'Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review', International Journal of Human–Computer Interaction, vol. 35, no. 4–5, pp. 356–373, Mar. 2019. https://doi.org/10.1080/10447318.2018.1543084
[20] B. Fecher and S. Friesike, 'Open science: one term, five schools of thought', in Opening science, Springer, 2014, pp. 17–47. https://doi.org/10.1007/978-3-319-00026-8_2
[21] P. Kraker, D. Leony, W. Reinhardt, and G. Beham, 'The case for an open science in technology enhanced learning', International Journal of Technology Enhanced Learning, vol. 3, no. 6, pp. 643–654, 2011. https://doi.org/10.1504/IJTEL.2011.045454

[22] B. Abdelghani, 'Trends in MOOCs Research: Analysis of Educational Technology Journals', Int. J. Emerg. Technol. Learn., vol. 15, no. 17, p. 47, Sep. 2020. https://doi.org/10.3991/ijet.v15i17.14637
[23] L. Colledge, F. de Moya-Anegón, V. Guerrero-Bote, C. López-Illescas, M. E. Aisati, and H. Moed, 'SJR and SNIP: two new journal metrics in Elsevier's Scopus', Serials, vol. 23, no. 3, pp. 215–221, Nov. 2010. https://doi.org/10.1629/23215
[24] 'Content Policy and Selection - Content - Scopus - Solutions | Elsevier'. https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection (accessed May 24, 2019).
[25] D. M. Blei, A. Y. Ng, and M. I. Jordan, 'Latent Dirichlet Allocation', Journal of Machine Learning Research, vol. 3, no. Jan, pp. 993–1022, 2003.
[26] T. Schmiedel, O. Müller, and J. vom Brocke, 'Topic Modeling as a Strategy of Inquiry in Organizational Research: A Tutorial With an Application Example on Organizational Culture', Organizational Research Methods, p. 1094428118773858, May 2018. https://doi.org/10.1177/1094428118773858
[27] 'SciVal', Jan. 02, 2020. https://www.scival.com/ (accessed Jan. 14, 2019).
[28] L. Colledge, Snowball Metrics Recipe Book. Elsevier, 2017. Accessed: Jun. 18, 2019. [Online]. Available: https://www.snowballmetrics.com/metrics/snowball-metrics-recipe-book/
[29] A. Littlejohn, N. Hood, C. Milligan, and P. Mustain, 'Learning in MOOCs: Motivations and self-regulated learning in MOOCs', The Internet and Higher Education, vol. 29, pp. 40–48, 2016. https://doi.org/10.1016/j.iheduc.2015.12.003
[30] R. F. Kizilcec, M. Pérez-Sanagustín, and J. J. Maldonado, 'Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses', Computers & education, vol. 104, pp. 18–33, 2017. https://doi.org/10.1016/j.compedu.2016.10.001
[31] K. S. Hone and G. R. El Said, 'Exploring the factors affecting MOOC retention: A survey study', Computers & Education, vol. 98, pp. 157–168, 2016. https://doi.org/10.1016/j.compedu.2016.03.016
[32] K. F. Hew, 'Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS', British Journal of Educational Technology, vol. 47, no. 2, pp. 320–341, 2016. https://doi.org/10.1111/bjet.12235
[33] W. Xing, X. Chen, J. Stein, and M. Marcinkowski, 'Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization', Computers in human behavior, vol. 58, pp. 119–129, 2016. https://doi.org/10.1016/j.chb.2015.12.007
[34] Denise K. D.K and Edward M. E.M, 'Adventuring into MOOC writing assessment: Challenges, results, and possibilities', Coll. Compos. Commun. College Composition and Communication, vol. 67, no. 3, pp. 318–359, 2016.
[35] M. Badali, J. Hatami, S. K. Banihashem, E. Rahimi, O. Noroozi, and Z. Eslami, 'The role of motivation in MOOCs' retention rates: a systematic literature review', RPTEL, vol. 17, no. 1, p. 5, Dec. 2022. https://doi.org/10.1186/s41039-022-00181-3
[36] P. R. Trowler, Academic tribes and territories. McGraw-Hill Education (UK), 2001.
[37] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, 'Systematic review of research on artificial intelligence applications in higher education – where are the educators?', Int J Educ Technol High Educ, vol. 16, no. 1, p. 39, Dec. 2019. https://doi.org/10.1186/s41239-019-0171-0
[38] G. J. Biesta, 'Giving teaching back to education: Responding to the disappearance of the teacher', Phenomenology & Practice, vol. 6, no. 2, pp. 35–49, 2012. https://doi.org/10.29173/pandpr19860

[39] P. Dillenbourg, 'The Evolution of Research on Digital Education', Int J Artif Intell Educ, vol. 26, no. 2, pp. 544–560, Jun. 2016. https://doi.org/10.1007/s40593-016-0106-z

[40] L. Nguyen, 'Topic modeling with more confidence: A theory and some algorithms', presented at the Pacific-Asia Knowledge Discovery and Data Mining, Ho Chi Minh City, 2015.

[41] E. Zeide and H. Nissenbaum, 'Learner Privacy in MOOCs and Virtual Education', Theory and Research in Education, vol. 16, no. 3, pp. 280–307, Nov. 2018. https://doi.org/10.1177/1477878518815340

[42] M. Khalil, P. Prinsloo, and S. Slade, 'User Consent in MOOCs – Micro, Meso, and Macro Perspectives', The International Review of Research in Open and Distributed Learning, vol. 19, no. 5, Nov. 2018. https://doi.org/10.19173/irrodl.v19i5.3908

10 Authors

Dr Eamon Costello is an Associate Professor of Digital Learning at Dublin City University. He is curious about how we think, learn, and work online, offline, and everywhere in between. He has taught, researched, and published on a variety of topics in digital learning, STEM education, open education, and post digital social science fiction. His work in this field has been widely cited and he has successfully delivered several externally funded research projects.

Dr Tiziana Soverino holds a PhD in Irish Folklore and Early Irish from University College Dublin, and is a Teacher and an Editor.

Richard Bolger holds a Master of Science in Management of Internet Enterprise Systems, from DCU, and he is the Chair of IT Programmes in DCU Open Education.

Article submitted 2021-12-08. Resubmitted 2022-05-15. Final acceptance 2022-05-23. Final version published as submitted by the authors.