



A bibliometric review on latent topics and trends of the empirical MOOC literature (2008–2019)

Caixia Liu^{1,4} · Di Zou² · Xieling Chen¹ · Haoran Xie³ · Wai Hong Chan¹

Received: 4 October 2020 / Revised: 28 March 2021 / Accepted: 30 March 2021 / Published online: 17 April 2021
© Education Research Institute, Seoul National University, Seoul, Korea 2021

Abstract

Massive Open Online Courses (MOOCs) have become a popular learning mode in recent years, especially since the outbreak of COVID-19 in late 2019, which had resulted in a significant increase in associated research. This paper presents a bibliometric review of 1078 peer-reviewed MOOC studies between 2008 and 2019. These papers are extracted from three influential databases, the Web of Science (WOS), Scopus, and the Education Resources Information Center (ERIC). The MOOC literature analysis with a bibliometric approach identified the research trends, journals, countries/regions, and institutions with high H-index, scientific collaborations, research topics, topic distributions of the prolific countries/regions and institutions, and annual topic distributions, after which the representative research and research implications were discussed. This review gives researchers a deep and comprehensive understanding of current MOOC research and identifies potential research topics and collaborative partners, which supports MOOC-related future research.

Keywords MOOC · Structural topic modeling · Bibliometric analysis · Literature review

Introduction

The COVID-19 outbreak in late 2019 put online learning back in the spotlight and made online education one of the hottest topics in education. In the past few years, the rapid development of information and communication technology has resulted in major changes in education delivery, with online learning developing rapidly. Compared with traditional learning, online learning has fewer time and space constraints, making learning more flexible for both teachers and learners. As a typical online education form and a powerful substitute for the classroom, MOOC, an acronym

for Massive Open Online Course, is an online course for the public and the latest development of distance education (Deng & Benckendorff, 2021). MOOCs originated in Canada in 2008 when the 12-week “Connectivism and Connective Knowledge” course was facilitated by Stephen Downes and George Siemens at the University of Manitoba (Boyatt et al., 2014; De Waard et al., 2012). The “Massive” in the MOOC acronym indicates that there are no enrollment limitations and the “Open” indicates that learners are free from geographical constraints, course sizes, temporal boundaries, entry requirements, or financial restraints (Dodson et al., 2015). “Online,” of course, refers to learning through the internet (Thompson, 2011). Downes (2008) categorized two main types of MOOCs: networks of distributed online resources (cMOOCs) and structured learning pathways centralized on digital platforms (xMOOCs). cMOOCs are based on connectivism learning theory (Siemens, 2004), which emphasizes creation, creativity, autonomy, social networking, and connected and collaborative learning (Saadatdoost et al., 2015), whereas xMOOCs have more traditional classroom settings, the instructor and learner roles are differentiated, and the courses are similar to formal university courses, with a combination of pre-recorded video lectures with quizzes, tests, and other assessments (Rabin et al., 2019). In sum, xMOOCs are centered on professors rather

✉ Di Zou
dizoudaisy@gmail.com

¹ Department of Mathematics and Information Technology, The Education University of Hong Kong, 10 Lo Ping Road, Hong Kong SAR, China
² Department of English Language Education, The Education University of Hong Kong, 10 Lo Ping Road, Hong Kong SAR, China
³ Department of Computing and Decision Sciences, Lingnan University, 8 Castle Peak Road, Hong Kong SAR, China
⁴ Institute of EduInfo Science and Engineering, Nanjing Normal University, Nanjing, China

than a community of students (Online Education Blog of Touro College, 2013) and focus on knowledge duplication (Dodson et al., 2015; Siemens, 2012), and cMOOCs focus on knowledge creation and generation.

In the past few years, there has been increased research interest in MOOCs. This study took a bibliometrics approach to review the MOOCs academic research with the aim of providing a deeper understanding of the research status, trends, and priority topics, and to provide guidance for future research. Therefore, this study was driven by the following research questions.

- (1) What was the annual trend of MOOC research?
- (2) Which journals, countries/regions, and institutions were the major MOOC research contributors?
- (3) What were the scientific collaborations among major countries/regions and institutions?
- (4) What were the main research topics of empirical MOOC studies?
- (5) What have the topic distributions and the annual topic distributions been in the prolific countries/regions and institutions?

After reviewing MOOC-related research in the section of Literature review, the Methods section introduced the bibliometrics review method. Then the Results section presented the analysis of the descriptive and qualitative statistics, such as the article and citation counts, the most prolific countries/regions and institutions, the scientific collaborations, the main topics, trends, and correlations, annual topic distributions in the most prolific countries/regions and institutions. The Discussion section provided an in-depth discussion, the limitations of this study, and the possible areas for future research, and the Conclusion briefly reviewed the main points of this paper.

Literature review

MOOC research was analyzed from macro- and micro-perspectives to identify the macro-development trends and the specific (micro) research directions or issues, respectively.

Macro-perspective of the MOOC review

The macro-perspective of the MOOC review focused on the issues of MOOC itself, such as the related literature number, MOOC classification, research methods, topics, annual trend, and social ethics. Liyanagunawardena et al. (2013) conducted the first review of 45 MOOC research articles published from 2008 to 2012 in academic journals, for which a quantitative analysis was conducted on article classification, contributor distribution, annual research trends, MOOC

classifications, and possible future research directions. Similarly, Veletsianos and Shepherdson (2016) analyzed 183 articles published from 2013 to 2015 using both qualitative and quantitative methods and came to three main conclusions: (1) most articles were by American and European researchers; (2) only a few papers were widely cited with nearly half not cited, and (3) quantitative methods were more favored with the data mainly collected using surveys and automated methods. However, the research was based on a very small portion of the available data, which restricted the understanding of MOOCs. Different from these two reviews, Saadatdoos et al. (2015) explored and analyzed 32 MOOC research studies from education and information system perspectives, from which a holistic MOOCs definition was derived and relevant theories and issues extracted, which significantly contributed to the creation of a MOOCs research domain structure; however, this study lacked any broader, deeper analysis of MOOC research institutions, collaborations, and other factors. Ebben and Murphy (2014) analyzed 25 empirical studies from 2009 to 2013 that chronologically conceptualized MOOC scholarship themes under (1) connectivist MOOCs, engagement, and creativity from 2009 to 2011/2012; and (2) xMOOCs, learning analytics, assessment, and critical discourses on MOOCs from 2012 to 2013. However, the research only had a MOOC scholarship perspective and only a limited number of papers were reviewed. With a focus on MOOC research methods and topics, Zhu et al. (2018a, b) conducted a systematic review of 146 empirical MOOC studies in five key journals from 2014 to 2016, for which they divided the research methods into quantitative, qualitative, and mixed methods to reveal the relationships between research topics and research methods, and then comprehensively analyzed the trends, research methods, author locations, MOOC delivery countries, and primary journals; however, only a limited number of articles were extracted from Scopus and only a three year time period was examined, which limited the research findings. Deng et al. (2019) conducted a narrative review of 102 MOOC research articles published between 2014 and 2016 using a Perceive, Process, Perform (3P) Model focused on learner factors, teaching contexts, learner engagement, and learning outcomes, and found that there was little evidence-based research on the non-mainstream MOOCs consumers, there was an oversimplification of the role of the learner factors in the evidence-based MOOC research, and that research between teaching and learning helped progress the understanding of MOOC research. However, the focus of the analysis was on the findings rather than on the methodological approaches and the use of the 3P model introduced some ontological constraints. Rolfe (2015) conducted a systematic review of 68 pre-2014 MOOC focused articles from a socio-ethical perspective and developed a socio-ethical dimensional MOOCs framework that encompassed

MOOC pedagogy and quality, the social inclusion afforded by MOOCs, learner diversity and equality, and the digital and social media literacy of the open learners. However, as there have been many more MOOC articles since 2015, further reviews and research are needed.

Detailed information on the review articles: research keywords, databases, article types, time ranges, methodologies, and article numbers, is shown in Table 1.

Micro-perspective of the MOOC review

The micro-perspective of the MOOC review studies mainly focused on particular aspects/topics related to MOOC users, such as student participation, active learning strategies, engagement and retention, academic engagement, and self-regulated learning, with the greater number of these studies being conducted since 2018. For example, based on 38 articles from 2012 to 2015, Joksimović et al. (2018) conducted a systematic review of the approaches to model learning in MOOCs that specifically examined the approaches to defining and measuring learning outcomes, learning contexts, student engagement, and the association between the identified metrics and measured outcomes, after which a framework was suggested to study the associations between the contextual factors such as demographics and classrooms and individual needs, student engagement, and learning outcomes. Paton et al. (2018) analyzed 38 articles from 2013 to 2017 focused on learner engagement and retention in vocational MOOCs education and training, from which six functional approaches were identified to improve learner retention and promote engagement: (1) good quality instructional course design; (2) well-developed assessment tasks aligned with course objectives; (3) learner collaboration opportunities; (4) instructor commitment to timely contextualized communication; (5) course achievement certifications, and (6) further study pathways. As both the above reviews analyzed 38 articles within a narrow time frame of about 3–4 years, there was a need to extend current empirical knowledge to explore more findings. Davis et al. (2018) investigated 126 MOOC studies published between 2009 and 2017 from an active learning perspective and found that the three most effective active learning strategies were cooperative learning, simulations and gaming, and interactive multimedia, and Guajardo Leal et al. (2019) focused on MOOC learning engagement and reviewed 176 articles published from 2015 to 2018, finding that most related articles were from the United States, Australia, and the United Kingdom and most had employed qualitative exploratory methods.

Since 2019, there has been a greater MOOC research focus on self-regulated learning (SRL). For example, Lee et al. (2019) presented a systematic review of empirical research on SRL in MOOCs focused on the effects of SRL on learning, SRL strategies, and SRL interventions, and

suggested some MOOC designs to promote SRL. Wong et al. (2019) conducted a systematic review on SRL that paid greater attention to the human factors in SRL, such as prompt feedback, integrated support systems, and other human factors, finding that human factors (e.g., gender, cognitive abilities, prior knowledge) played an important role in effective SRL, which suggested that to provide the support that best fits each individual learner, learning analytics could be used. However, these SRL-related reviews only examined between 21 and 42 articles; therefore, as there has been an increase in SRL-related MOOC literature in recent years, this topic needs further exploration.

Table 2 lists the related review article information: topic, scope, methodology, number, and journals.

Therefore, while there have been significant MOOC review research studies, there have been some limitations. First, most studies employed systematic rather than bibliometrics reviews, and although Zheng and Yang (2017) claimed that their study was a bibliometrics MOOC review, the research mainly focused on the development trends and popular topics in a four-year span, and while they tracked the evolution of MOOC studies using statistics and identified the popular subjects using a co-word network atlas based on keywords, the research lacked a larger scope or a deeper exploration of MOOCs. Second, past reviews have tended to examine less than 100 papers and only three-to-four-year time periods, which could have hindered the effectiveness of any statistical analyses. As shown in Table 1, almost all reviews were published before 2017, with very few studies from macro-perspectives having been conducted in the past three years. Third, little past review research has systematically conducted topic analyses, topic distributions, and the cooperative research.

In contrast to these earlier studies, this bibliometrics study examined 1078 studies over 10 years and, therefore, provides a more detailed picture of MOOC topics, development trends, cooperative partners, collaborative organizations, topic distributions of the prolific countries/regions and institutions, annual topic distributions, and further discussed the representative research work, and research implications.

Methods

The most used method in MOOC-related review research has been a systematic review approach that describes and justifies the paper identification methods in such a way that it can be replicated (Fink, 2010; Liyanagunawardena et al., 2013). The bibliometrics approach focuses on “the application of mathematics and statistical methods to books and other media of communication (Pritchard, 1969, p. 349).” Howkins (1981) claimed that Bibliometrics implied the quantitative analysis of the bibliographical features of the

Table 1 Macro-perspective MOOC review studies

Author	Keywords	Database	Article type	Scope	Methodology	Number
Liyangunawardena et al. (2013)	MOOC, Massive Open Course and Massively Open Online Course	Academic journals, Google Scholar, ISI Web of Knowledge, ProQuest, JSTOR, IEEE Explorer, and Scopus	Journals, conference publications, academic magazine articles, reports, and workshop presentations	2008–2012	Systematic review	45
Ebben and Murphya (2014)	MOOC	Academic Search Complete, Communication and Mass Media Complete, Directory of Open Access Journals, Education Full Text, ERIC, Google Scholar, MathSciNet, Science Direct, and Web of Science	Academic journals	2009–2013	Systematic review	25
Raffaghelli et al. (2015)	MOOC	SCOPUS, Web of Science (WOS), Educational Resources Information Centre (ERIC) and Directory of Open Access Journals (DOAJ)	Journal article	2008–2014	Systematic review	60
Veletianos and Shepherdson (2015)	MOOC or Massive Open Online Course	Scopus, Journal of Online Learning and Teaching, Summon, Google Scholar, EdITLib Digital Library, and Educause Library	Empirical, English, journal, conference proceedings, or Educause Review	2013–2015	Systematic review	183
Saadatdoost et al. (2015)	MOOC, massive open online courses, xMOOC and cMOOC	Google Scholar, EDUCAUSE, ScienceDirect, Taylor & Francis Online	–	–	Structural literature review	32
Veletianos and Shepherdson (2016)	MOOC or Massive Open Online Course	Scopus, Summon, Google Scholar, EdITLib Digital Library and the Educause Library	Journals, conference proceedings, Educause Review	2013–2015	Systematic methods	183
Bozkurt et al. (2017)	MOOC, MOOCs, Massive Open Online Course and Massively Open Online Course	Multiple academic databases	Journal articles, English	2008–2015	systematic review	362
Duart et al. (2017)	MOOC, Massively Open Online Course, and Massive Open Online Course	Journal Citation Reports (WOS) and Scopus (Scimago) databases	Journal	2013–2015	Systematic review	33
Zheng and Yang (2017)	MOOCs and education	China Academic Journal Network Publishing Database (CAJD)	Journal articles	2013–2016	Bibliometric analysis	445
Deng and Benckendorff (2017)	Massive Open Online Course(s) and MOOC(s) et al	ERIC, Scopus, ProQuest, and Web of Science	Journal articles, conference papers, book chapters, and unpublished theses and dissertations	2014–2016	Systematic review	53

Table 1 (continued)

Author	Keywords	Database	Article type	Scope	Methodology	Number
Zhu et al. (2018a,)	MOOC and Massive Online Open Course(s)	Five key journals in Scopus, and several other targeted journals not indexed by Scopus	Journal articles	2014–2016	Systematic review	146
Moreno-Marcos et al. (2018)	(Predict or prediction or predictive or forecasting) and (MOOC or MOOCs or Massive Open Online Course or Massive Open Online Courses)	Scopus and ISI Web of Knowledge	Conferences and journals articles	2014–2017	Systematic review	88
Deng et al. (2019)	Massive Open Online Course(s), MOOC(s) and names of leading MOOC platforms	ERIC, Google/Google Scholar, ProQuest, Scopus, and Web of Science	Peer-reviewed journals, conference papers and proceedings, book chapters, and unpublished theses and dissertations	2014–2016	Narrative review	102
Rolfe (2015)	MOOC*, cMOOX*, xMOOC*, "massiv* online open course*," "massiv* open online course*," "free online learning," Coursera, MITx, edX, CCK, PLENK, "student experience"; social*; ethic*; inclusiv*; divers*; accessibil*; value*; quality; pedagogy, literacy	Web of Knowledge, PubMed, IEEE Xplore, Scopus, Google Scholar, Journal of Online Learning and Teaching, The International Review of Research in Open and Distance Learning and the European Journal of Open, Distance and eLearning. 'OER' conferences and the 'Annual Open Education Conference'	Journals and conference proceedings	2004–2014	Systematic Review	68

Table 2 Micro-perspective MOOC review studies

Author	Topic	Scope	Methodology	Number	Journal
Joksimović et al. (2018)	Model learning	2012–2015	Systematic review	38	Review of Educational Research
Davis et al. (2018)	Active learning strategies	2009–2017	Systematic review	126	Computers & Education
Costello et al. (2018)	Twitter's use	2011–2017	Systematic review	34	International Journal of Educational Technology in Higher Education
Paton et al. (2018)	Engagement and retention	2013–2017	Systematic review	38	Computers & Education
Fang et al. (2019)	Language learning	2009–2018	Technology-based learning model systematic	33	Interactive Learning Environments
Guajardo Leal et al. (2019)	Academic engagement	2015–2018	Systematic mapping study (SMS)	176	International Review of Research in Open and Distributed Learning
Lee et al. (2019)	Self-regulated learning	2008–2016	Systematic reviews	21	Australasian Journal of Educational Technology
Wong et al. (2019)	Self-regulated learning	2006–2016	Systematic review	35	International Journal of Human–Computer Interaction
Alonso-Mencía et al. (2020)	Self-regulated learning	2014–2017	Systematic reviews	42	Educational Review
Almatrafi and Johri (2018)	Discussion forums	2013–2017	Systematic review	84	IEEE Transactions on Learning Technologies

Table 3 Inclusion and exclusion criteria for manually verifying retrieved publications

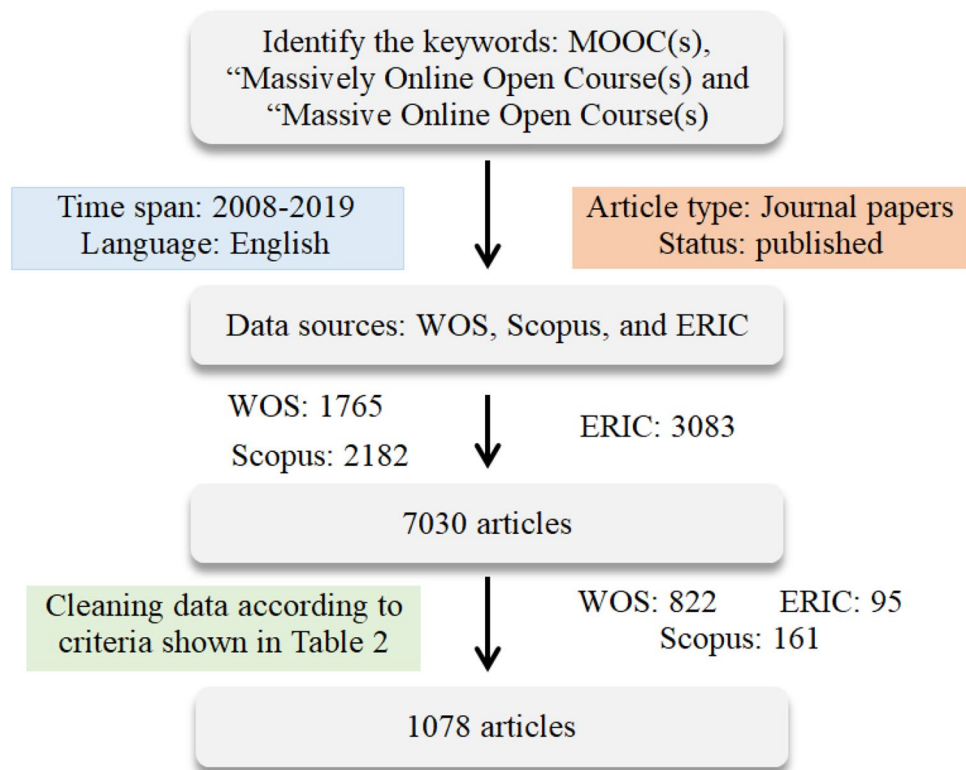
Inclusion criteria	
I1	Empirical study
I2	Tools, design, models, frameworks designed for MOOCs
I3	Research foci on factors in MOOC, such as learners, instructors, assessments, courses, techniques, modality
I4	Variants of MOOC, such as sMOOC, TOOC, and SPOC
I5	MOOC studies based on phenomenological interviews, qualitative interview-type surveys, group discussions, descriptive and qualitative methodologies, content analysis methods, etc.
I6	Case studies or MOOC experience introductions
I7	Education studies with a data source from a set of MOOC platforms or courses
I8	Case studies, such as flipped classrooms and disease advocacy
Exclusion criteria	
E1	Literature review article
E2	Topics related to libraries or library issues
E3	Generalized research on Digital/open/digital/distance /online/blended/flipped education/course/learning
E4	Opinions, impacts, views, and perspectives on MOOCs
E5	Not written in English

body of literature. That is to say, bibliometrics utilizes quantitative analysis and statistics to describe patterns of publication within a given field or body of literature and has been considered as an effective statistical method for evaluating scientific publications (Chen et al., 2018). This paper, therefore, adopted a bibliometrics approach to conduct a qualitative analysis of related MOOC research.

Data collection

The data were collected from a search of journal articles published from 2008 to 2019 in three electronic databases: the Web of Science (WOS), Scopus, and the Education

Resources Information Center (ERIC). The search strings “MOOC(s),” “Massively Online Open Course(s),” and “Massive Online Open Course(s)” were used to screen the titles, abstracts, and keywords, and specific criteria applied to ensure relevance. For example, the selected studies had to be MOOC English language empirical studies in peer-reviewed journals. The specific inclusion and exclusion criteria are listed in Table 3 and the flowchart for the dataset acquisition is shown in Fig. 1.

Fig. 1 Dataset acquisition flowchart

Data analysis

The data search from WOS, Scopus, and ERIC identified 1078 articles with the analysis focused on the five research questions. To answer question #1, the annual number of MOOC empirical articles published between 2008 and 2019 is calculated and the curves of 11 annual numbers fitted to determine the MOOC research rules and trends. To answer question #2, the major contributors to MOOC research were calculated and programmed. To answer question #3, a social network analysis approach (Bastian et al., 2009) was taken to analyze and visualize the collaborative scientific research relationships and the prolific countries/regions and institutions. To answer question #4, structural topic modeling (STM) based on the R package was employed (Chen et al., 2020a, b, c; Roberts et al., 2014a, b) to identify the topics of the 1078 articles from the abstracts. To answer question #5, a graphing tool named Cluster Purity Visualizer (Swamy, 2016) was first implemented to obtain a basic distribution graph of the topic distributions for prolific countries/regions and institutions. Then, the JavaScript packages `d3.v3.js`¹ and `clusterpurityChart.js`² were used to conduct the layout adjustment and coloring of the basic graph.

¹ <https://d3js.org/d3.v3.js>.

² <https://bl.ocks.org/nswamy14/raw/e28ec2c438e9e8bd302f/clusterpurityChart.js>.

Results

The analysis results were displayed by article and citation counts, the prolific countries/regions and institutions, topic identification, trends, and correlations, prolific country/regional and institutional distributions, scientific collaborations, and the annual topic distributions.

Analysis of article counts and citation counts

Figure 2 shows the annual empirical MOOC research counts from 2008 to 2019, from which it can be seen that before 2013, there were significantly fewer MOOC-related articles as MOOC theory was developing at this time; however, from 2013, there was increasing academic interest as MOOC theory was evolving. By 2019, the number of published papers was around three times greater than in 2014. When the annual number of published articles was fitted ($y = 3.305128x^2 - 12,879.64x + 12,939,030$ with $R^2 = 0.974$, $p = 4.605 \times 10^{-7}$), the results showed a parabolic function with the right part of the curve exhibiting a galloping increasing trend.

Table 4 lists the top journals ranked by the H-index. Four bibliometric indicators were employed to evaluate the most prolific countries/regions and institutions: H for Hirsch index (Hirsch 2005); A for article count; C for citation count; and ACP for average citations per article. H

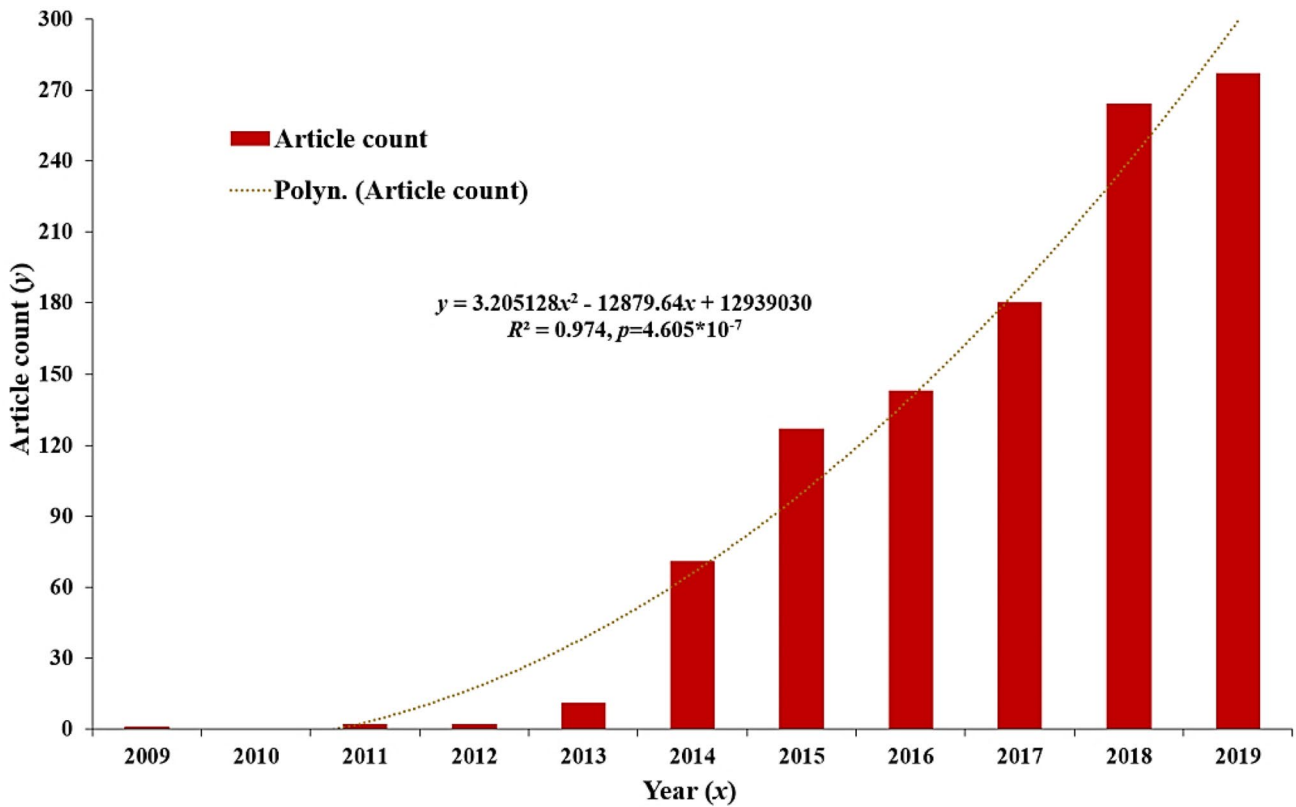


Fig. 2 Annual article counts

Table 4 Top journals ranked by H-index

Journals	H	C	A	ACP
International Review of Research in Open and Distributed Learning	29	4510	68	66.32
Computers & Education	22	3092	32	96.63
British Journal of Educational Technology	13	1127	17	66.29
Computers in Human Behavior	13	1084	17	63.76
International Review of Research in Open and Distance Learning	10	2184	10	218.40
Online Learning	10	369	20	18.45
Distance Education	10	699	14	49.93
Internet and Higher Education	9	1061	10	106.10
Open Praxis	9	247	17	14.53
International Journal of Emerging Technologies in Learning	9	239	34	7.03
Australasian Journal of Educational Technology	9	181	13	13.92
Journal of Computer Assisted Learning	8	558	11	50.73
Journal of Universal Computer Science	8	492	17	28.94
Journal of Computing in Higher Education	8	342	10	34.20

was used to evaluate the quantity and level of academic output, from which it was found that the International Review of Research in Open and Distributed Learning (IRRODL) had the highest MOOC research H-index with H (29), C (4510), and A (68), followed by Computers & Education with H (22), C (3092), and A (32), which indicated that these two journals have had a significant

influence on MOOC research. The International Review of Research in Open and Distance Learning had the highest ACP, and although this journal had only ten MOOC articles, the citation counts were 2184, indicating that these MOOC articles were of high quality and had significant influence. The Internet and Higher Education ranked

Table 5 Top countries/regions ranked by H-index

C/R	H	A	C	ACP
USA	51	266	10,732	40.35
UK	28	103	5125	49.76
Spain	27	116	2633	22.70
China	22	172	1838	10.69
Australia	22	49	2396	48.90
Canada	21	41	3080	75.12
Netherlands	16	37	978	26.43
Taiwan	13	33	747	22.64
Turkey	12	22	614	27.91
Hong Kong	11	18	627	34.83
Mexico	10	19	258	13.58

second in terms of the ACP, followed by Computers & Education.

Prolific countries/regions and institutions

The top 11 countries/regions ranked by the published article numbers are listed in Table 5. The 11 most prolific countries/regions contributed 876 articles or 81.26% of the total (1078). The USA (51), UK (28), and Spain (27) were the most prolific, and China, Australia, and Canada each contributed 22 papers. Canada (75.12) ranked first for the ACP, followed by the UK (49.76), Australia (48.90), and the USA (40.35).

Table 6 shows the top 11 institutions ranked by H-index, which together contributed 15.58% of the total articles. Of these 11 institutions, five were from the USA and two were from the UK. Purdue University (PU), the Massachusetts Institute of Technology (MIT), Pennsylvania State University (PSU), and Harvard University (HU) were the most prolific institutions, further indicating the USA's dominant MOOC research position. MIT, PSU, and HU had the top three H-indices, MIT ranked first for the citation count (2393), followed by HU (2058) and the Open University (OU) (1720), and MIT ranked first for the ACP (about 132.9), followed by HU (about 114.3) and the OU (about 122.9).

Scientific collaborations

Social network analysis (Bastian et al., 2009) was used to visualize the collaborative scientific research relationships between the most prolific countries/regions and institutions. The collaboration networks were built using Gephi,³ open-source software for graph and network analysis. The

Table 6 Top institutions ranked by H-index

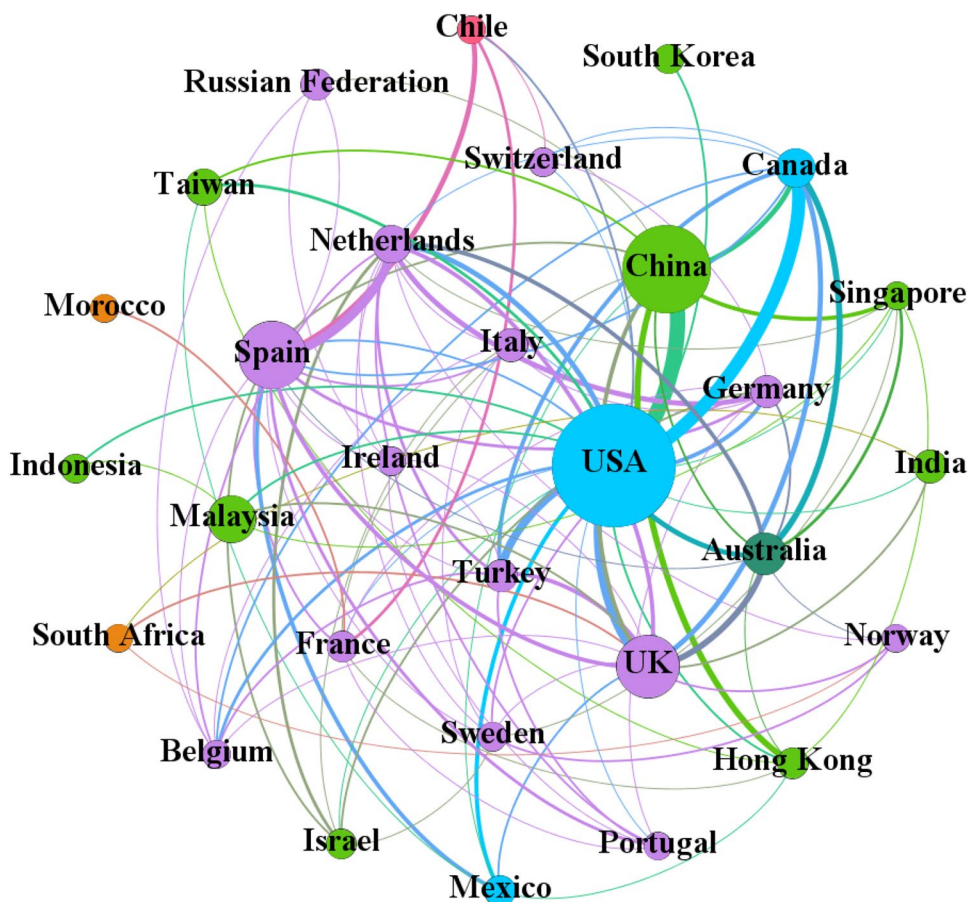
Institutions	C/R	H	A	C	ACP
Massachusetts Institute of Technology	USA	14	18	2393	132.9444
Pennsylvania State University	USA	12	20	860	43
Harvard University	USA	11	18	2058	114.3333
Universidad Carlos III de Madrid	Spain	11	18	599	33.27778
Open University of the Netherlands	Netherlands	10	15	423	28.2
Purdue University	USA	10	21	556	26.47619
University of Edinburgh	UK	10	12	390	32.5
Anadolu University	Turkey	9	11	490	44.54545
Duke University	USA	9	10	477	47.7
The Open University	UK	9	14	1720	122.8571
Pontificia Universidad Catolica de Chile	Chile	9	11	680	61.81818

analysis was conducted in three steps. First, the input data covering a node sheet and an edge sheet were prepared. The node sheet had four columns: id number, label for countries/regions and institutions, group for indicating continent of countries/regions, or countries/regions of institutions and authors, and article count size; and the edge sheet had three columns: the source and the target for the corresponding co-authorship pairs, and the weight of the collaborative article numbers. Second, the node and edge sheets were used to visualize the co-authorship network using the Fruchterman Reingold algorithm. Finally, the node size and node color were configured based on the article count and group data. The countries/regions and institutions were represented using different node sizes and colors, with the node size denoting the corresponding article number, and the node color indicating the continents to which the corresponding country/region belonged.

Figure 3 shows the collaborative network for the 30 most prolific countries/regions, each of which had greater than ten published articles. The collaborative network had 30 nodes and 112 links. It can be seen from the node size that the USA had the largest number of articles (266) and had collaborated with 23 countries/regions, followed by China (172), Spain (116), and the UK (103), each of which had respectively collaborated with 11, 18, and 15 countries/regions. The USA collaborated with the most countries/regions, followed by Spain, the UK, the Netherlands, China, Australia, and Germany, with the number of collaborated articles being, respectively, 87, 46, 48, 43, 43, 31, and 22. Therefore, the USA collaborated on the most articles, with China and Canada being the most important partners with 15 and 13 articles. China collaborated mostly with the USA, followed by Hong Kong and Canada. Spain collaborated

³ <https://gephi.org/>.

Fig. 3 Collaborations between the 30 most prolific countries/regions



with 18 countries/regions; nine articles with the Netherlands and five with Chile; and in addition to the countries already mentioned, Canada, Turkey, France, Sweden, and Belgium also closely collaborated with other countries/regions.

The collaborative scientific research relationships between the 37 most prolific institutions are illustrated in Fig. 4. The most prolific institute was the University of Technology Malaysia (UTM) with 26 articles, followed by PU (21), PSU (20), MIT (18), University Carlos III of Madrid (UCM) (18), and HU (18), each of which had respective collaborations of 0, 4, 1, 10, 10, and 6 articles. Therefore, although the UTM ranked first for the number of published articles, it had no collaborative relationships with other institutes. Of the 37 most prolific institutions, the University of Edinburgh (UE) collaborated with the most institutions (6), with the University of South Australia (UniSA) being its main partner for five of these six articles.

Topic identification, trends, and correlations

Semantic coherence is based on the frequency of individual words and the co-occurrence of the frequency of different word pairs and is maximized when the most probable words in a given topic frequently co-occur (Silge, 2018). If words

have a high probability of appearing in a topic and a low probability of appearing in other topics, the corresponding topic is considered exclusive (Kuhn, 2018). Figure 5 shows the semantic coherence and exclusivity scores for 26 topics with the topic numbers ranging from five to 30. In the figure, each point represents a model with its name and indicates how many topics were considered. For example, the point labeled "15-topic model" represents a model fitted with 15 topics. It can be seen that 14 and 15 topics achieved higher semantic coherence and exclusivity values, which indicated that more potential terms within the topic occurred in the same document and more terms were exclusively affiliated with the single topic. Two domain experts independently compared models with different numbers of topics by inspecting the representative terms and articles (Jiang et al., 2018), and finally a 15-topic model was identified for the qualitative evaluation as this number of topics was found to have the greatest semantic consistency within the topics and exclusivity between the topics. Based on the estimated article-topic and topic-term distributions, the probability of an article or term belonging to a topic was determined, with the most representative articles and terms in a single topic receiving the highest assignment probabilities.

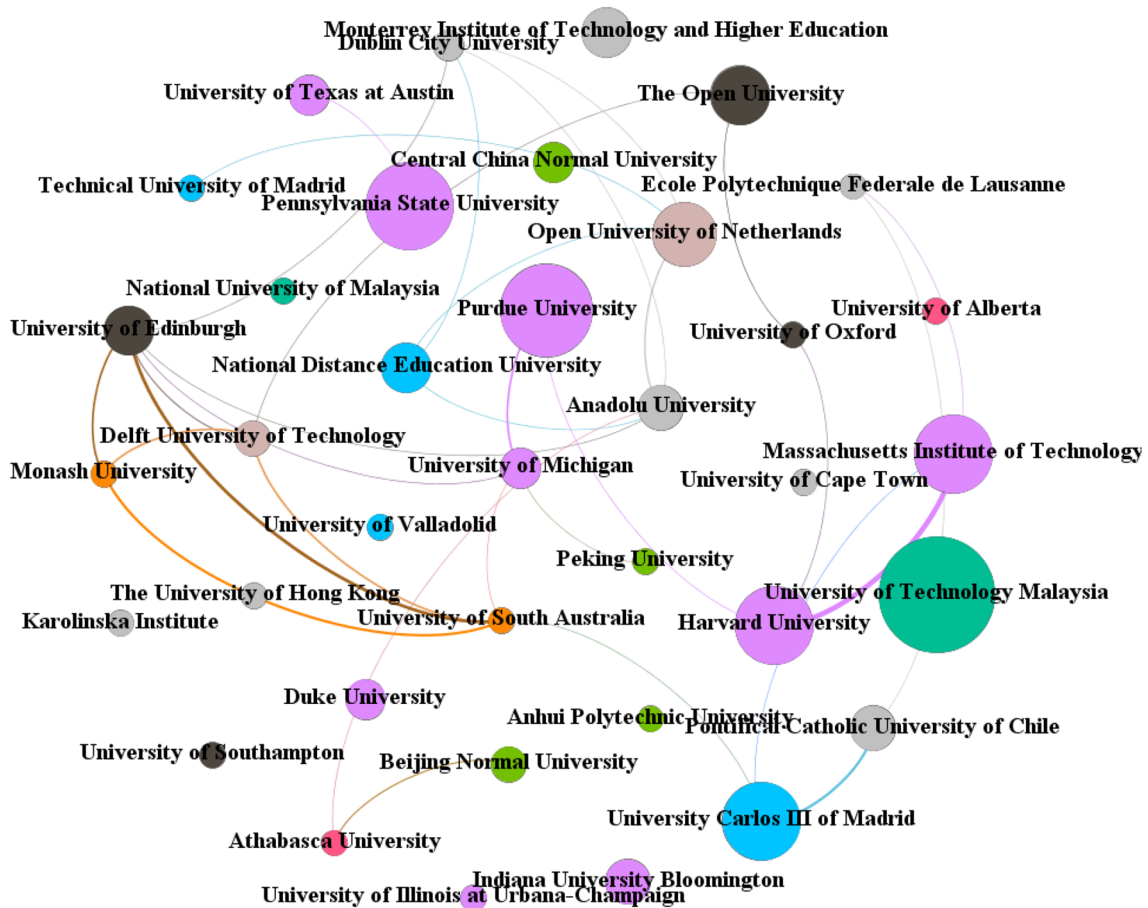


Fig. 4 Collaborations between the 37 most prolific institutions

Table 7 shows the 15-topic STM analysis results with the representative terms, the topic proportions within the whole corpus, the suggested topic labels, and the topical trends. The six most-discussed topics were educational data mining and visualization (10.51%); cMOOCs and healthcare MOOCs (8.37%); MOOCs for languages (7.64%); demographic features of MOOC learners (7.62%); peer and formative assessment (7.30%); and flipped learning for MOOCs (7.03%). In Table 7 $\uparrow(\downarrow)$ indicates not significant ($p > 0.05$) increasing (decreasing) trends, and $\uparrow\uparrow(\downarrow\downarrow)$, $\uparrow\uparrow\uparrow(\downarrow\downarrow\downarrow)$, and $\uparrow\uparrow\uparrow\uparrow(\downarrow\downarrow\downarrow\downarrow)$ indicate significant increasing (decreasing) trends at, respectively, $p < 0.05$, $p < 0.01$, and $p < 0.001$. Therefore, educational data mining and visualization, learner perceptions and satisfaction, business and entrepreneurship for MOOCs, and SRL had significantly increasing trends while the remainders were not significantly increasing and some were decreasing. MOOCs for languages, regional and local MOOC practices and research, flipped learning for MOOCs, teacher education, course gamification and recommendations, peer and formative assessments, and xMOOCs were found to have increasing trends, but MOOCs for institutions, demographic features of MOOC learners, semantic

data and finance MOOCs, and cMOOCs and healthcare MOOCs were found to have decreasing trends.

In Table 7, educational data mining and visualization had the most significant increasing trend. To mine specific, deeper content on this topic, the representative terms in the dataset were further analyzed, from which it was statistically found that the educational data mining and visualization topic was focused on three main factors: analytics (analysis), behavior, and prediction. The related analytics (analysis) factors could be divided into two: the technique, methodology or tools, such as big data analyses and qualitative and quantitative analyses; and research content, such as study pattern analysis and video learning analytics. In terms of behavior, the studies showed a particular interest in learner community behaviors, learning behaviors, behavior modeling, and video-watching behaviors. As to the prediction, the related research included student dropout and performance predictions, learning behavior predictions, predictive analytics, retention rate predictions, grade predictions, and student retention predictions. Increasingly, more research was focused on data

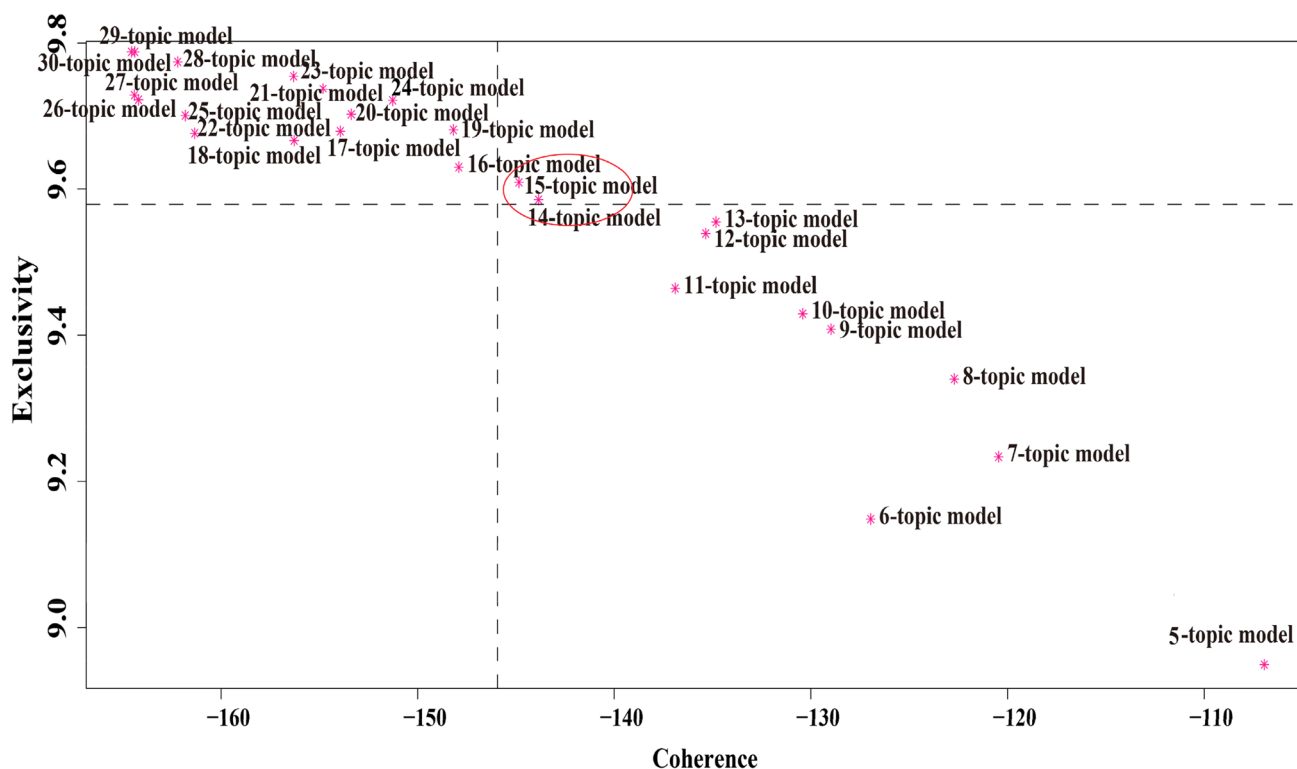


Fig. 5 Semantic coherence and exclusivity in the MOOCs-related topics

mining to analyze and predict MOOC-related aspects and visualize the corresponding results.

The annual proportions within the whole corpus of identified topics are visualized in Fig. 6. The first focus was on the topic evolution, which was identified using a topic model. The main significantly increasing trends were learner perceptions and satisfaction, educational data mining and visualization, and SRL. The evolution curve for peer and formative assessment reached a peak in 2013, which indicated that this MOOC topic had attracted the most research attention in 2013 but had fallen out of favor by 2015. Research interest in MOOCs for institutions, the demographic features of MOOC learners, business and entrepreneurship for MOOCs, and MOOCs for institutions first fell, then rose, and then fell again. There were two distinct peaks for flipped learning for MOOCs, MOOCs for language, and xMOOCs in 2013 and 2018, 2013 and 2015, and 2013 and 2017.

Topic distributions

Figure 7 shows the topic distributions for the top nine countries/regions and institutions ranked by the H-index and the annual topic distributions. Figure 7a shows the particular research topics for each prolific country/region or institution. Educational data mining and visualization was the most active

topic in the USA, at UCM, and at PSU. The research interest in the UK, Canada, and Turkey was cMOOCs and healthcare MOOCs, in China was flipped learning for MOOCs, in Taiwan was learning perceptions and satisfaction, in the Netherlands and PU was teacher education, and in Spain, the OU of the Netherlands, and Anadolu University (AU) was the demographic features of MOOC learners.

The annual topic distributions are shown in Fig. 7c, in 2009, the demographic features of MOOC learners had the greatest research focus; in 2011, cMOOCs and healthcare MOOCs were the most popular topics; and in 2012, the demographic features of MOOC learners and semantic data and finance MOOCs were the most popular.

Discussion

This review examined 1078 studies to reveal the interesting trends and hidden relationships in MOOC research up to 2019. Research methods and data collection methods were examined using descriptive and quantitative statistics, which included analyses of article counts, citation counts, prolific countries/regions and institutions, scientific collaborations, topic identification, trends and correlations, prolific country/regional and institutional topic distributions, and annual topic distributions. These findings provide important

Table 7 Topic labels with their representative terms, proportions in the whole corpus, and trends

Label	Representative terms	%	<i>p</i>	trend
Educational data mining and visualization	Visual, prediction, analytics, temporal, clustering, thread, log, behavior, visualization, predictive	10.5	0.000677	↑↑↑↑
cMOOCs and healthcare MOOCs	Death, VPS, patient, LCA, crowd workers, healthcare, virtual, VP, cMOOC, community	8.37	0.858	↓
MOOCs for languages	Teacher, language, ESL, English, polytechnic, blend, Sino-foreign, curriculum, speaker, Japanese	7.64	0.2105	↑
Demographic features of MOOC learners	Woman, competence, men, single-gender, soft, identity, didactic, adolescent, trait, region	7.62	0.7205	↓
Peer and formative assessment	Essay, peer, grading, assignment, feedback, mini-MOOC, grader, assessment, formative, automated	7.30	0.858	↑
Flipped learning for MOOCs	Flipped, teaching, mode, MOOC-based, reform, component, autonomous, classroom, construction, ability	7.03	0.07364	↑
Learner perceptions and satisfaction	Continuance, intention, acceptance, perceived, ease, usefulness, satisfaction, expectancy, tam, structural	6.73	0.004208	↑↑↑
Self-regulated learning	Self-regulated, SRL, motivation, engagement, self-efficacy, motivational, retention, persistence, achievement, intrinsic	6.76	0.0491	↑↑
Teacher education	Welfare, attitudinal, change, instructional, lesson, facilitation, dissonance, impactful, scope, learned	6.48	0.1524	↑
Business and entrepreneurship for MOOCs	Business, management, system, efficiency, cloud, embedded, MyMOOCspace, entrepreneurship, remote, innovation	6.42	0.01227	↑↑
Course gamification and recommendation	Gamification, success, indicator, cost, pre-course, accessed, subsequent, subgroup, recommendation, game	5.79	0.2831	↑
xMOOCs	Antibiotic, stewardship, crowdsourced, school, cultural, market, xMOOCs, act, health, efficacy	4.94	0.2831	↑
Regional and local MOOC practices and research	Malaysian, readiness, wearable, mandarin, Malaysia, culture, smart, city, administrator, interface	4.89	0.4743	↑
MOOCs for institutions	Registered, volunteer, astronomy, transactional, rated, iMOOC, trainee, unit, facilitator, lecture	4.82	0.7205	↓
Semantic data and finance MOOCs	MCQS, emotional, flaw, borrower, query, vocabulary, credit, financial, semantic, verbal	4.71	1	↓

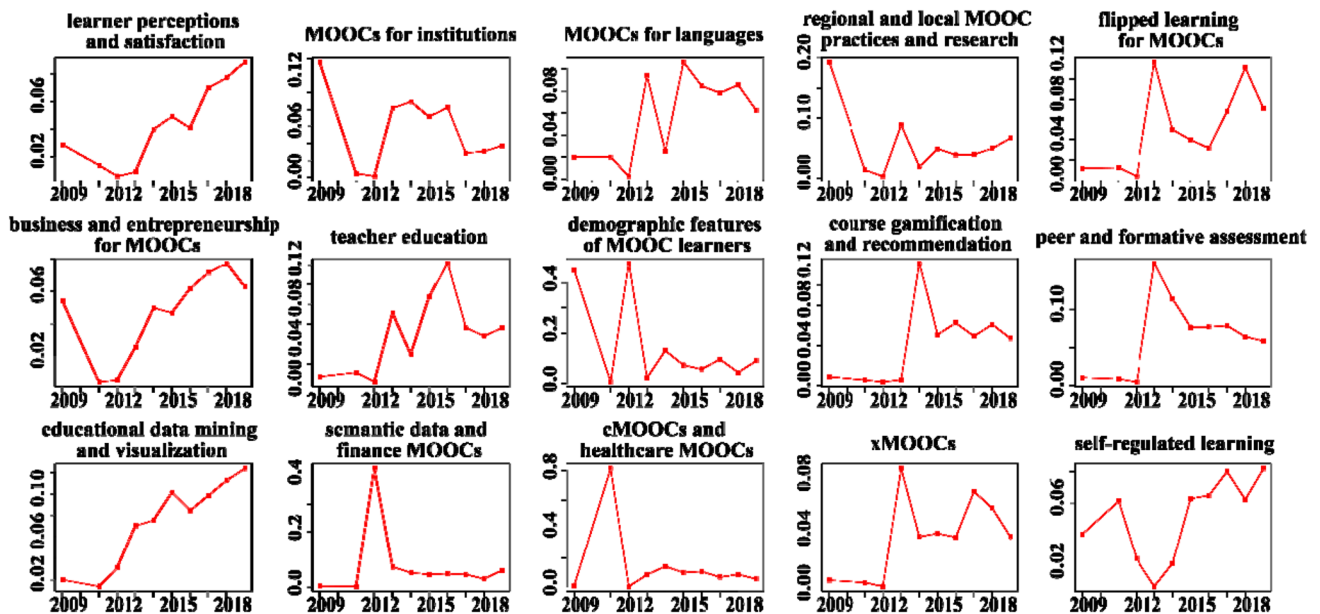


Fig. 6 Annual topic proportions within the whole corpus for 15 topics

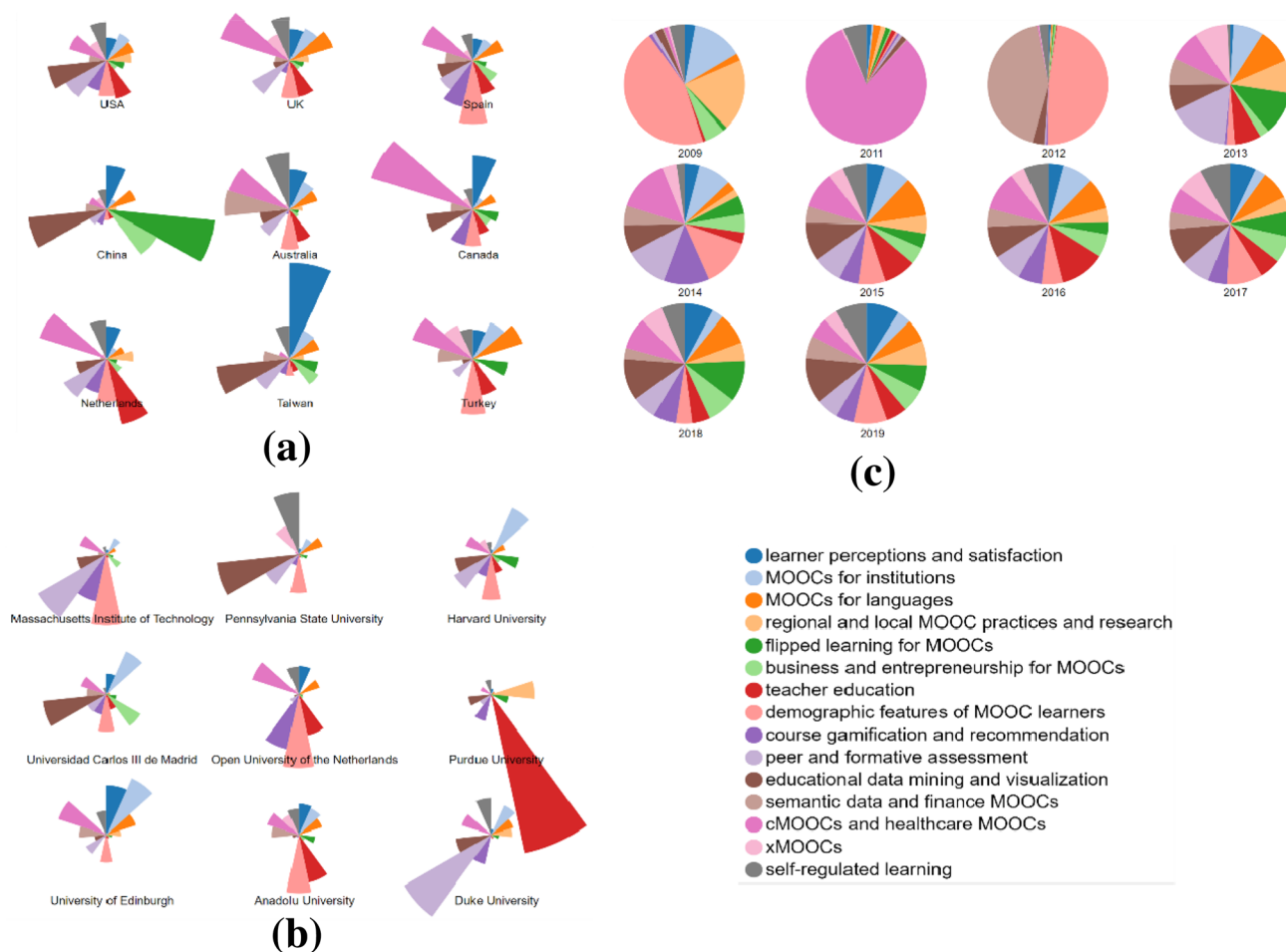


Fig. 7 Topic proportion distributions by prolific countries/regions, institutions, and years

information to enhance MOOC researchers' understanding of current MOOC research status and trends.

Representative research work

The representative empirical MOOC research articles revealed the MOOCs research trends; therefore, to better understand each topic, in this section, the most representative research work in each topic is further analyzed.

The research on educational data mining and visualization was mainly focused on using educational data mining techniques to predict, analyze, or explore the issues related to MOOCs, such as academic performances or behaviors. For example, An et al. (2019) explored the learning resource mention identification in MOOC forums using an LSTM-CRF model and evaluated the strategies using a dataset from the Coursera online forum. This paper provided solutions to identifying resource mentions for real learning resources and demonstrated a classic educational data mining research mode. The research focus for MOOCs for languages was

on the language learning MOOC users or the courses. For example, Mustikasari (2017) used a descriptive qualitative approach to investigate MOOC English teaching materials and the professional teaching development provided by joining MOOC, concluding that developing a MOOC for Madrasah English teachers was challenging and providing suitable teaching materials was vital; therefore, this paper was useful in highlighting the importance of MOOC materials and the willingness of English teachers to develop MOOCs, and was beneficial to MOOC language research. Peer and formative assessment focused on MOOC peer reviews, assessments, and assessment tool development. Meek et al. (2017) discussed MOOC peer reviews by investigating student participation, performance, and opinions in a MOOC peer-review task by evaluating student topic summary data using a qualitative peer-review process that compared the summarizes to student demographic data and performance, and found that the student opinions regarding the usefulness of the peer-review tasks were mixed, concluding that instructional design strategies were needed to improve

the usefulness of peer-review tasks. Flipped learning for MOOCs was focused on the development, applications of MOOCs or the effectiveness in the flipped classroom. For example, Zhu et al., (2018a, b) developed a small private online course-based flipped classroom teaching model that was driven by curriculum ontology, which they applied to a teaching plan and verified in the Electronic Commerce MOOC, which is a valuable reference for hybrid teaching. Learner perceptions and satisfaction research has mainly tended to examine the perceived behaviors, satisfaction, and intentions associated with MOOC use; for example, Wu and Chen (2016) used a framework that integrated a technology acceptance model and a task fit technology model to examine the factors influencing MOOC adoption and investigate MOOC continuance intentions. Therefore, this focus assisted researchers to gain a better understanding of learner perceptions and satisfaction. SRL research was focused on how learners guide their learning in terms of effectiveness, strategies, etc., in the MOOC learning environment. For example, Onah and Sinclair (2017) investigated and assessed SRL using a MOOC platform (eLDa) to compare self-directed learning and instructor-led learning, and concluded that self-directed learning was able to provide learners with better SRL skills. Business and entrepreneurship for MOOCs research has tended to focus on entrepreneurship and business courses, with most articles using empirical cases to design entrepreneurship MOOCs or verify suitable MOOC platforms to teach or develop entrepreneurship. To understand how the inclusion of issues related to entrepreneurship in MOOCs could positively impact participants, Beltrán Hernández de Galindo et al. (2019) analyzed the incorporation of entrepreneurial competencies in MOOCs to develop educational innovation and collaborative project attributes and investigated whether MOOC discussion forum interactions had resulted in entrepreneurial opportunities. Teacher education research has looked at various elements associated with teacher development. For example, Kennedy and Laurillard (2019) examined the use of co-design models in MOOC projects to deliver teacher professional development (TPD) and developed a ToC model that could be applied to TPD for mass displacement, which could assist in the professional development needs of MOOC teachers with MOOC in mass displacement. Course gamification and recommendation research has mainly examined the elements associated with MOOC course improvements, such as gamification and recommendations. For example, in a classic research case focused on using information technologies to generate content recommendations, Pang et al. (2018) proposed an adaptive recommendation for MOOC that had scoring and learning durations as features and combined collaborative filtering techniques and time series to improve recommendation accuracy, which better satisfied the learners and reduced dropouts. xMOOC research has focused on the development

of evaluation criteria. For example, Nkuyubwatsi (2016) examined the learning materials, activities, assessments, and scalability in ten xMOOCs, the findings from which could inform open education policies and practices. Regional and local MOOC practices and research has generally focused on empirical or case studies in a specific region. For example, Aljaraideh (2019) conducted a case study on respondents from universities in Jordan to explore the challenges and benefits of using MOOCs in higher education, identified the possible barriers to MOOCs at Jerash University, and found there was general acceptance by faculty that the MOOCs would be an advantage for users. cMOOCs and healthcare MOOCs research is specifically focused on cMOOCs and healthcare education. For example, Li et al. (2016) analyzed the content of messages posted by learners and instructors in online course learning spaces for a case study, the findings from which provided valuable information on student difficulties and needed support strategies for cMOOC learning. Research on the demographic features of MOOC learners has examined the specific characteristics of MOOC learners. For example, Lee and Chung (2019) analyzed K-MOOC learner data: number of participants, average completion rate, and participant backgrounds, provided by the National Lifelong Learning Agency and compared Korea's K-MOOC and the United States' edX. Research on MOOC learners' demographic features can reveal the current state of MOOC programs and address possible issues. The MOOCs for institution research focus have been on the development or application of MOOCs in some institutions and institutional cooperation. For example, Glencross and St Denny (2017) investigated the MOOC application for voting in the UK referendum on EU membership, which contributed to the public understanding of and engagement with EU-related politics and policy issues. With a focus on semantic data and finance MOOCs, Siddike et al. (2017) explored current microfinance MOOC education using a semi-structured interview research strategy, identified the current advantages and possible drawbacks for the adoption of MOOCs for microfinance education, and presented a MOOC framework to offer financial literacy to the poor, with the main findings being able to be extended to other courses.

Research implications

The STM analysis provided future MOOC topic directions. Because of the growth in big data, the most potential new topic is educational data mining (EDM) and visualization. EDM is the analysis of various types of educational data by using statistical, machine learning, and deep learning algorithms (Chen et al., 2020a, b, c; Romero & Ventura, 2010). EDM and analysis have brought new ways to solve long-term research problems in the field of traditional educational technology. MOOCs are able to continuously record

all static and dynamic data throughout the entire teaching activity, such as the number of logins, interactive responses, and the time cost of learning each video, without affecting the activities of either the teachers or the students. Therefore, MOOCs provide effective big data for EDM. EDM harnesses the power of emerging artificial intelligence technologies (i.e., machine learning and neural networks) to mine the MOOCs' big data (i.e., logs) (Chen et al., 2020a, b, c) and conduct practical educational assessments, predictions, and interventions, research employing the MOOC data is expected to remain a research hotspot. However, as learners work directly with the MOOC platforms, their satisfaction is a significant factor affecting the continuous use of such platforms (Lu et al. 2019); therefore, to improve service quality, improve evaluation systems, and enhance teaching quality, it is expected that research into learner perceptions and satisfaction will continue to be important.

Based on the statistics shown in Table 7, another potential hot topic in the future is SRL, which is how learners can become masters of their own learning (Zimmerman & Schunk, 2012). It is an internal mechanism that is composed of learners' attitudes, abilities, and learning strategies. Self-regulated students have been found to select and use self-regulating learning strategies to achieve their desired academic outcomes on the basis of feedback about learning effectiveness and skills (Zimmerman, 1990). SRL also has a profound influence on the way of teachers interact with the students and learning content organization (Yang, 2020). Because of MOOCs' time, space, supervision, and management constraints, it is particularly important to ensure students have self-regulating learning abilities. Therefore, exploring the issues surrounding these skills when approaching MOOC learning and content development strategies (e.g., video production, classroom management, and organizational forms) could improve these SRL abilities.

The global changes in the research trends inferred from the annual topic distributions and article counts can help researchers assess how government policies, technological developments, and major life changes impact and drive change in the topic of research. In the first two or three years since MOOCs were first proposed in 2008, MOOCs have been in the stage of exploration and development. The emergence of cMOOCs gave rise to innovative pedagogical and technical approaches (Ebben & Murphy, 2014), which then attracted focused research. As shown in this study, the cMOOCs and healthcare MOOCs ranked first in the 2009–2012 research stage, followed by the demographic features of MOOC learners and semantic data and finance MOOCs. The increasing globalization of health care has highlighted the inadequacy of many health care services around the world; therefore, to meet this need, health care education needs to have an international perspective (Hovenga, 2004). Online courses such as MOOCs provide a

useful platform for the delivery of this type of healthcare education, which is why medical MOOC development has been increasing rapidly, especially in developed countries such as Australia, Canada, and the United Kingdom (Maxwell et al., 2018). For example, MEDU (www.med-u.org) has hundreds of medical educators and offers virtual patient content. Accordingly, MOOC research on medical education course development, implementation, engagement outcomes, and other practical considerations has been a popular research area since 2009 and since 2013, the number of MOOC studies has increased significantly. Besides the increase in the cMOOC and healthcare MOOC research, from 2013 to 2016, teacher education and educational data mining and visualization research were the second and third most researched areas. In October 2012, the US Department of Education published "Enhancing Teaching and Learning through Educational Data Mining and Learning Analysis," which pointed out that the mining and analysis of educational big data could promote teaching system reform at US colleges, universities and K-12 schools, educational data mining (EDM) has received wide attention, and the research into EDM and visualization ranked first from 2017 to 2019. Research into learner perceptions, satisfaction, and SRL also received significant attention during this period, indicating that the needs, feelings, and experiences of learners and MOOC learning methods were being seen more important, which was consistent with the growth in person-centered or student-centered education (Zuconi, 2016) and the use of smart technology to facilitate interactive teaching and learning (Leverage Edu, 2020).

The identification of the scientific collaborations and topic distributions could also assist MOOC researchers to find research partners and funding. For example, the researchers at PU have been primarily focused on teacher education. The university launched a teacher education program (TEP) that includes online Master's degree programs and online certificate programs for teachers and potential teachers. They may provide more practical experience in teacher education in scientific collaboration. For example, Watson from PU provided some research on MOOCs' attitudinal learning in MOOCs by examining the instructors' attitudinal dissonance (Watson et al., 2017) and the learner's attitudinal change in a MOOC (Watson & Kim, 2016). PSU and UCM have had a greater research focus on education data mining and visualization. For example, Wong et al. (2015) from PSU used a keyword taxonomy approach to analyze large quantities of MOOC forum data and identify the types of learning interactions taking place in forum conversations, and Moreno-Marcos et al. (2018) from the UCM analyzed the predictive power for anticipating assignment grades in a MOOC. MOOC research from MIT, HU, the OU of the Netherlands, and AU has mainly focused on the demographic features of MOOC learners. For

example, Hansen from HU and Reich from MIT collaborated to analyze course participant features, such as economic background, education, and age using data from 68 MOOCs offered by HU and MIT between 2012 and 2014 (Hansen & Reich, 2015). DU and MIT collaborated to research MOOC assessment; for example, Comer and White (2016) from DU designed and deployed an English MOOC writing assessment course, concluding that writing assessment could be effectively adapted to the MOOC environment. The UE has published research on MOOCs for institutions, learner perceptions and satisfaction, and cMOOCs and healthcare MOOCs. For example, Skrypnik, (2015) analyzed the roles of course facilitators, learners, and technology in the flow of information in a cMOOC, and Murray (2014) examined participant perceptions toward a MOOC conducted by UE.

The H-index ranking of the top journals for MOOC research identified the main journals to be the IRRODL, Computers & Education, the British Journal of Educational Technology, Computers in Human Behavior, the International Review of Research in Open and Distance Learning, and Online Learning, and Distance Education.

Conclusions

Since its emergence in 2008, MOOCs have become a popular education-focused research topic. In particular, the online education demand generated by the global COVID-19 pandemic has elevated MOOCs to the forefront of education delivery. This study examined 1078 empirical MOOC articles published between 2008 and 2019 with the aim of assisting MOOC researchers to gain a deeper and more diverse understanding of the current MOOC research foci, trends, and hidden relationships by analyzing the annual article and citation counts, the most prolific countries, regions, institutions, and scientific collaborations, etc. This review provides researchers and educators with a detailed and comprehensive picture of the MOOC research trends and topics up to 2019, which could help them to build upon MOOC studies, address novel and popular topic areas, and find collaborative research partners.

However, as this review only focused on the empirical articles published before 2019, there have been many more MOOC focused papers published in 2020, and therefore, further systemic analyses of MOOCs research methods and research topics will be conducted in future research.

Funding This research is supported by the FLASS Internationalization and Exchange Scheme (FLASS/IE_A03/18-19) at The Education University of Hong Kong, and the Teaching Development Grant (102489) at Lingnan University, Hong Kong.

Data availability Requests for the data can be addressed to the corresponding author.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Approval for conducting this research was received from the anonymous organization.

References

- Aljaraidh, Y. (2019). Massive open online learning (MOOC) benefits and challenges: A case study in Jordanian context. *International Journal of Instruction*, 12(4), 65–78. <https://doi.org/10.29333/iji.2019.1245a>.
- Almatrafi, O., & Johri, A. (2018). Systematic review of discussion forums in massive open online courses (MOOCs). *IEEE Transactions on Learning Technologies*, 12(3), 413–428. <https://doi.org/10.1109/TLT.2018.2859304>.
- Alonso-Mencía, M. E., Alario-Hoyos, C., Maldonado-Mahauad, J., Estévez-Ayres, I., Pérez-Sanagustín, M., & Delgado Kloos, C. (2020). Self-regulated learning in MOOCs: Lessons learned from a literature review. *Educational Review*, 72(3), 319–345. <https://doi.org/10.1080/00131911.2019.1566208>.
- An, Y. H., Pan, L., Kan, M. Y., et al. (2019). Resource mention extraction for MOOC discussion forums. *IEEE Access*, 7, 87887–87900. <https://doi.org/10.1109/ACCESS.2019.2924250>.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *The third international ICWSM conference* (Vol. 3, pp. 361–362). California: Association for Computational Linguistics. <https://www.aaii.org/ocs/index.php/ICWSM/09/paper/view/154>.
- Boyatt, R., Joy, M., Rocks, C., & Sinclair, J. (2014). What (use) is a MOOC? *The 2nd international workshop on learning technology for education in cloud* (pp.133-145). Berlin: Springer
- Bozkurt, A., Akgün-Özbek, E., & Zawacki-Richter, O. (2017). Trends and patterns in massive open online courses: Review and content analysis of research on MOOCs (2008–2015). *International Review of Research in Open and Distributed Learning: IRRODL*, 18(5), 118–147. <https://doi.org/10.19173/irrodl.v18i5.3080>.
- Chen, X., Hao, J., Chen, J., Hua, S., & Hao, T. (2018). A Bibliometric Analysis of the Research Status of the Technology Enhanced Language Learning. In *International symposium on emerging technologies for education* (pp.169–179). Berlin: Springer.
- Chen, X., Xie, H., & Hwang, G. J. (2020a). A multi-perspective study on artificial intelligence in education: Grants, conferences, journals, software tools, institutions, and researchers. *Computers and Education: Artificial Intelligence*. <https://doi.org/10.1016/j.caeai.2020.100005>.
- Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020b). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education Artificial Intelligence*. <https://doi.org/10.1016/j.caeai.2020.100002>.
- Chen, X., Zou, D., Cheng, G., & Xie, H. (2020c). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of computers & education. *Computers & Education*, 151, 1–21. <https://doi.org/10.1016/j.compedu.2020.103855>.

- Comer, D. K., & White, E. M. (2016). Adventuring into MOOC writing assessment: Challenges, results, and possibilities. *College Composition and Communication*, 67(3), 318–359.
- Costello, E., Brown, M., Mhichil, M., & Zhang, J. (2018). Big course small talk: Twitter and MOOCs—A systematic review of research designs 2011–2017. *International Journal of Educational Technology in Higher Education*, 15, 1–16. <https://doi.org/10.1186/s41239-018-0127-9>.
- Davis, D., Chen, G., Hauff, C., & Houben, G. J. (2018). Activating learning at scale: A review of innovations in online learning strategies. *Computers & Education*, 125, 327–344. <https://doi.org/10.1016/j.compedu.2018.05.019>.
- De Waard, I., Koutropoulos, A., Hogue, R. J., Abajian, S. C., Keskin, N. Ö., Rodriguez, C. O., & Gallagher, M. S. (2012). Merging MOOC and mLearning for increased learner interactions. *International Journal of Mobile and Blended Learning*, 4(4), 34–46. <https://doi.org/10.4018/jmbl.2012100103>.
- Deng, R., & Benckendorff, P. (2017). A contemporary review of research methods adopted to understand students' and instructors' use of massive open online courses (MOOCs). *International Journal of Information and Education Technology*, 7(8), 601–607. <https://doi.org/10.18178/ijiet.2017.7.8.939>.
- Deng, R., & Benckendorff, P. (2021). What are the key themes associated with the positive learning experience in MOOCs? An empirical investigation of learners' ratings and reviews. *International Journal of Educational Technology in Higher Education*, 18(1), 1–28. <https://doi.org/10.1186/s41239-021-00244-3>.
- Deng, R., Benckendorff, P., & Gannaway, D. (2019). Progress and new directions for teaching and learning in MOOCs. *Computers & Education*, 129, 48–60. <https://doi.org/10.1016/j.compedu.2018.10.019>.
- Dodson, M. N., Kitburi, K., & Berge, Z. L. (2015). Possibilities for MOOCs in corporate training and development. *Performance Improvement*, 54(10), 14–21. <https://doi.org/10.1002/pfi.21532>.
- Downes, S. (2008). Places to go: Connectivism and connective knowledge. *Innovate: Journal of Online Education*, 5(1), 1–6.
- Duart, J. M., Roig-Vila, R., Mengual-Andres, S., & Durán, M. Á. M. (2017). The pedagogical quality of MOOCs based on a systematic review of JCR and Scopus publications (2013–2015). *Revista española de pedagogía*, 75(266), 29–46. <https://doi.org/10.22550/rep75-1-2017-09>.
- Ebben, M., & Murphy, J. S. (2014). Unpacking MOOC scholarly discourse: A review of nascent MOOC scholarship. *Learning, Media and Technology*, 39(3), 328–345. <https://doi.org/10.1080/17439884.2013.878352>.
- Leverage Edu. (2020). Modern education: A significant leap forward. Retrieved October 29, 2020, from <https://leverageedu.com/blog/modern-education/>.
- Fang, J. W., Hwang, G. J., & Chang, C. Y. (2019). Advancement and the foci of investigation of MOOCs and open online courses for language learning: A review of journal publications from 2009 to 2018. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2019.1703011>.
- Fink, A. (2010). *Conducting research literature reviews: From internet to paper*. (3rd ed.). Thousand Oaks, CA: Sage.
- Glencross, A., & St Denny, E. (2017). Remain or leave? Reflections on the pedagogical and informative value of a Massive Open Online Course on the 2016 UK referendum on EU membership. *Journal of Contemporary European Research*, 13(4), 1422–1436.
- Guajardo Leal, B. E., Navarro-Corona, C., & Valenzuela González, J. R. (2019). Systematic mapping study of academic engagement in MOOC. *International Review of Research in Open and Distributed Learning*, 20(2), 113–139. <https://doi.org/10.19173/irrodl.v20i2.4018>.
- Hansen, J. D., & Reich, J. (2015). Democratizing education? Examining access and usage patterns in massive open online courses. *Science*, 350(6265), 1245–1248. <https://doi.org/10.1126/science.aab3782>.
- Hernández, B., de Galindo, M. D. J., Romero-Rodríguez, L. M., & Ramirez Montoya, M. S. (2019). Entrepreneurship competencies in energy sustainability MOOCs. *Journal of Entrepreneurship in Emerging Economies*, 11(4), 598–616. <https://doi.org/10.1108/JEEE-03-2019-0034>.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572. <https://doi.org/10.1073/pnas.0507655102>.
- Hovenga, E. J. (2004). Globalisation of health and medical informatics education—what are the issues? *International Journal of Medical Informatics*, 73(2), 101–109. <https://doi.org/10.1016/j.ijmedinf.2003.11.004>.
- Howkins, D. T. (1981). Unvocational used of online information retrieval systems: Online bibliometric study. *Journal of American Society for Information Science*, 28(1), 13–18.
- Jiang, H., Qiang, M., Fan, Q., & Zhang, M. (2018). Scientific research driven by large-scale infrastructure projects: A case study of the Three Gorges Project in China. *Technological Forecasting and Social Change*, 134, 61–71. <https://doi.org/10.1016/j.techfore.2018.05.012>.
- Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., Dawson, S., Graesser, A. C., & Brooks, C. (2018). How do we model learning at scale? A systematic review of research on MOOCs. *Review of Educational Research*, 88(1), 43–86. <https://doi.org/10.3102/0034654317740335>.
- Kennedy, E., & Laurillard, D. (2019). The potential of MOOCs for large-scale teacher professional development in contexts of mass displacement. *London Review of Education*, 17(2), 141–158. <https://doi.org/10.18546/LRE.17.2.04>.
- Kuhn, K. D. (2018). Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transportation Research Part C: Emerging Technologies*, 87, 105–122. <https://doi.org/10.1016/j.trc.2017.12.018>.
- Lee, D., Watson, S. L., & Watson, W. R. (2019). Systematic literature review on self-regulated learning in massive open online courses. *Australasian Journal of Educational Technology*, 35(1), 28–41. <https://doi.org/10.14742/ajet.3749>.
- Lee, S., & Chung, J. Y. (2019). Lessons learned from two years of K-MOOC experience. *Educational Media International*, 56(2), 134–148. <https://doi.org/10.1080/09523987.2019.1614245>.
- Li, S., Tang, Q., & Zhang, Y. (2016). A case study on learning difficulties and corresponding supports for learning in cMOOCs. *Canadian Journal of Learning and Technology*, 42(2), 1–26. <https://doi.org/10.21432/T2GS4S>.
- Liyaganawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs: A systematic study of the published literature 2008–2012. *International Review of Research in Open and Distributed Learning*, 14(3), 202–227. <https://doi.org/10.19173/irrodl.v14i3.1455>.
- Lu, Y. F., Wang, B., & Lu, Y. B. (2019). Understanding key drivers of MOOC satisfaction and continuance intention to use. *Journal of Electronic Commerce Research*, 20(2), 105–117.
- Maxwell, W. D., Fabel, P. H., Diaz, V., Walkow, J. C., Kwiek, N. C., Kanchanaraksa, S., Wamsley, M., Chen, A., & Bookstaver, P. B. (2018). Massive open online courses in US healthcare education: Practical considerations and lessons learned from implementation. *Currents in Pharmacy Teaching and Learning*, 10(6), 736–743. <https://doi.org/10.1016/j.cptl.2018.03.013>.
- Moreno-Marcos, P. M., Munoz-Merino, P. J., Alario-Hoyos, C., Estevez-Ayres, I., & Delgado Kloos, C. (2018). Analysing the predictive power for anticipating assignment grades in a massive open online course. *Behaviour & Information Technology*, 37(10–12), 1021–1036. <https://doi.org/10.1080/0144929X.2018.1458904>.

- Meek, S. E., Blakemore, L., & Marks, L. (2017). Is peer review an appropriate form of assessment in a MOOC? Student participation and performance in formative peer review. *Assessment & Evaluation in Higher Education*, 42(6), 1000–1013. <https://doi.org/10.1080/02602938.2016.1221052>.
- Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., & Kloos, C. D. (2018). Prediction in MOOCs: A review and future research directions. *IEEE Transactions on Learning Technologies*, 12(3), 384–401. <https://doi.org/10.1109/TLT.2018.2856808>.
- Murray, J. A. (2014). Participants' perceptions of a MOOC. *Insights: The UKSG Journal*, 27(2), 154–159.
- Mustikasari, D. W. (2017). Developing massive open online course (MOOC): Need analysis of teaching materials for madrasah English teachers. *Register Journal*, 10(2), 170–184. <https://doi.org/10.18326/rgt.v10i2.170-184>.
- Nkuyubwatsi, B. (2016). Positioning extension massive open online courses (xMOOCs) within the open access and the life long learning agendas in a developing setting. *Journal of Learning for Development*, 3(1), 14–36.
- Onah, D. F. O., & Sinclair, J. E. (2017). Assessing self-regulation of learning dimensions in a stand-alone MOC platform. *International Journal of Engineering Pedagogy*. <https://doi.org/10.3991/ijep.v7i2.6511>.
- Online Education Blog of Touro College. (2013). What is the difference between xMOOCs and cMOOCs?. Retrieved from <http://blogs.onlineeducation.touro.edu/distinguishing-between-cmoocs-and-xmoocs/>.
- Pang, Y., Liu, W., Jin, Y., Peng, H., Xia, T., & Wu, Y. (2018). Adaptive recommendation for MOOC with collaborative filtering and time series. *Computer Applications in Engineering Education*, 26(6), 2071–2083. <https://doi.org/10.1002/cae.21995>.
- Paton, R. M., Fluck, A. E., & Scanlan, J. D. (2018). Engagement and retention in VET MOOCs and online courses: A systematic review of literature from 2013 to 2017. *Computers & Education*, 125, 191–201. <https://doi.org/10.1016/j.compedu.2018.06.013>.
- Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation*, 25(4), 348–349.
- Rabin, E., Kalman, Y. M., & Kalz, M. (2019). An empirical investigation of the antecedents of learner-centered outcome measures in MOOCs. *International Journal of Educational Technology in Higher Education*, 16(1), 1–20. <https://doi.org/10.1186/s41239-019-0144-3>.
- Raffaghelli, J. E., Cucchiara, S., & Persico, D. (2015). Methodological approaches in MOOC research: Retracing the myth of Proteus. *British Journal of Educational Technology*, 46(3), 488–509. <https://doi.org/10.1111/bjjet.12279>.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2014a). stm: R package for structural topic models. *Journal of Statistical Software*, 10(2), 1–40. <https://doi.org/10.18637/jss.v000.i00>.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., & Rand, D. G. (2014b). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082. <https://doi.org/10.1111/ajps.12103>.
- Rolfe, V. (2015). A systematic review of the socio-ethical aspects of Massive Online Open Courses. *European Journal of Open, Distance and E-Learning*, 18(1), 52–71. <https://doi.org/10.1515/eurodl-2015-0004>.
- Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601–618. <https://doi.org/10.1109/TSMCC.2010.2053532>.
- Saadatdoost, R., Sim, A. T. H., Jafarkarimi, H., & Mei Hee, J. (2015). Exploring MOOC from education and Information Systems perspectives: A short literature review. *Educational Review*, 67(4), 505–518. <https://doi.org/10.1080/00131911.2015.1058748>.
- Siddike, M. A. K., Kohda, Y., & Hoque, M. (2017). Application of MOOCs for borrowers' financial education in microfinance. *Knowledge Management & E-Learning: An International Journal*, 9(2), 160–176. <https://doi.org/10.34105/j.kmel.2017.09.010>.
- Siemens, G. (2004). Connectivism: A learning theory for the digital age. Retrieved October 29, 2020, from <https://jotamac.typepad.com/jotamacsweblog/files/Connectivism.pdf>.
- Siemens, G. (2012). MOOCs are really a platform. Retrieved October 29, 2020, from <http://www.elearnspace.org/blog/2012/07/25/moocs-are-really-a-platform/>.
- Silge, J. (2018). Training, evaluating, and interpreting topic models. Retrieved October 29, 2020, from <https://juliasilge.com/blog/evaluating-stm/>.
- Skrypnik, O., Joksimović, S., Kovanović, V., Gašević, D., & Dawson, S. (2015). Roles of course facilitators, learners, and technology in the flow of information of a cMOOC. *The International Review of Research in Open and Distributed Learning*, 16(3), 187–217. <https://doi.org/10.19173/irrodl.v16i3.2170>.
- Swamy, N. (2016). Cluster purity visualizer. <https://bl.ocks.org/nswamy14/e28ec2c438e9e8bd302f>.
- Thompson, K. (2011). 7 things you should know about MOOCs. Retrieved October 29, 2020, from <https://www.chronicle.com/article/8-things-you-should-know-about-moocs/>.
- Veletsianos, G., & Shepherdson, P. (2015). Who studies MOOCs? Interdisciplinarity in MOOC research and its changes over time. *International Review of Research in Open and Distributed Learning*, 16(3), 1–17. <https://doi.org/10.19173/irrodl.v16i3.2202>.
- Veletsianos, G., & Shepherdson, P. (2016). A systematic analysis and synthesis of the empirical MOOC literature published in 2013–2015. *International Review of Research in Open and Distributed Learning*, 17(2), 198–221. <https://doi.org/10.19173/irrodl.v17i2.2448>.
- Watson, S. L., & Kim, W. (2016). Enrolment purposes, instructional activities, and perceptions of attitudinal learning in a human trafficking MOOC. *Open Learning*, 31(3), 1–15. <https://doi.org/10.1080/02680513.2016.1230845>.
- Watson, S. L., Watson, W. R., Janakiraman, S., & Richardson, J. (2017). A team of instructors' use of social presence, teaching presence, and attitudinal dissonance strategies: An animal behaviour and welfare MOOC. *International Review of Research in Open & Distributed Learning*, 18(2), 68–91. <https://doi.org/10.19173/irrodl.v18i2.2663>.
- Wong, J. S., Pursel, B., Divinsky, A., & Jansen, B. J. (2015). Analyzing MOOC discussion forum messages to identify cognitive learning information exchanges. *Proceedings of the Association for Information Science and Technology*, 52(1), 1–10. <https://doi.org/10.1002/pra2.2015.145052010023>.
- Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G. J., & Paas, F. (2019). Supporting self-regulated learning in online learning environments and MOOCs: A systematic review. *International Journal of Human-Computer Interaction*, 35(4–5), 356–373. <https://doi.org/10.1080/10447318.2018.1543084>.
- Wu, B., & Chen, X. (2016). Continuance intention to use MOOCs: Integrating the technology acceptance model (TAM) and task technology fit (TTF) model. *Computers in Human Behavior*, 67, 221–232. Doi:<https://doi.org/10.1016/j.chb.2016.10.028>.
- Yang, J. (2020). Cultivation of college english self-regulated learning ability and adjustment of teachers' roles. *Overseas English*, 18, 175–176.
- Zheng, Y., & Yang, R. Y. (2017). The rise of MOOCs: The literature review of research progress and hot spots of MOOCs education in mainland China. *EURASIA Journal of Mathematics Science and Technology Education*, 13(9), 6165–6174. <https://doi.org/10.12973/eurasia.2017.01056a>.

- Zhu, M., Sari, A., & Lee, M. M. (2018a). A systematic review of research methods and topics of the empirical MOOC literature (2014–2016). *The Internet and Higher Education*, 37, 31–39. <https://doi.org/10.1016/j.iheduc.2018.01.002>.
- Zhu, Y., Zhang, W., He, Y., et al. (2018b). Design and implementation of curriculum knowledge ontology-driven SPOC flipped classroom teaching model. *Educational Sciences: Theory & Practice*, 18(5), 1351–1374. <https://doi.org/10.12738/estp.2018.5.034>.
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3–17. https://doi.org/10.1207/s15326985ep2501_2.
- Zimmerman, B. J., & Schunk, D. H. (2012). *Self-regulated learning and academic achievement: Theory, research, and practice*. Berlin: Springer.
- Zucconi, A. (2016). The need for person-centered education. *Cadmus*, 3(1), 1–26.

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