



## Review article

# Take a MOOC and then drop: A systematic review of MOOC engagement pattern and dropout factor

Hao Huang<sup>a,b,\*</sup>, Lihjen Jew<sup>a</sup>, Dandan Qi<sup>b</sup><sup>a</sup> International College, Krirk University, No. 3, Soi Ramindra 1 Anusawari Subdistrict, Bang Khen District, Bangkok, Thailand<sup>b</sup> Chongqing Business Vocational College, 81 Daxuechengzhong Road, Shapingba District, Chongqing, China

## ARTICLE INFO

## Keywords:

MOOC

Higher education

Engagement pattern

Dropout factor

## ABSTRACT

Massive Open Online Course (MOOC) plays an important role in education equity and lifelong learning without entrance barriers, time limitations, and geographical constraints. However, MOOC is criticized by researchers for its high dropout rate. The purpose of this paper is to summarize the engagement patterns of MOOC learners and factors affecting dropouts in higher education in the scholarly literature, published during the period of 2008 to 2021. Twenty-one available published studies in seven major academic databases had been reviewed and presented in the systematic literature review following the PRISMA methodology. The results showed that the engagement patterns of MOOC learners could be grouped to Start, Mid, and End according to the stage of MOOC learning. The factors affecting MOOC dropout could be grouped into categories of Course Attributes, Social Status, Cognitive Ability, Emotional Factor and Learning Behavior.

## 1. Introduction

Massive Open Online Course (MOOC) is a way of sharing and learning without enrollment and pre-requirement limitation [1]. MOOC was originally named by Canadian scholars Bryan Alexandra and Dave Cormier in 2008 for the online course run by Siemens and Downs [2]. Driven by the growth of information technology and communication media in recent years, MOOC has developed rapidly with world-class educational institutes involved. Since COVID-19 broke out all over the world in 2020, conventional face-to-face education stopped, and MOOC is becoming a popular teaching and learning method in most countries and regions during this specific time [3–6]. Compared with traditional on-campus education, MOOC is open to all learners with internet connection but without enrollment restrictions. In this case, MOOC plays an important role in education equity and lifelong learning because anyone can learn any educational resources without entrance barriers, time limitations, and geographical constraints [7]. Data collected in 2020 by Class Central ([www.class-central.com](http://www.class-central.com)) showed that Coursera, the largest MOOC provider, reached 31 million new registered users followed by edX with 10 million and FutureLearn with 5 million [8].

While MOOC has already been utilized as a complemented education tool, especially in higher education, researchers criticized it for its low completion rate. According to researchers, the MOOC registrants were massive but the completion rates were lower than 10% [9]. It means that MOOC was well-known by massive learners but few of them got betterment from it. Low completion rate caused many problems and hurt the benefits from people involved. MOOC learners' time was wasted, MOOC providers' effort was wasted and

\* Corresponding author. International College, Krirk University, No. 3 Soi Ramindra 1 Anusawari Subdistrict, Bang Khen District, Bangkok, Thailand.

E-mail addresses: [realhuanghao@hotmail.com](mailto:realhuanghao@hotmail.com) (H. Huang), [Jewlihen@gmail.com](mailto:Jewlihen@gmail.com) (L. Jew), [dandan526458920@163.com](mailto:dandan526458920@163.com) (D. Qi).

<https://doi.org/10.1016/j.heliyon.2023.e15220>

Received 21 February 2023; Received in revised form 25 March 2023; Accepted 30 March 2023

Available online 11 April 2023

2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

MOOC platform companies' profit was lost. Researchers wondered why people took a MOOC, but after a while, they quitted. To solve the problem of low completion rate, many researchers put much effort into looking for the engagement patterns and finding out factors affecting dropouts [10–13]. Student engagement patterns in these research were used to classify MOOC learners into different types according to their learning actions such as videos watched, quizzes answered and posts made to forums [14]. Many scholars also have systematically examined the literatures associated with MOOC teaching and learning. Zhu and her team did two systematic reviews on MOOC research techniques and trends in 2018 and 2020 respectively [15,16]. They found the research topics about MOOC primarily focused on student perceptions and course designs. Dalipi et al. [17] examined literatures on MOOC student dropout prediction techniques of machine learning. Ogunyemi et al. [18] explored factors affecting MOOC learning engagement from previous studies and they found ten indicators such as learner's personality and learning interaction were identified as important for fostering the learner engagement in MOOC. Despite these scholarly contributions, previous reviews did not focus primarily on the study of how college students learned MOOC and why they dropped out.

The purpose of this study is to offer a comprehensive systematic literature review of the research on MOOC learner engagement patterns and dropout factors to contribute for MOOC instructor and institute management. In this study, 21 studies on MOOC published between 2008 and 2021 were selected and reviewed. The research questions listed below guided this study:

- RQ1. What are the engagement patterns of MOOC learners in higher education?
- RQ2. What factors affect MOOC dropouts in higher education?

## 2. Research method

A systematic literature review (SLR) method was utilized for this study. SLR refers to a scientific approach of collecting data from multiple studies for a certain topic, evaluating by inclusion and exclusion criteria, integrating most related evidence, and presenting the findings of the study [19]. The SLR has features of (a) objectivity in selecting literature, (b) completeness in searching literature, and (c) the presentation of the findings from synthesizing. To this end, Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) was used for this systematic review. The review was based on the four stages of information flow described in the PRISMA model with the support of software EndnoteX9, Nvivo11, and Excel [20].

### 2.1. Identification

Considering the completeness in searching, the WEB OF SCIENCE, SCOPUS, IEEE, JSTOR, SAGE, Springer Link, and EBSCO were chosen as databases in this study. Since the WEB OF SCIENCE has a different search engine from other databases, different search strategies were used for initial search. For WEB OF SCIENCE, "MOOC", "higher education" and "learner OR student OR participant OR participation" were chosen as the search words for steps 1 to 3 independently, and then we combined step 1 to 3 as step 4. For other databases, "MOOC" AND "higher education" AND "learner OR student OR participant OR participation" were used as keywords in the title and abstract search. We searched in September 2021 and a total of 4814 papers were identified. 3722 papers were left after duplicates were removed through Endnote. The search results of each database in the identification stage were shown in Table 1.

### 2.2. Screening

At the screening stage, we manually analyzed these papers using the codebook, as shown in Table 2, and filtered the duplicate and non-compliant articles by browsing their titles and abstracts. Because previous studies implemented their experiments with various definitions of dropout [12], to cover as many relevant literatures as possible, all papers with different definitions of dropout were included for further analysis. We screened out 3697 papers at this stage and 25 papers were left.

### 2.3. Eligibility

The eligible studies had to meet both of the following criteria: 1) studies focused on MOOC engagements or dropout factors; and 2)

**Table 1**  
Search results in Identification Stage.

Database	Result
Web of Science	604
SCOPUS	1184
IEEE	387
JSTOR	462
SAGE	602
Springer Link	931
EBSCO	644
Total	4814
Duplicated	-1092
<b>After removing duplicates</b>	<b>3722</b>

**Table 2**  
Inclusion and exclusion criteria code book.

Main Criteria	Definitions
Inclusion	MOOC in higher education
Exclusion	not in the context of higher education
	not about MOOC engagement analysis
	not about dropout factor analysis
	not a paper article
	not in English
	literature review
	repeated study

Note. MOOC: massive open online course.

in the context of higher education. Three researchers (HH, LJ, DQ) reviewed the full-text of 77 papers independently. Disagreements about eligibility were discussed until consensus was obtained.

2.4. Included

Finally, 25 research papers were left for final selection process. After further analysis on the methods and results of these articles, 4 papers were found included neither MOOC engagement patterns, nor factors affecting the students’ MOOC performances. Therefore, 21 papers were eligible for synthesis and review. The PRISMA flowchart was shown in Fig. 1.

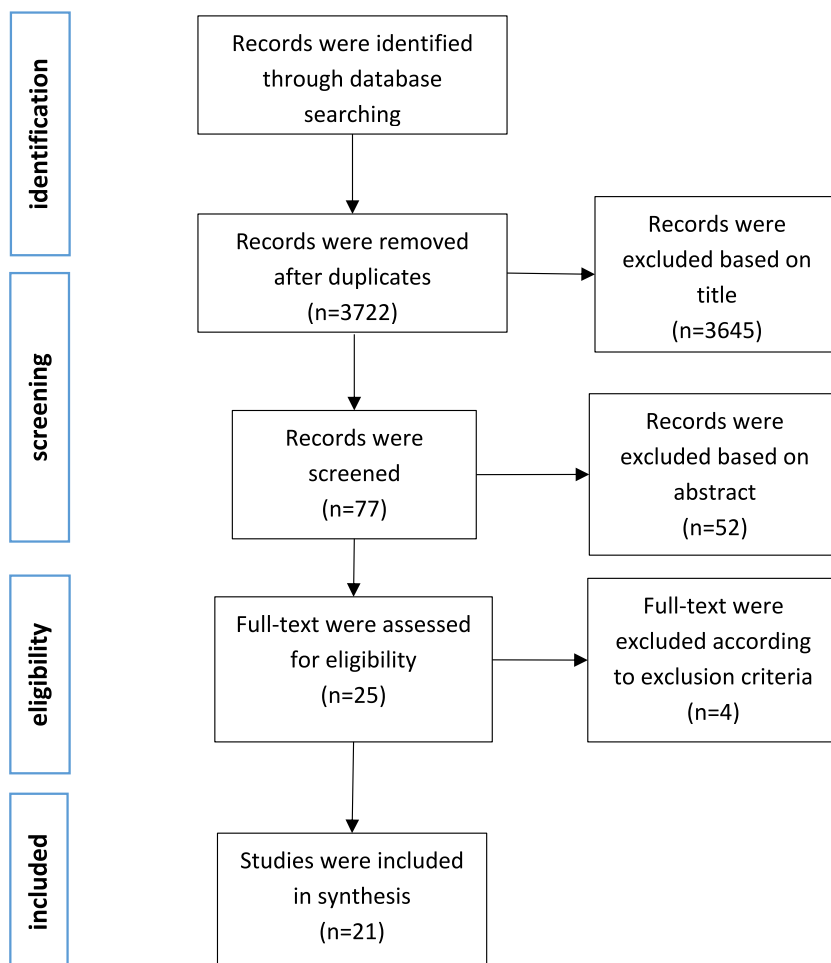


Fig. 1. Results of literature search based on the PRISMA statement.

## 2.5. Study quality assessment

To assess the quality of the 21 selected papers, three researchers (HH, LJ, DQ) reviewed the full texts independently according to the study quality assessment checklists [21], as shown in Table 3, and answered Yes/No/Partially. In case of disagreement for these questions, researchers discussed inconsistencies until consensus was reached.

The answers to the questions in Table 3 were assigned with three scores: Yes (1), No (0), and Partially (0.5). The overall quality score for a paper ranged from 0 to 5, representing poor and excellent quality respectively [21]. Each researcher assessed papers against study quality assessment checklists and results were presented in Table 4.

## 2.6. Data extraction

According to our research question, two review authors (LJ and DQ) used to extract data from selected papers. Extracted data were compared and discussed with any discrepancies. HH entered data into Excel and double checked this for accuracy. Finally, we extracted 9 categories of relevant data from the 21 research papers. Table 5 shows the categories of extraction data and their descriptions.

## 3. Results

### 3.1. Year of publication and geographical distribution

We counted the countries of the first authors and found that the United Kingdom (6 papers) had the most publications, followed by Australia (3 papers), Mexico (2 papers), and Austria (2 papers). China, Egypt, Greece, India, Malaysia, Morocco, Norway, and Turkey each had 1 paper each from 2014 to 2021 on this particular topic, as shown in Figs. 2 and 3. These figures indicate that developed countries paid more attention to the problem of MOOC dropout than developing countries. Specifically, as shown in Fig. 3, the interest in MOOC engagement pattern had increased since 2014. After 2017, the study of MOOC engagement pattern decreased and the machine learning for dropout prediction was getting more concerns from researchers.

### 3.2. MOOC engagement pattern

Understanding how learners access MOOC resources and exploring the relationship between MOOC engagement pattern and learning performance will provide important evidence for instructors and stakeholders to improve MOOC design and provide suitable support.

#### 3.2.1. Engagement pattern and classification criteria

After analyzing the selected papers, we can find that the MOOC learner engagement patterns were quite different, as shown in Table 6, based on their research purpose. To reveal a more detailed story of student engagement, Coffrin et al. [10] divided MOOC learners into groups of *Auditors*, *Active and Qualified* based on the total weekly participation across two distinct MOOC courses. Auditor students watched videos but did few assessment. Active students involved in assessments according to instructor's requirement. Qualified students watched a videos or participated assessments and met the assessment qualification criteria as well. Glance and Barrett [22] analyzed student activities in terms of the two main components of watching videos and taking quizzes, and the results showed that there were two groups of people taking MOOC: *Auditors* and *Engagers*. Rather than analyzing students based on accessing certain course materials, other researchers used not only content and assessment within MOOC courses but also discussion and interaction among different involved parties [11,12,23]. Ferguson and Clow [11] used the data from FutureLearn which applied socio-constructivist pedagogy and they classified learner as *Samplers*, *Strong Starters*, *Returns*, *Mid-way Dropouts*, *Nearly There*, *Late Completers* and *Keen Completers*. Sunar et al. [12] conducted a study based on the social behaviors of learners from FutureLearn. From the analysis of the selected papers, it is found that most studies classified MOOC engagement patterns, according to the learners' final performances, to *Completer* and *Dropout* [24–30]. Duru et al. [31] specifically focused on MOOC learners with language problems, so they divided them into *English as an Official and Primary Language (EOPL)*, *English as an Official but Not Primary Language (EONPL)* and *English as a Second Language (ESL)* groups. By analyzing the engagement patterns of these groups, the findings indicated that language used in MOOC affected the completion rates among different groups of MOOC learners.

Although the MOOC engagement patterns were quite different from the studies according to their various research purposes, they

**Table 3**  
Study quality assessment checklists.

NO.	Question
1	Is the analysis process description complete?
2	Is it clear what methods were used in study?
3	Is it clear how accuracy of the result was measured?
4	If cross-validation method was used?
5	Were all variables fully defined?

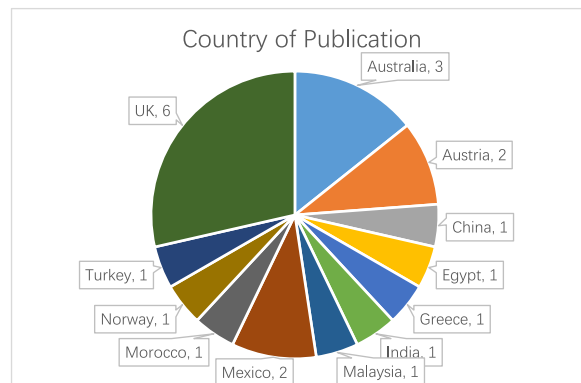
**Table 4**  
Quality criteria and paper ID

Score Range*	Quality	Paper ID
[0, 1.5)	Poor Quality	
[1.5, 3)	Fair Quality	
[3, 5)	Good Quality	3, 5, 8, 9, 10, 15, 16, 21
5	Excellent Quality	1, 2, 4, 6, 7, 11, 12, 13, 14, 17, 18, 19, 20

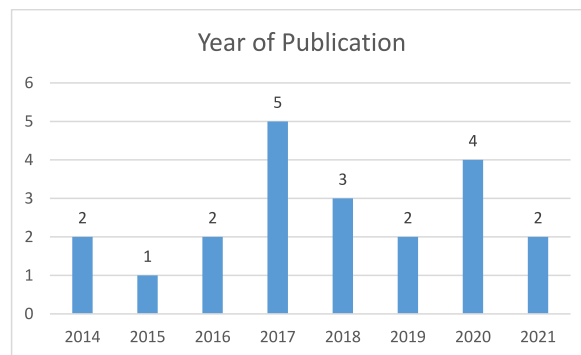
Note. \*A bracket means the number is included and a parenthesis means the number is excluded. For example, [1.5, 3) means scores from 1.5(included) to 3(excluded).

**Table 5**  
Data category and description from 21 papers.

Category	Description
Paper ID	ID number of the selected paper
Title	the title of the selected paper
Author	the authors of the selected paper
Year	paper publication year
Country	the country of the first author
Method	The method used to find out factors
Feature	the features used for engagement classification
Engagement	engagement patterns of MOOC learners
Factor	factors affecting MOOC performance
Conclusion	conclusion of the study



**Fig. 2.** Country of publication.



**Fig. 3.** Year of publication.

could be summarized as *Demographics (DE)*, *Learner Behaviors (LB)*, and *Final Performances (FP)* according to the classification criteria. DE refers to the MOOC learner’s personal information of age, gender, nationality, native language, economical status, etc. LB includes MOOC learners studying behaviors such as login time and numbers, watching the video, taking quizzes, interacting with fellows, etc.

**Table 6**  
MOOC engagement patterns and classification criteria.

NO.	Author (year)	Engagement Patterns	Classification criteria
1	Coffrin, Corrin, De Barba and Kennedy (2014)	Auditors, Active, Qualified	Learner Behavior
2	Glance and Barrett (2014)	Auditors, Engagers	Learner Behavior
3	Ferguson and Clow (2015)	Samplers, Strong Starters, Returners, Mid-way Dropouts, Nearly There, Late Completers, and Keen Completers.	Learner Behavior
4	Gomez and Aleman (2016)	Incomplete & Complete	Final Performance
5	Jane and Sara (2016)	Active and Passive	Learner Behavior
6	Shabandar, Hussain, Laws, Keight, and Lunn (2017)	In & Out learners over the course of the first week In & Out learners over the course of last week	Final Performance
7	Duru, Sunar, Dogan and White (2017)	“English as an Official and Primary Language”, “English as an Official but Not Primary Language”, and “English as a Second Language” group.	Learner Demographic
8	El-Said (2017)	Completers and Dropouts	Final Performance
9	Sunar, White, Abdullah, and Davis (2017)	Follower, Not Follower	Learner Behavior
10	Vitiello, Walk, Chang, Hernandez, Helic, and Guetl (2017)	Completers and Dropouts	Final Performance
11	Shabandar, Hussain, Liatsis, and Keight (2018)	Retention, Completion and Attrition, Learner groups	Final Performance
12	Vázquez, Ramirez and González (2018)	participants who completed both tools (completers) and those who completed the initial tool only (non-completers)	Final Performance
13	Vitiello, Walk, Helic, Chang, and Guetl (2018)	Inactive users , Active users , Completers , Dropouts	Learner Behavior
14	Imran, Dalipi and Kastrati (2019)	Complete and Dropout	Final Performance
15	Youssef, Mohammed, Hamada and Wafaa (2019)	the learners at risk of dropping out, those who are likely to fail, and those who are on the road to success.	Learner Behavior
16	Aldowah, Samarraie, Alzahrani and Alalwan (2020)	Complete and Dropout	Final Performance
17	Dai, Teo, Rappa, and Huang (2020)	Continuance Intention and Dropout	Learner Behavior
18	Lai, Zhao, and Yang (2020)	Complete and Dropout	Final Performance
19	Narayanasamy and Elçi (2020)	Dropout Class and Retention Class	Final Performance
20	Crane and Comley (2021)	Social Learners , Non-social Learners	Learner Behavior
21	Panagiotakopoulos, Kotsiantis, Kostopoulos, Iatrellis, and Kameas (2021)	Complete and Dropout	Final Performance

FP is always expressed by the grade that the learner achieved at the end of the MOOC.

### 3.2.2. Research trend in engagement patterns

By analyzing the papers in Table 6, sorted by publication year, we can see the research trend on engagement pattern was changed. The analysis of selected papers shows that studies were mainly on MOOC learner's behavior from 2014 to 2016 [10,12,14,23,25,32–34]. Various resources on MOOC have different values, and the way student accessing to these resources reflects their learning preference to some extent. Existing studies showed that the types of resources that students accessed and the way they accessed would affect students' final performance.

After 2016, most of the studies started to focus on if there was any difference in demographics and cognitive status between completers and dropouts [24–26]. Having so much research evidence and thanks to the development of information technology, researchers tried to utilize machine learning as a suitable way to identify learners at risk, and then they can provide educational intervention in advance to solve the problem of high dropout rate. In this case, the classification of learners was for prediction purposes rather than engagement pattern analysis [9,30,34].

### 3.3. Factor affecting dropout

Finding out what factors causes MOOC dropouts in higher education has always been the focus of researches. The most involved theories in the analysis of MOOC dropout factor were Cognitive Dissonance Theory [35], Self-regulation Theory and Incentive Motivation Theory [28]. These theories explained the reasons why learners continued or abandoned MOOC. It can be seen from Table 7 that 20 out of 21 papers used the quantitative analysis method and only one paper used qualitative analysis in factor analysis. Coffrin et al. [10] conducted an exploratory investigation of students' learning processes in two MOOC courses which had a different curriculum and assessment designs. A clear finding from this investigation was that the student activities and success in the first couple of weeks were significantly associated with students' outcomes at the end of the course. Glance and Barrett [22] did a novel analysis using activity data from log records collected from 42 MOOCs run by Stanford University between 2011 and 2013 on Coursera and the research results demonstrated that attrition was not related to the level of engagement and was a deterministic process.

We classified factors affecting MOOC performance extracted from each paper into categories of (1) learner behavior, (2) learner

**Table 7**  
Factors and methods.

NO	Author (year)	Factors	Category	Method	Classification
1	Coffrin, Corrin, De Barba and Kennedy (2014)	student activity and success in the first couple of weeks	learner behaviors	clustering and State Transition Diagrams	quantitative
2	Glance and Barrett (2014)	duration of the course, number of activities, external factors such as lack of time	Course attributes	video and quiz activity is plotted against the normalized sequence of the video and quiz	quantitative
3	Ferguson and Clow (2015)	social-constructivist pedagogy course context, course design, and course pedagogy	Course attributes	Replicating, a k-means algorithm to extract clusters	quantitative
4	Gomez and Aleman (2016)	educational level, previous experience, self-regulation, IT level, language, social status	learner demographics	quantitative methodology, descriptive statistics, and probabilistic models	quantitative
5	Jane and Sara (2016)	the level of meaningful, high impact, engagement activity	learner behaviors	North American National Survey of Student Engagement and UKES survey	quantitative
6	Shabandar, Hussain, Laws, Keight, and Lunn (2017)	clickstream attribute	learner behaviors	Exploratory Data Analysis (EDA), Conventional supervised machine learning, Voting classification algorithms	quantitative
7	Duru, Sunar, Dogan and White (2017)	language	learner demographics	machine learning techniques and natural language processing	quantitative
8	El-Said (2017)	Interactivity, technology employment, Course design, Perceived Value, social behaviors	comprehensive	Thematic analyses of semistructured interviews	qualitative
9	Sunar, White, Abdullah, and Davis (2017)		learner behaviors	The algorithm in Python, GLE, and PowerPoint tools	quantitative
10	Vitiello, Walk, Chang, Hernandez, Helic, and Guetl (2017)	Session Length, Timespan Clicks, Requests, and Active Time	comprehensive	Support Vector Machine and Boosted Decision Tree	quantitative
11	Shabandar, Hussain, Liatsis, and Keight (2018)	engagement level, demographic and behavioral features	comprehensive	statistical and machine learning	quantitative
12	Vázquez, Ramirez and González (2018)	motivation, previous general knowledge, specific knowledge, technology practices involved, behavioral and cognitive elements	comprehensive	The survey, <i>t</i> -test analyses, exploratory factor analysis	quantitative
13	Vitiello, Walk, Helic, Chang, and Guetl (2018)	learning behaviors	learner behaviors	Boosted Decision Tree	quantitative
14	Imran, Dalipi and Kastrati (2019)	user provided data, system generated events	learner behaviors	evaluated 15 different combinations of architectures for DNN hidden-layers	quantitative
15	Youssef, Mohammed, Hamada and Wafaa (2019)	11 categories of 61 features from the dataset	comprehensive	machine learning algorithms	quantitative
16	Aldowah, Samarraie, Alzahrani and Alalwan (2020)	academic skills and abilities, prior experience, course design, feedback, social presence, and social support.	comprehensive	a multiple-criteria decision-making method , DEMATEL method	quantitative
17	Dai, Teo, Rappa, and Huang (2020)	Confirmation, Attitude, Satisfaction, Continuance Intention, Curiosity	comprehensive	Equation modeling	quantitative
18	Lai, Zhao, and Yang (2020)	students' behavior logs	learner behaviors	Broad Learning System (Random Vector Function-Link Neural Network)	quantitative
19	Narayanasamy and Elçi (2020)	course structure, IT level, course materials, family reason, and language	comprehensive	machine learning methods	quantitative
20	Crane and Comley (2021)	engagement with the teacher and fellow students	learner behaviors	descriptive analysis	quantitative
21	Panagiotakopoulos, Kotsiantis, Kostopoulos, Iatrellis, and Kameas (2021)	personal and demographic information of students	learner demographics	9 classification algorithms compared	quantitative

demographic, (3) course attribute and (4) comprehensive. Fig. 4 indicates that the most common and widely accepted factor categories are comprehensive (8 papers) and learner behaviors (8 papers).

Besides finding out the factors affecting MOOC performance, many scholars began to pay more attention to the prediction of students with dropout risks. To develop a powerful and accurate predictive model, researchers usually used the information captured from learners' clicks and their interaction with course resources. Clickstream data reflected the various behaviors of learners on the MOOC platform, such as video clicks, duration, conversion, etc. Clickstream was found as the most commonly used data source for prediction since they could be extracted from the system with a large volume and could also be used for further research by machine learning. Al-Shabandar et al. [25] applied Learning Analytics to address dropout issues by comparing the attrition rates in different periods of learning. The results showed that the participants were more likely to dropout during the first week.

Previous studies had shown that course completers involved more in course contents and social discussion forums. Some MOOC platforms were constructed by social constructivist approach to help participants to easily express their thoughts and interact with

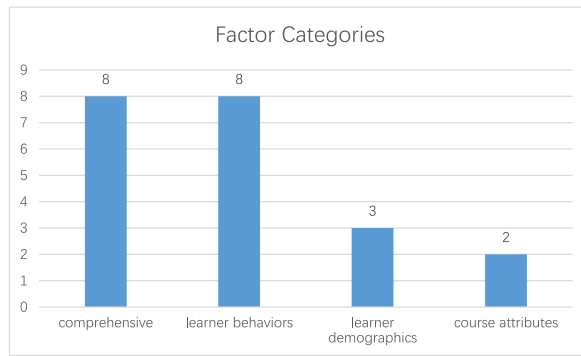


Fig. 4. Factor categories.

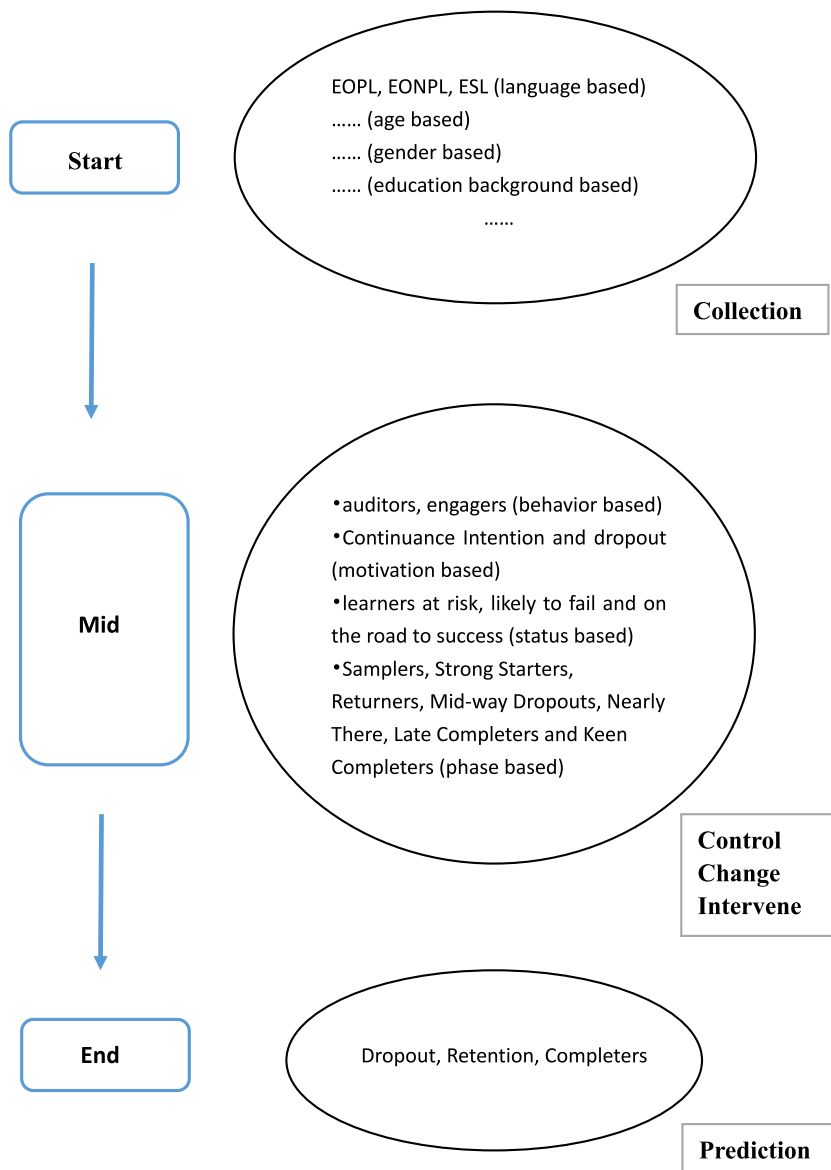


Fig. 5. Categories of MOOC engagement patterns.



peers [12]. In this case, researchers added variables such as the forum data, frequency of completing the subject test, accuracy rate and other learning behavior data, demographic information data, personal attitude and emotion data (collected through survey questionnaires or interviews) into the prediction model [36]. With the development of information technology, educational data could be collected and stored, then processed by a machine easily. According to the statistical results in Table 7, more than half of the articles used various machine learning methods to make predictions. Quantitative analysis was the main analysis strategy to do the research in this area. Vitiello et al. [32] experimented with a Boosted Decision Tree and considered data from users' active time and interactions with the MOOC. To evaluate the performance of machine learning algorithms, Imran et al. [33] investigated the performance of various feedforward Deep Neural Network (DNN), which was a machine learning algorithms implemented by stacking layers of neural network, architectures to develop a student dropout prediction model. The results showed that 7 hidden layers with 64 neurons DNN architecture had the highest accuracy.

#### 4. Discussion and direction for future studies

To the best of our knowledge, this is the first systemic literature review examining and synthesizing the MOOC engagement patterns and factors affecting dropouts in higher education. This study contributes to addressing a gap in literature as MOOC learner engagement patterns and dropout factors have yet to be comprehensively reviewed before. In this study, analysis from 21 selected papers, we found that although the MOOC engagement patterns were quite different among the studies, they can be classified into three basic categories: learner demographic, learner behavior, and final performance. The factors affecting MOOC performances that were extracted from the selected paper include categories of learner behavior, learner demographic, Course attribute, and comprehensive.

##### 4.1. Engagement patterns and factors affect dropouts

###### RQ1. What are the engagement patterns of MOOC learners in higher education?

From the analysis of the selected papers and data extracted, we found that the engagement patterns of MOOC learners in higher education were various according to their research purposes and methods used. In other words, because of the differences in MOOC learners, course designs, material providers, instructors, etc., all of these influenced the way people interact with MOOC. Therefore, the engagement patterns of MOOC learners were various from one to another. However, by synthesizing the engagement patterns from each paper, emerging the same patterns, extracting the similarity elements and considering the nature of MOOC learning, we can group the MOOC engagement patterns to three categories based on the learning phases: Start, Mid and End, as shown in Fig. 5. The engagement patterns in the Start group are the background information of MOOC learners which could be collected at the beginning of the course. The Mid group includes the learning behavior data which can be analyzed to control the learning environment, change the teaching method and provide necessary intervention. The End group of complete and dropout is usually analyzed to identify learner at risk.

The MOOC is open to learners with different educational backgrounds and languages, so the demographic information of learners is always one of the most important elements that researchers have to be concerned about. Duru et al. [31] specifically focused on identifying second language English speakers and grouped learners to EOPL, EONPL, and ESL. According to the research, 80% of learners who spoke English as their first language wrote comments that were longer than 200 characters. The MOOC engagement patterns classified by learner demographics were also implemented in many previous studies [37–39]. To identify demographic information of learners, the embedded questionnaire when registering on MOOC and machine learning techniques will be useful [31].

When learners accessing course materials, taking assignments, interacting with instructors or fellows, huge number of learning data are generated. With the fortune of so much information in this period, researchers can investigated MOOC learner behaviors, motivation, and learning status for detailed analysis. Based on the learner behaviors, Gance and Barrett [22] did a novel analysis and found a significant difference in watching videos and taking quizzes. From the perspective of motivation, participants with different continuance intentions would benefit differently from MOOC [35]. Applying the clustering method, Ferguson and Clow [11] found seven distinct patterns of engagement: Samplers, Strong Starters, Returners, Mid-way Dropouts, Nearly There, Late Completers and Keen Completers. The engagement patterns reflected the differences in the numbers and time of visiting content, posting comments and doing assessments. Normally, MOOC will last for a few weeks, and during this period, by analyzing learner's engagement patterns, stakeholders can do much work to improve dropout rate by controlling course materials and structure, providing intervention to learners at risk and changing their study behaviors and motivations.

**Table 8**  
Factors affecting MOOC performance.

Category	Factor
Course Attribute	contracture, design, content, duration of the course, number of activities
Social Status	The family reason, social support
Cognitive Ability	demographics, educational level, previous experience, IT level, language,
Emotional Factor	attitude, satisfaction, continuance intention, curiosity, motivation, perceived value, self-regulation,
Learning Behavior	engagement with the teacher and fellow students, students' behavior logs, system generated events, social behaviors, click stream attribute, student activity, and success in the first couple of weeks

Most studies classified MOOC learner engagement patterns, based on their final performances, to completers and dropouts. Gomez-Zermeno and Aleman De La Garza [24] conducted research and grouped learners to complete and dropout. Generally, the research of engagement patterns based on the final performance was conducted to look for the factors affecting them, so instructors could identify learners at risk and provide necessary support in advance.

#### **RQ2.** What factors affect MOOC dropouts in higher education?

We group the factors affecting MOOC completion or dropout into five categories, as shown in Table 8, based on the research results. Categories of Social Status, Cognitive Ability, Emotional Factor and Learning Behavior are related to the concept of MOOC learner, while the other category is related to the course attribute.

As an innovational way of learning, previous researches found the attributes of MOOC, such as pedagogy and duration, would affect the learners' final performances. It is found that attrition was affected by the number of activities. If MOOC courses were instructor-paced, the participant would give up when it was difficult to catch up [22]. The MOOC having the right construct pedagogy will improve final performances. For example, FutureLearn incorporated not only contents and assessments but also discussion boards within MOOC design because learners with social activities were more likely to complete MOOC [11,23].

According to the literatures, the categories which relate to the students' Social Status, Cognitive Ability, Emotional Factor and Learning Behavior could affect the students' MOOC performances. Social Status, Cognitive Ability and Emotional Factor playing an important role in MOOC performance could be explained that MOOC learning is a self-regulated activity which rely heavily on learner's personal attributes. Learning Behavior was also found to have a significant impact on final performance. This indicates that the more students engage in learning activities (accessing materials, taking quizzes and interacting with peers), the better performances they will achieve. Gomez-Zermeno and Aleman De La Garza [24] found the similar results that students who completed the course had higher education level and better personal attributes, while the participants who failed with problems social status and cognitive ability.

Additional finding from this study is that surveys and machine learning were the most frequent data collection techniques for MOOC analysis and prediction. The analysis showed that Decision Trees, Naive Bayes classifiers and Artificial Neural Networks were the most commonly used methods to predict students' performances. Imran et al. [33] have focused on the application of feedforward deep neural network architectures and the result showed that the number of videos viewed, the content accessed, the number of an assignment taken, the day's login and the posts on the forum were related to the final performance. This finding comes in line with the findings of previous studies, where it was found that machine learning is the most frequent data mining and predicting technique [40–42].

#### **4.2. Implications**

For the higher education institutions, the SLR results could provide the education stakeholders with an overview of how learners were participating with MOOC and what factors affected the final performances. The results can assist stakeholders to focus on the specific problems related to the students at risk and then provide intervention in advance to improve the students' educational outcomes. Considering the three categories of engagement patterns at the Start, Mid and End stages, with the factors affecting performances in mind, stakeholders should notice that the MOOC has its target participants with specific attributes (e.g. language level, education background). If any MOOC registrant is not in this target, there would be a risk of dropout. In this case, it is suggested that MOOC on the platform should state all the information about the course and pre-requirement clearly [43]. In addition, since the data of learning behaviors are prediction indicators for the final performances, the platform server should store as much educational data as possible for detailed analysis. Last but not least, machine learning is a powerful tool for analysis and prediction, each MOOC is suggested to have a suitable prediction model to find out learners at risk [44].

Practically, these results could assist MOOC instructors to concentrate on course design and student learning activities to increase the students' performances and education quality. Because of the nature of MOOC, all the information that instructors can access is data from the platform. Therefore, only instructor with a high level of IT ability could teach effectively in a MOOC environment. The suggestion for the instructors is to analyze from the beginning of registration to the end of the course. In the beginning, instructors should analyze the learners' background information [45]. During the course, MOOC instructors should analyze the learning activities and help students at risk to enhance their intrinsic motivations and foster learners' social skills [46]. At the end of MOOC, the instructor should analyze the performance of the learner, figure out the problem for the next run.

Researchers can benefit from the findings of this SLR for their future researchers. By analyzing learner engagement patterns in MOOC, there has been few study on course design, teaching methods from the perspectives of instructors and social environment. Therefore, in the future studies, researchers could particularly focus on these aspects.

#### **4.3. Limitation**

First, we only identified papers written in English in seven well-known electronic databases. Thus, other types of researches with other languages were omitted. Second, we did not contact any author to ask for additional data that may lead to publication bias. Finally, this research study was unfunded, and both time and resources were limited.

#### 4.4. Direction for future studies

The most common approach in previous studies was quantitative analysis. Therefore, further research is suggested to do the qualitative analysis looking for the reason behind learning behaviors. The SLR results showed the engagement patterns and factors affecting dropout, but little was known about how to increase the engagements and solve the problem of dropout. So the future work could be on the teaching personalization and educational intervention; exploring use of MOOC forums; adopting badges and other incentive mechanisms.

#### 5. Conclusion

This study explored the engagement patterns of MOOC learners and factors affecting dropout in higher education using a systemic literature review followed by PRISMA methodology. Based on the results of the review, the engagement patterns could be grouped to categories of Start, Mid and End according to the stage of MOOC learning. At the Start stage, the engagement patterns were classified based on the demographic information, such as EOPL, EONPL and ESL. At the Mid stage, the engagement patterns were classified as learning behavior based, motivation based, learning status based and learning phase based. At the End stage, the engagement patterns were usually classified to complete, retention and dropout according to the learners' final performances.

The factors affecting MOOC dropout we extracted from the selected papers could be grouped into categories of Course Attributes, Social Status, Cognitive Ability, Emotional Factor and Learning Behavior. Concerning the data collection techniques, the results indicated that surveys and machine learning were the most frequent data collection techniques and Decision Trees, Naive Bayes classifiers and Artificial Neural Networks were the most commonly used method to predict students' performances.

#### Declaration of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### References

- [1] I.S. Zinovieva, et al., The Use Of Moocs As Additional Tools For Teaching Nosql In Blended And Distance Learning Mode, 2021 vol. 1946, no. - 1.
- [2] I. De Waard, et al., Using mLearning and MOOCs to understand chaos, emergence, and complexity in education, *Int. Rev. Res. Open Dist. Learn.* 12 (7) (2011).
- [3] Y. Yilmaz, et al., RE-AIMing COVID-19 online learning for medical students: a massive open online course evaluation, *BMC Med. Educ.* 21 (2021) 1.
- [4] Y. Shi, Research of the Development of Distance Learning under the COVID-19 Circumstances Based on Video Conferencing Software and MOOCs, 2nd International Conference on Education, 2021 presented at the.
- [5] M. Pérez-Sanagustín, et al., A MOOC-based flipped experience: scaffolding SRL strategies improves learners' time management and engagement, *Comput. Appl. Eng. Educ.* 29 (4) (2020) 750–768.
- [6] T. Agasisti, G. Azzone, M. Soncin, Assessing the effect of Massive Open Online Courses as remedial courses in higher education, *Innovat. Educ. Teach. Int.* 59 (4) (2021) 462–471.
- [7] S.N. Mohd Hamid, T.T. Lee, H. Taha, N.A. Rahim, A.M. Sharif, E-Content module for chemistry massive open online course (MOOC): development and students' perceptions, *J. Technol. Sci. Educ.* 11 (1) (2021).
- [8] D. Shah, The Second Year Of The MOOC: A Review Of MOOC Stats And Trends In 2020, 2021. Available: <https://www.classcentral.com/report/the-second-year-of-the-mooc/>.
- [9] S.K. Narayanasamy, A. Elçi, An effective prediction model for online course dropout rate, *Int. J. Dist. Educ. Technol.* 18 (4) (2020) 110.
- [10] C. Coffrin, L. Corrin, P. De Barba, G. Kennedy, Visualizing Patterns of Student Engagement and Performance in MOOCs, Association for Computing Machinery, 2014 presented at the.
- [11] R. Ferguson, D. Clow, Examining engagement: analysing learner subpopulations in massive open online courses (MOOCs), in: presented at the International Learning Analytics and Knowledge Conference, 2015.
- [12] A.S. Sunar, S. White, N.A. Abdullah, H.C. Davis, How learners' interactions sustain engagement: a MOOC case study, *J. Latex Class Files* 10 (- 4) (2017) 487.
- [13] H. Aldowah, H. Al-Samarraie, A.I. Alzahrani, N. Alalwan, Factors affecting student dropout in MOOCs: a cause and effect decision-making model, *J. Comput. High Educ.* 32 (- 2) (2020) 454.
- [14] J. Sinclair, S. Kalvala, Student engagement in massive open online, *Int. J. Learn. Technol.* 11 (3) (2016) 218–237.
- [15] M. Zhu, A.R. Sari, M.M. Lee, A comprehensive systematic review of MOOC research: research techniques, topics, and trends from 2009 to 2019, *Educ. Technol. Res. Dev.* 68 (4) (2020) 1685–1710.
- [16] M. Zhu, A. Sari, M.M. Lee, A systematic review of research methods and topics of the empirical MOOC literature (2014–2016), *Internet High Educ.* 37 (2018) 31–39.
- [17] F. Dalipi, A.S. Imran, Z. Kastrati, MOOC Dropout Prediction Using Machine Learning Techniques: Review and Research Challenges, IEEE Global Engineering Education Conference, 2018 presented at the 2018.
- [18] A.A. Ogunyemi, J.S. Quaicoe, M. Bateurs, Indicators for enhancing learners' engagement in massive open online courses: a systematic review, *Comp. Educ. Open* 3 (2022).
- [19] Y. Xiao, M. Watson, Guidance on conducting a systematic literature review, *J. Plann. Educ. Res.* 39 (1) (2017) 93–112.
- [20] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, S. Linkman, Systematic literature reviews in software engineering – a systematic literature review, *Inf. Software Technol.* 51 (1) (2009) 7–15.
- [21] B. Kitchenham, E. Mendes, G.H. Travassos, A systematic review of cross- vs. Within-company cost estimation studies, in: presented at the 10th International Conference on Evaluation and Assessment in Software Engineering, 2007.
- [22] D.G. Gance, P.H.R. Barrett, Attrition patterns amongst participant groups in massive open online courses, *Crit. Persp. Educ. Techn.* (2014) 20.
- [23] R.A. Crane, S. Comley, Influence of social learning on the completion rate of massive online open courses, *Educ. Inf. Technol.* 26 (- 2) (2021) 2293.
- [24] M.G. Gomez-Zermeno, L. Aleman De La Garza, Research analysis on mooc course dropout and retention rates, *Turk. Online J. Dist. Educ.* 17 (2) (2016) 14.
- [25] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn, Towards the differentiation of initial and final retention in massive open online courses, *LNCS, Lect. Notes Comput. Sci.* 10361 (2017) 36.
- [26] G.R. El-Said, Understanding how learners use massive open online courses and why they drop out, *J. Educ. Comput. Res.* 55 (- 5) (2017) 752.
- [27] M. Vitiello, S. Walk, V. Chang, R. Hernandez, D. Helic, C. Guetl, MOOC dropouts: a multi-system classifier, in: presented at the European Conference on Technology Enhanced Learning, 2017.

- [28] R. Al-Shabandar, A.J. Hussain, P. Liatsis, R. Keight, Analyzing learners behavior in MOOCs: an examination of performance and motivation using a data-driven approach, *IEEE Access* 6 (2018), 73685.
- [29] J.A.V. Vázquez, M.S. Ramirez-Montoya, J.R.V. González, Motivation and knowledge: pre-assessment and post-assessment of MOOC participants from an energy and sustainability project, *Int. Rev. Res. Open Dist. Learn.* 19 (- 4) (2018) 132.
- [30] T. Panagiotakopoulos, S. Kotsiantis, G. Kostopoulos, O. Iatrellis, A. Kameas, Early dropout prediction in moocs through supervised learning and hyperparameter optimization, *Electronics* 10 (2021) 14.
- [31] I. Duru, A.S. Sunar, G. Dogan, S. White, Challenges of identifying second language English speakers in MOOCs, in: presented at the European Conference on Massive Open Online Courses, 2017.
- [32] M. Vitiello, S. Walk, D. Helic, V. Chang, C. Guetl, User behavioral patterns and early dropouts detection: improved users profiling through analysis of successive offering of MOOC, *J. Univers. Comput. Sci.* 24 (- 8) (2018) 1150.
- [33] A.S. Imran, F. Dalipi, Z. Kastrati, Predicting Student Dropout in a MOOC: an Evaluation of a Deep Neural Network Model, *Association for Computing Machinery*, 2019 presented at the.
- [34] S. Lai, Y. Zhao, Y. Yang, Broad Learning System for Predicting Student Dropout in Massive Open Online Courses, *Association for Computing Machinery*, 2020 presented at the.
- [35] H.M. Dai, T. Teo, N.A. Rappa, F. Huang, Explaining Chinese university students' continuance learning intention in the MOOC setting: a modified expectation confirmation model perspective, *Comput. Educ.* 150 (2020).
- [36] M. Youssef, S. Mohammed, E.K. Hamada, B.F. Wafaa, A predictive approach based on efficient feature selection and learning algorithms' competition: case of learners' dropout in MOOCs, *Educ. Inf. Technol.* 24 (- 6) (2019) 3618.
- [37] Q. Zhang, F.C. Bonafini, B.B. Lockee, K.W. Jablokow, X. Hu, Exploring demographics and students' motivation as predictors of completion of a massive open online course, *Int. Rev. Res. Open Dist. Learn.* 20 (- 2) (2019) 161.
- [38] B. Yuru, L. Shi, A.M. Toda, Demographical Changes of Student Subgroups in MOOCs: towards Predicting At-Risk Students, *Information Systems Development*, 2019 presented at the.
- [39] K. Li, MOOC learners' demographics, self-regulated learning strategy, perceived learning and satisfaction: a structural equation modeling approach, *Comput. Educ.* 132 (2019) 30.
- [40] W. Zou, Z. Pan, C. Li, M. Liu, Does social presence play a role in learners' positions in MOOC learner network? A machine learning approach to analyze social presence in discussion forums, in: presented at the International Conference on Quantitative Ethnography, 2021.
- [41] S.S. Hanswal, A. Pareek, G. Vyas, A. Sharma, Sentiment analysis on E-learning using machine learning classifiers in Python, in: presented at the Rising Threats in Expert Applications and Solutions, 2021.
- [42] S. Geng, B. Niu, Y. Feng, M. Huang, Understanding the focal points and sentiment of learners in MOOC reviews: a machine learning and SC-LIWC-based approach, *Br. J. Educ. Technol.* 51 (5) (2020) 1803.
- [43] R.H. Rizzardini, M.M. Chan, C. Guetl, An attrition model for MOOCs: evaluating the learning strategies of gamification, in: *Formative Assessment, Learning Data Analytics and Gamification*, 2016, p. 311.
- [44] Z. Xie, Modelling the dropout patterns of MOOC learners, *Tsinghua Sci. Technol.* 25 (3) (2020) 324.
- [45] S. Assami, N. Daoudi, R. Ajhoun, Personalization Criteria for Enhancing Learner Engagement in MOOC Platforms, *IEEE Global Engineering Education Conference*, 2018 presented at the.
- [46] K.F. Hew, W.S. Cheung, Students' and instructors' use of massive open online courses (MOOCs): motivations and challenges, *Educ. Res. Rev.* 12 (2014) 58.