

Thematic and Sentiment Analysis of Learners' Feedback in MOOCs

Harsh Vardhan Pant¹, Manoj Chandra Lohani¹ and Jeetendra Pande²

¹*Graphic Era Hill University*

²*Uttarakhand Open University*

Abstract: In recent years, sentiment analysis has gained popularity among researchers in various domains, including the education domain. Sentiment analysis can be applied to review the course comments in Massive Open Online Courses (MOOCs), which could enable course designers' to easily evaluate their courses. The objective of this study is to explore the influential factors that affect the completion rate of MOOCs and unravel the sentiments of dropout learners by evaluating learners' feedback. In the present study, sentiment analysis was performed using Python programming and NVivo tools on the feedback of the learners enrolled in three MOOCs entitled Introduction to Cyber Security, Digital Forensics and Development of Online Courses for SWAYAM, which was hosted on the SWAYAM platform (www.swayam.gov.in). Two instruments were used for data collection: (1) a structured questionnaire using a 5-point Likert scale was administrated using Google Forms – the questionnaires have also some additional open-ended questions – and (2) semi-structured interview schedules with the domain experts. The feedback was collected using Google Forms and a total of 324 responses were received between April 23, 2022 to May 31, 2022. The non-probability sampling method served as the sampling approach in the quantitative phase in this study.

During analysis, the findings of the feedback uncovered important dimensions of some peculiar factors that may be responsible for retention of learners, i.e., content localisation, credit mobility and latest trend courses that were less explored in the earlier literature.

Keywords: thematic, sentiment, analysis, feedback, MOOC, retention.

Introduction

MOOCs were considered as panacea for higher education during COVID-19 by providing flexible education for learners. Many governments in different countries are facing financial pressure in the field of higher education and its infrastructure, in addition to scholarships and student loans. So, several countries, including India, have advocated for Massive Open Online Courses, not only to educate students who work but also who live remotely or cannot access traditional university campuses for other reasons. However, the literature shows that online distance education has dropout rates higher than traditional education (Xavier & Meneses, 2020). Understanding factors that influence learners' education is a fundamental process to promote student retention. To do this learners' feedback represents a vital source of information that can also be used by e-instructors in order to improve teaching pedagogies and training activities. The popularity and significance of learners' feedback and reviews have increased, particularly in the COVID-19 pandemic times, when most educational institutions have changed their teaching approach from traditional face-to-face learning to the online learning method.



As we know, the demographic nature of MOOCs is very complex, with learners coming from various countries, having different languages, castes and creeds. A large volume of information in the form of reviews and comments are produced through the MOOC forums by them, regarding the course and other aspects of MOOCs. So, for the MOOCs analyst it is a complex task to analyse the sentiment of learners' comments, so that he/she can take the necessary action to help dropout learners and increase the retention rate in the MOOC platform. One way to overcome these challenges is by leveraging the advantages of sentiment analysis and opinion mining techniques (Dalipi et al, 2021).

Sentiment analysis systems use natural language processing (NLP) and machine learning (ML) techniques to discover and retrieve information and opinions from vast amounts of textual information (Cambria et al., 2013). The main goal of the current study is to evaluate the feedback and systematic reviews of learners enrolled in MOOCs entitled Introduction to Cyber Security, Digital Forensics and Development of Online Courses for SWAYAM, offered through the SWAYAM platform (www.swayam.gov.in) and to find out the actual factors that influence the learners' retention in MOOCs using sentiment analysis. The result of the study could assist MOOC designers to identify the crucial factors responsible for retention and, therefore, reduce the learners' dropout rate by addressing those factors.

Related Research

Sentiment Analysis is not a new term. The origin of sentiment analysis can be traced to the 1950s, when sentiment analysis was primarily used for written paper documents (Scott & Buzzlogix, 2015). We have seen a massive increase in the number of papers focusing on sentiment analysis and opinion mining in recent years (Mika et al., 2018). In Table 1 the authors have listed some related work which has been published in top-ranking journals.

Table 1: Recent Published Work and Literature in Sentiment and Thematic Analysis on MOOCs

Authors	Title	Objective/Findings
(Aytuğ, 2019)	Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach	The purpose of our research is to present an efficient sentiment classification scheme with high predictive performance in MOOC reviews, by pursuing the paradigms of ensemble learning and deep learning.
(Dalipi et al., 2021)	Sentiment analysis of students' feedback in MOOCs: A systematic literature review	This article is a systematic literature review on the use of sentiment analysis for evaluating students' feedback in MOOCs, exploring works published between January 1, 2015, and March 4, 2021. This study found that the research has revolved around six areas: MOOC content evaluation, feedback contradiction detection, SA effectiveness, SA through social network posts, understanding course performance and dropouts, and MOOC design model evaluation.

Authors	Title	Objective/Findings
(Greenland & Moore, 2021)	Large qualitative sample and thematic analysis to redefine student dropout and retention strategy in open online education	This study finds a probability-based qualitative sample of 226 participants captured dropout subthemes down to a 5% level of incidence or frequency of occurrence.
(Sraidi, et al., 2022)	A sentiment analysis based approach to fight MOOCs' drop out	This article explores new tools and methods (quality and machine learning) to analyse the different limitations and difficulties learners face in MOOCs.
(Xavier & Meneses, 2020)	Dropout in online higher education: A scoping review from 2014 to 2018	This article presents a scoping review of the recent literature on the theme, focusing on dropout definitions, concepts, and models, study domains and themes, methodological approaches, and findings.

Table 1 shows the last three years of published research papers regarding the related work. Although, Aytuğ (2019) analysed a corpus containing 66,000 MOOC reviews, the study is silent with regard to specified research questions. Studies like Dalipi et al. (2021) and Xavier and Meneses (2020) have found the various themes but these articles are based on literature reviews. Although there have been numerous studies conducted on the topic, such as those by Sraidi et al. (2022) and Greenland & Moore (2021), sentiment analysis has been able to identify the primary reasons for high dropout rates but identifying crucial factors that contribute to student retention remains a challenging area of research. So, it seems a further detailed investigation is needed.

Christensen et al. (2013) reported that the motivations for learners from diverse national cultures are different. In a country like India, where many languages and dialects are spoken, as well as in the SWAYAM portal, where the learner enrolment graph comes first in all reputable MOOC platforms, doing research in such an environment could strengthen the credibility of all the above research. Content localisation, latest trend course and credit mobility may also be another factor for strengthening and retaining the learners in MOOCs (Pant et al., 2021). These motivational factors are identified and published, after the findings of the various literature reviews (Pant et al., 2021). Given the earlier findings from the literature above, the researchers wanted to verify whether or not addressing these factors was effective in practice. To do this, qualitative and sentiment methods seemed to be suitable methods to understand the sentiments of learners regarding the given factors. So, this study mainly focused on exploring the influential factors that affect learners' intention to retention in MOOCs using sentiment analysis with following research questions:

RQ1: What are the influential factors that affect learners' intention to retention in MOOCs?

RQ2: Does retention in MOOCs depend on content localisation, trendiest course and credit mobility factors?

Methods

A mixed-methods approach with both qualitative and quantitative methods was adopted for the study.

Sample

The population of the current study was the learners enrolled in the MOOCs Introduction to Cyber Security (10,440 enrolments), Digital Forensics (3,883) and Development of Online Course for SWAYAM (879 enrolments), which were offered through the SWAYAM platform between January and June 2022, with a total of 15,202 enrolments. A self-selected, non-probability sampling method was used as a sample methodology in the quantitative phase of this study due to the vastness of the target population and the difficulty in collecting the sampling frame. Feedback from 324 students was collected from the three courses of the SWAYAM. The objective of the questionnaire was to collect the learners' opinions about the factors that motivated them to complete these MOOCs. A structured questionnaire using a 5-point Likert scale was administered using Google Forms. The questionnaires additionally had some open-ended questions. The language of the questions was in English and Hindi and they were all translated by a language expert.

Instruments

Two instruments were used for data collection: a survey questionnaire and interview schedules. An online survey was administered to learners from different SWAYAM courses viz. Introduction to Cyber Securityⁱ, Digital Forensicsⁱⁱ, and Development of Online Course for SWAYAMⁱⁱⁱ through a Google Form shared with the students between March and June 2022. Quantitative data was collected through the survey with respect to the following dimensions: role of instructor for learners (RIL), perceived usefulness (PU), behavioural intention (BI), content localisation (CL), credit mobility (CM) to promote e-education, perceived job performance, social influence and latest-trend course impact on e learning. The demographic data of age, gender and education was also recorded. Qualitative data was collected through open-ended questions and interview schedules prepared for the experts who were actively engaged in the development of MOOCs for SWAYAM. Cronbach's Alpha Coefficient was calculated for questionnaire reliability through the SPSS tool. The statistical value of Cronbach's α (given in Table 2) was between 0.751 to 0.873.

Table 2: Cronbach's Alpha Coefficient of Reliable Variables used in the Study

Predictor	Cronbach Alpha Coefficient
CL	0.873
CM	0.851
SI	0.809
LTC	0.787
RET	0.831
PU	0.825
BI	0.751

Beside the survey questionnaire, a focus group discussion was conducted for qualitative research. A semi-structured interview, with five experts who had designed and developed in various MOOCs

hosted on the SWAYAM platform, was conducted to investigate the factors affecting the retention rate in MOOCs.

Data Refining

Data refining is an essential process when analysing a large volume of text. This extensive procedure consists mainly of converting unstructured text data into structured forms (Ban et al., 2019). The whole data refining process can be divided into five steps as follows:

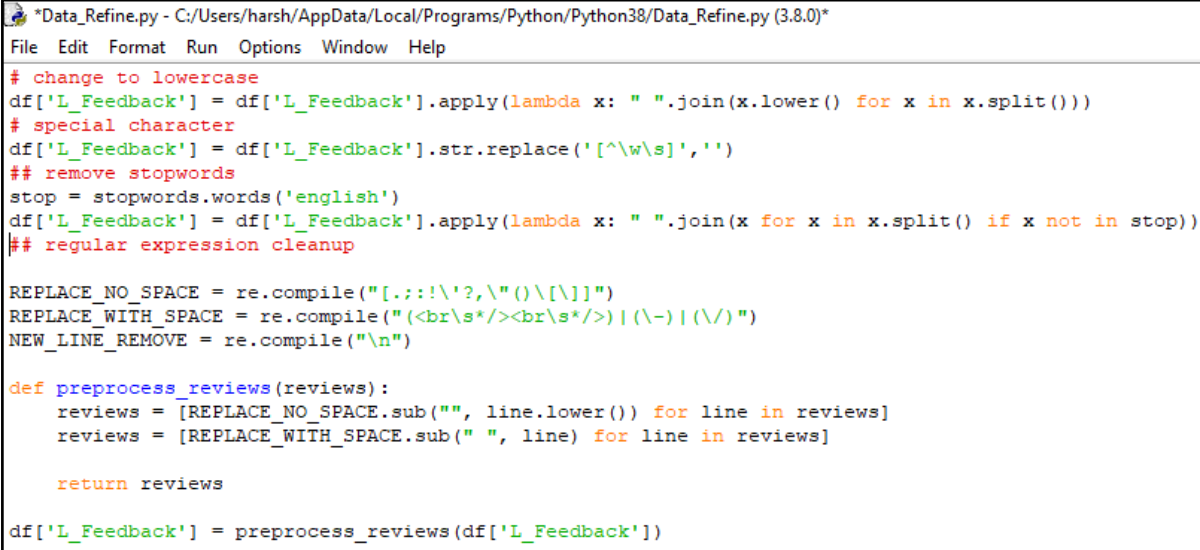
Step 1: Segment text reviews by language in an Excel spreadsheet to facilitate further analysis.

Step 2: Change the case of the entire feedback to lower case and removing special characters, stop words (words that do not add meaning, such as he, have, etc.) and the new line character in Python (see Fig. 1).

Step 3: Exclude from the concerned subject all reviews, filled in by the users, that were not valid.

Step 4: With the help of an online translator, all reviews covered by this study were translated into English no matter in what language they were originally written.

Step 5: Finally, with the assistance of an online spell checker, all reviews were revised in order to be correctly interpreted by further analysis software.

A screenshot of a Python script titled "Data_Refine.py" in a code editor. The script performs text preprocessing on a DataFrame column named 'L_Feedback'. It includes comments for each step: changing to lowercase, removing special characters, removing stopwords, and regular expression cleanup. The cleanup step involves three regex patterns: one to remove non-alphanumeric characters, one to replace line breaks with spaces, and one to remove newlines. A function 'preprocess_reviews' is defined to apply these steps to a list of reviews, and it is then applied to the DataFrame column.

```
*Data_Refine.py - C:/Users/harsh/AppData/Local/Programs/Python/Python38/Data_Refine.py (3.8.0)*
File Edit Format Run Options Window Help

# change to lowercase
df['L_Feedback'] = df['L_Feedback'].apply(lambda x: " ".join(x.lower() for x in x.split()))
# special character
df['L_Feedback'] = df['L_Feedback'].str.replace('[^\w\s]', '')
## remove stopwords
stop = stopwords.words('english')
df['L_Feedback'] = df['L_Feedback'].apply(lambda x: " ".join(x for x in x.split() if x not in stop))
## regular expression cleanup

REPLACE_NO_SPACE = re.compile("[.:!'\?,\"(){}[\]]")
REPLACE_WITH_SPACE = re.compile("<br\s*/><br\s*/>|(\-)|(\/)")
NEW_LINE_REMOVE = re.compile("\n")

def preprocess_reviews(reviews):
    reviews = [REPLACE_NO_SPACE.sub("", line.lower()) for line in reviews]
    reviews = [REPLACE_WITH_SPACE.sub(" ", line) for line in reviews]

    return reviews

df['L_Feedback'] = preprocess_reviews(df['L_Feedback'])
```

Figure 1: Code for Data Pre-processing

Procedure for Data Analysis

This work is part of a research programme to investigate the learners' sentiments regarding the retention factor in MOOCs. The challenge was to develop a better and deeper understanding from the data collected, to improve the student experience of joining and learning through a MOOC in the SWAYAM portal. In order to accomplish the tasks online, a feedback form and reviews were collected from the enrolled course learners. A mixed-methods approach, i.e., using quantitative and qualitative data, was used and it collected data from Google Forms that were constituted in two parts. The first part was a 5-point Likert-scale survey, with 25 statements about expected factors influencing retention

or dropout in MOOCs. The second part consisted of open-ended questions. The approach used for qualitative analysis was a systematic and novel multi-methodological procedure that combined word cloud visualisation (see Fig. 2); automated thematic analysis and sentiment analysis in NVivo 12 and Python. This integration of visualisations enabled us to identify five themes in order to analyse the factors of retention or dropout in MOOCs, and 21 sub-themes of relationships between MOOC motivation and retention. The main themes of the learners' feedback were identified by a manual study of each learner's feedback. Thereafter, an analysis of the text as such was performed. This was performed by using word frequency analysis (WFA) through NVivo, which gave us the number of times a word was used in the entire sample set of learner's feedback. The more frequently a word was used in the feedback, the bigger was its size in the word cloud (Figueiredo et al, 2019).

Sentiment Analysis

Sentiment analysis is the act of computationally recognising and classifying opinions stated in a text, particularly to ascertain if the writer had a favourable, negative, or neutral viewpoint on a certain subject, item, etc. (Onan, 2021). Sentiment analysis may be applied to analysing learners' reviews, opinions, sentiments and emotions towards provided education and services.

This study contains review and feedback statements about MOOC course experiences. Thereby, as a means to reach the goals proposed by this research, it will be possible to uncover and classify any relevant semantic and emotional information regarding the retention or dropout rates in MOOCs selected by this study.

Results

This paper's analysis and results reporting format adheres to theme analysis principles (e.g., Maguire & Delehunt, 2017). This paper has described the report of qualitative findings by using percentages. Table 3 shows the result, including percentages relating to theme frequency. These results have then been further interpreted in the subsequent discussion section, which includes verbatim comments from some participants.

Table 3: Main Reasons and Associated Sub-themes to Increase Retention in MOOCs

Key Themes and Subthemes	Initial Main Motive %	Motive Overall %
Base: 324 Participants	N	N
Job Promoted Trendy Course	30.5	45.4
	18	147
High demand of trendy course.	8	52
Perceived usefulness of latest trend course.	4	44
Retention of complete course.	6	51
University Policies and Processes	20.3	19.1
	12	62
Credit Mobility	5	24
Placement Policy	4	20
Certificate Credential	3	18

Key Themes and Subthemes	Initial Main Motive %	Motive Overall %
Base: 324 Participants	N	N
Social Effect	18.6	13.2
	11	43
General family commitments (like health, geographical area, economic, etc.)	4	15
Relationship and affinity.	2	8
Past experiences of friends and family circle.	5	20
Content and Language	15.25	12.3
	9	40
Workability of English language.	1	4
Localisation of content.	3	13
Localisation makes translation skill redundant.	1	4
Comprehension and retention in regional language.	2	8
English as a barrier in MOOCs.	1	2
Content must have depth in knowledge as per market requirement.	1	9
Technical Support motivation	15.3	10
	9	32
Improving assignment techniques.	3	12
Improving broadband connection.	6	18
Key Themes and Subthemes		
Behavioural Intention as a predictors in respect of the intention to continue using MOOCs		
Enhancement in learning interest due to the MOOC.		
Usefulness of MOOC material in completion of course.		
Prospects of local content in MOOC course.		
Prefer online learning		

Table 3 presents the five main reasons (themes) for MOOC retention, along with 21 associated sub-themes. In this table, the frequency or number of mentions for each theme and sub-theme is provided. The second column shows the frequency of learners that listed the themes and sub-themes as their initial main factors for continuing in the MOOCs. These themes are presented in the order of frequency of mention/occurrence. The third column in Table 3, referred to as “motive overall”, shows the frequency of all mentioned factors regarding learners’ retention in MOOCs, as well as those that the interviewer found worthy of probing. The percentages in Column 2, Initial Main Motive, therefore, relate the proportion of the total 59 participants that mentioned each theme as their main reason for retention in the MOOCs. In contrast, the percentages in Column 3, Reasons for Overall Motivation relate each theme's mention as a proportion of the total 324 learners’ motivation reasons given.

The thematic analysis identified key learner-informed motivations and associated sub-themes needed to increase retention in MOOCs (see Table 3).

Frequency Analysis

Frequency analysis was conducted in order to uncover the most mentioned words associated with motivation factors in MOOCs. This type of analysis is essential, especially at the beginning of the textual data analysis, since it can assist in identifying which factors are mostly discussed and assessed in the reviews.

Table 4: Frequency Analysis Results

S.N.	Job Related Motivation		Social and Personal Motivation		Course or Module Design Motivation		Technical Support Motivation	
	Word	Percentage	Word	Percentage	Word	Percentage	Word	Percentage
1	Job	3.30%	Advice	3.12%	Course	3.78%	Instructions	0.68%
2	Government	3.16%	Friends	2.38%	Trend	2.12%	Improve	0.81%
3	Policy	2.81%	Opinion	2.18%	Demand	1.78%	Assignment	0.94%
4	Organisation	1.35%	Family	0.78%	Module	1.15%	Connectivity	3.16%
5	Employment	0.97%	Relatives	0.72%	Content	1.11%	Instructor	2.12%
6	Profession	0.92%	Interest	0.70%	Regional	0.99%	Teacher	0.52%
7	Credit-Mobility	0.88%	Parents	0.65%	Localisation	0.98%	Interface	0.51%
8	Certificates	0.84%	Experience	0.64%	Syllabus	0.95%	Design	0.50%
9	Credit-Points	0.72%	Personal	0.58%	Knowledge	0.86%	Friendly	0.59%
10	Credibility	0.68%	Individual	0.47%	Popular	0.75%	Internet	1.21%

As shown in Table 4, after conducting the word search query, unrelated words, verbs, adjectives, articles, conjunctions, prepositions, pronouns, and others were removed. Only nouns related to MOOC attributes were maintained. Also, the word MOOCs, which was the most frequent noun in the reviews, was removed since it is not a MOOC attribute but the MOOC unit itself. The outcomes of this query support the selection of MOOC retention or MOOC dropout issues or factors.

Discussion

The findings reveal the complexities of learning difficulties in MOOCs. This study also illuminates multiple motivational factors that influence retention rate. It was observed that there is no singular factor responsible for learners' retention in MOOCs, nor is it influenced by only a single issue. When interviewed, many individuals mentioned the influence of multiple variables. However, as per the frequency analysis, Table 5 represents ten preliminary overall motivators of retention.

Table 5: The Preliminary Overall Motivators of Retention are Presented below (ranked in order of frequency)

Ten Preliminary Main Cause of Retention and Overall Motivation Factors	
1 st Job Oriented course	3 rd Perceived opinion regarding the course by learners
2 nd High demand of trendy course	4 th module designed as per market needs
5 th Content localisation	6 th Using broad band connection.
7 th Improve assignment instructions and support	8 th user friendly Interface of LMS
9 th Government Support	10 th Credit Mobility

These motivations (themes) were subsequently grouped into the following four broader categories, which are described below: job related motivation, social and personal motivation, course content or module design motivation and technical support motivation. Besides these, 21 sub-themes offer new motivation insights and are discussed below in order of their frequency of mention.

Job Related Motivation

Such type of motivation arises when learners take MOOCs to acquire skills and knowledge that will help them in their current or future job roles. Learners may be motivated by the prospect of gaining a competitive advantage in the job market or to maintain their job positions and advance their careers. Almost half of the learners related personal motivations as the main cause of their retention, which supports the observations of earlier online education researchers such as Beer and Lawson (2017). Within this category, a job or employment was the theme most frequently mentioned: 45.4% or 147 out of the 324 learners indicated that this was their most motivational factor (see Column 3 in Table 3). Six learners out of 18 (33.3%) verified the same during interviews (see Column 2 in Table 3). For example:

I prefer to join the skill oriented course to increase my job prospects.

A few of these categories have been touched on in the literature. For example, earlier literature indicated that most of the MOOC learners were working and their studies were related to their job and profession (Bayeck, 2016; Xiong et al., 2015). Lu et al. (2017) reported that there was a significant relationship between course relevancy and employment and MOOC users' satisfaction and depth of learning. Job satisfaction was a factor influencing turnover intention.

Government policies, like credit mobility and certification, were other motivations and themes mentioned, which accounted for 38.7% and 29%, respectively, of the students' initial main reasons for retention (Table 3). For example:

- (a) I prefer the course which facilitates course mobility.
- (b) The weight of the end-of-term examination in the MOOC adds value to an online course.

The National Education Policy (NEP) has made several realistic and reformative steps towards the encouragement of MOOC education. For example,

- (c) credit points are an important aspect in the National Educational Policy.
- (d) I think, I had sustained in the course due to credit point, because, my job is transferable.

As per our study no previous literature has been found regarding this particular motivation. It will be interesting to find out in future research what percentage of MOOC learners are employed and which percentage were motivated to acquire MOOC certificates or receive benefits from credit mobility.

Social and Personal Motivation

Some motives found in this study were personal and social reasons — linked to family and relationships — and perceived opinions regarding a course. Social motives were the most popular factor (about 13.2%, N = 324) to encourage participants to complete MOOCs. Family circumstances (Aldowah et al, 2020), personal interest (Watted & Barak, 2018), personal reasons (Petronzi & Hadi, 2016), and prior experiences (Greene et al., 2015) are some of the reasons that can personally motivate or de-motivate learners. Meeting new people (Uchidiuno et al., 2018), having friends join a course, and connection with others (Bayeck, 2016) affected MOOC learners' decisions to choose a course and complete it. In cases of concern about completion in a MOOC course, if any learners were influenced by the perception of their friends, family members, religion or culture, then such types of factors could be studied in the mentioned categories.

Some beginners chose courses based on a combination of what suited them and what seemed to offer good prospects after completion, according to the opinions of their families, friends or teachers. They seemed highly motivated to continue their courses as the following example shows:

My friends told me that the course is job oriented and very popular now a day, such motivation triggered me to remain in the MOOC.

During the interviews, four learners mentioned their main reason for not continuing their course as being general family commitments. In addition, five learners specified the category of perceived opinion regarding the course by other learners.

Course Content or Module Design Motivation

Course motivation refers to the potential of course structure, design, and content to cause learners to decide whether to take and complete MOOCs or not. In this study some motivations were revealed, such as module design as appropriate to market needs, high demand of trendy courses and content localisation, which 28.8% of learners (Fig. 3) mentioned as the main cause of retention in their MOOCs. In these categories learners considered many sub-themes also, such as content must have depth in knowledge as per market requirements, and the course was covered as described in the syllabus, etc. (see Table 3). One respondent explained that:

I have completed this course because it is more practical, interactive and based on the demand of market need. It also gives the facility of translation of scripts in to my regional language.

Goopio & Cheung (2020), Aldowah et al. (2020) and Said El, (2017) have emphasised the role of course design but all these studies provide much less information about specific motivation factors for module design, which 22.5% of learners mentioned as the main motivation cause when the course content had the depth of knowledge as per market requirements. For example:

Trendy Courses enhance our creativity so I will always prefer to join the MOOC course which are high in demand at market.

Thirteen percent of the students were motivated because the course covered the prescribed syllabus, and there were easy methods for communication with the e-instructor. Twenty-six percent were motivated through content localisation.

Technical Support Motivation

This refers to the extent to which technology encourages learners to engage in and finish MOOCs. The current study found the following technical motivations: improved assignment instructions and support, and a good broadband connection. Nine interviewees mentioned an internet connectivity issue and broadband network coverage issues, such as the following:

I'm not sure online education is for the country like India, where 4G speed is just like 2G and in the rural areas like my village no broadband connection is available. However, it may be specific for some region and due to more number of mobile users.

Behavioural Intention

Predicting the intention of human behaviour is an inherently complex topic that is related to the manner and reasons behind people's actions. During the study, it was observed that different learners demonstrated different behavioural intention in the same subject matter. For example, to increase the perceived effectiveness of MOOCs during course study, group discussions with students were conducted on related topics with e-instructors. The behavioural intentions of the learners were different during group discussions in different subjects. For example, in the group discussions in the Introduction to Cyber Security and Digital Forensics courses, it was seen that the rate of participation of learners was negligible in the Digital Forensics course.

To continue with this topic one interviewee noted that:

Group conversations encourage deeper understanding of a topic and improve long-term retention.

While one of the other participants countered that:

The group conversation tool used in MOOC confuses many learners.

Word Frequency Analysis

Cloud visualisation (Fig. 2) about learners' views indicated the most frequent words were as follows: job/Trend/popular/Market-demand/job-oriented (142), friends (45), opinions (40), content (29), and internet (26). The outcomes of the word frequency analysis validated the key themes and sub-themes which were investigated and have been discussed in the above section.

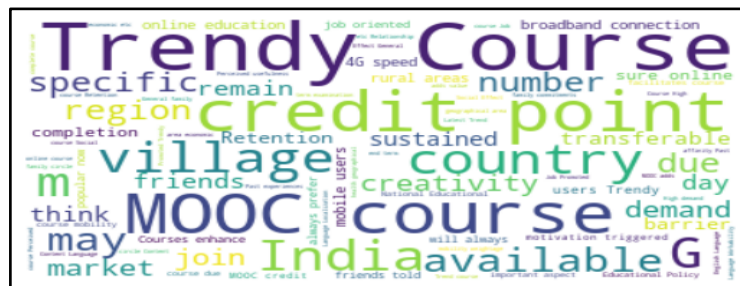


Figure 2: Word Cloud of MOOC Platform

MOOC Learners' Retention Strategy — A Framework for Evaluating and Responding to Retention Causes

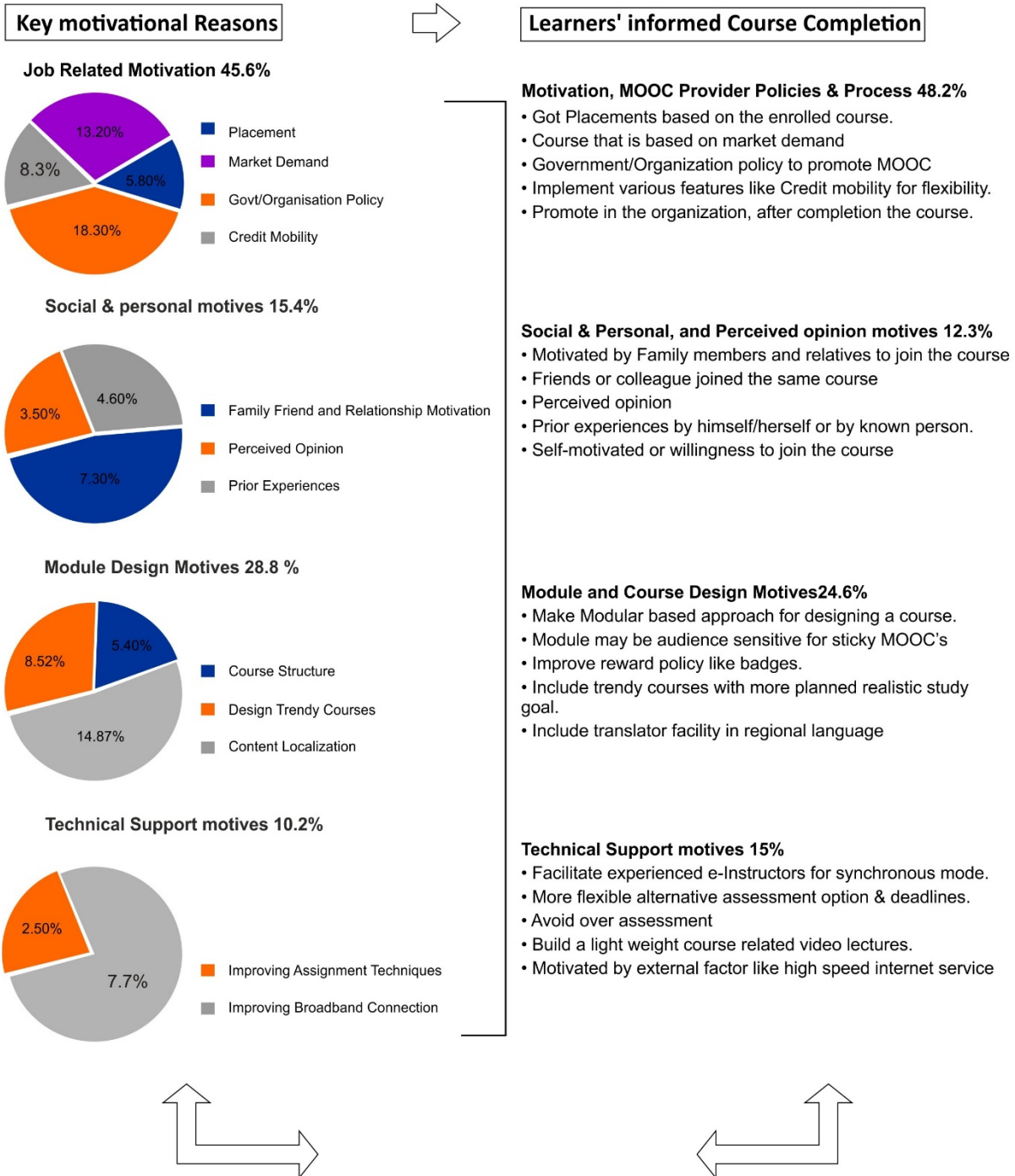


Figure 3: Conceptual Strategic Framework for Evaluating Retention in MOOCs

The literature provides a number of recommendations for improving the learner retention rate. For example, Diego et al. (2020) developed a modelling framework to maximise the effectiveness of retention efforts. However, their data set in this research was based on academic and pre-matriculation information. Additional data sets may supply information relevant to retention because student dropout is a context-specific phenomenon. In the context of developing effective retention strategies, de Oliveira et al. (2020) reviewed the factors that prevented students' dropout in higher education. But in more recent research authors have focused specifically on MOOC-based factors that prevent retention, based on the sentiment analysis of the learners' feedback. Adhering to this perspective, Figure 3, which is based on the outcome of this study, presents a strategic framework for evaluating and responding to the causes of retention within the context of a Massive Open Online Course. It presents a concise synthesis of this study's thematic analysis and proposed strategic retention perspective. It provides insights into the contribution of the factors that influenced retention in MOOCs. It also depicts the design and prioritisation of appropriate retention initiatives. The right-hand side of the framework lists the four broad motivational categories (i.e., job related, social and personal, module design and technical support motivation), which are made up of the four key retention themes and 16 associated sub-themes. The percentages and the physical size of the pie charts and their segments reflect the contribution of each category and theme to the retention of learners. The right-hand side lists the four retention themes, and the sub-themes that reflect learners' suggestions about ways that could have helped them complete their course. Lastly, the capturing of job-related and planned module design motivation, as defined in this study, could be used as a better motivational approach to students continuing and remaining in the course.

Evaluation of the above bulleted list in Figure 3 represents the main reasons for dropouts and motivations for retention in MOOCs. These learners' suggestions could prevent dropout from happening and could better enable universities to develop a student-informed retention strategy.

Conclusion

In a densely populated country like India, MOOCs have experienced tremendous growth. However, they suffer low retention rates that cause concern for online educators by creating negative perceptions about online learning (Stone & O'Shea, 2019; Tang & Chaw, 2019). A considerable amount of research has therefore been investigated to try to improve retention. Despite substantial research, a lack of in-depth understanding of retention in MOOCs has been cited as the essential problem, with more detailed investigation being called for (e.g., Lamon et al, 2020; Xavier & Meneses, 2020; Pant et al., 2021). During the COVID-19 pandemic there was more need for MOOCs, so it will be essential to find out the retention factors, if any, and predict the actual reasons for learner dropout. The current research represents the value of using a larger qualitative probability-based sample, in combination with in-depth interviewer probing and thematic analysis to find out the factors of retention. The thematic analysis identified five main themes for MOOC retention, along with 21 associated sub-themes and ten motivational factors. Recent studies have highlighted some particular themes, like government policies and trendy courses, along with some general factors, like content localisation and credit mobility. During our investigation, some common factors that had already been uncovered in earlier studies re-appeared, such as course content, social and technical factors, which validated the previous research with a different sample from a large context such as India (using a major platform like SWAYAM) and with different demographic data.

As with most research, there were limitations in this study that should be acknowledged. In particular, this study only examined feedback from specific courses in SWAYAM, which meant the results may not always be generalisable. Other limitations included the limited potential for statistical analysis (e.g., Cooper & Schindler, 2014) which is typical within qualitative investigations.

References

- Aldowah, H., Al-Samarraie, H., Alzahrani, A. I., & Alalwan, N. (2020). Factors affecting student dropout in MOOCs: A cause and effect decision-making model. *Journal of Computing in Higher Education*, 32, 1-26. doi:<https://doi.org/10.1007/s12528-019-09241-y>
- Aytuğ, O. (2019). Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach. *Computer Applications in Engineering Education*, 1-18. doi:10.1002/cae.22253
- Bayeck, R. (2016). Exploratory study of MOOCs learners' demographics and motivation: The case of students involved in groups. *Open Praxis*, 223–233. <https://doi.org/10.5944/openpraxis.8.3.282>
- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 18(2), 15-21. doi:10.1109/MIS.2013.30.
- Cooper, D.R., & Schindler, P. S. (2014). *Business research methods* (12th ed.). McGraw Hill International.
- Cooperman, L. (2014). Foreword. In A. Okada (Ed.), *Open Educational Resources and social networks: Co-Learning and professional development*. Scholio Educational Research & Publishing.
- Dalipi, F., Zdravkova, K., & Ahlgren, F. (2021). Sentiment analysis of students' feedback in MOOCs: A systematic literature review. *Frontiers in Artificial Intelligence*, 728708(4). doi:<https://doi.org/10.3389/frai.2021.728708>
- de Oliveira, C., Sobral, S., Ferreira, M., & Moreira, F. (2020). How does learning analytics contribute to prevent students' dropout in higher education: A systematic literature review. *Big Data and Cognitive Computing*, 5(64). doi:<https://doi.org/10.3390/bdcc5040064>
- Diego, O., Jonathan, V., Sebastián, M., Jaime, M., & Wouter, V. (2020). Uplift modeling for preventing student dropout in higher education. *Decision Support Systems*, 134. doi:<https://doi.org/10.1016/j.dss.2020.113320>.
- Figueiredo, R., Quelhas, O., Vieira Neto, J., & Ferreira, J. (2019). The role of knowledge intensive business services in economic development: A bibliometric analysis from Bradford, Lotka and Zipf laws. *Gestão & Produção*, e4356. doi:<https://doi.org/10.1590/0104-530X4356-19>
- Goopio, J., & Cheung, C. (2020). The MOOCs dropout phenomenon and retention strategies. *Journal of Teaching in Travel & Tourism*. doi:<https://doi.org/10.1080/15313220.2020.1809050>
- Greene, J.A., Oswald, C.A., & Pomerantz, J. (2015). Predictors of retention and achievement in a Massive Open Online Course. *American Educational Research Journal*, 52(5), 925–955. doi:<https://doi.org/10.3102/0002831215584621>
- Greenland, S.J., & Moore, C. (2021). Large qualitative sample and thematic analysis to redefine student. *British Journal of Educational Technology*, 1-21. doi:<https://doi.org/10.1111/bjet.13173>
- Lamon, S., Knowles, O., Hendy, A., Story, I., & Currey, J. (2020). Active learning to improve student learning experiences in an online postgraduate course. *Frontiers in Education*, 5, 1-10. doi:<https://doi.org/10.3389/feducation.2020.00010>
- Maguire, M., & Delehunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Teaching and Learning in Higher Education*, 9(3), 3351–3365.
- Mika, V.M., Daniel, G., & Miikka, K. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, 16-32. doi:<https://doi.org/10.1016/j.cosrev.2017.10.002>

- Onan, A. (2021). Sentiment analysis on Massive Open Online Course evaluations: A text mining and deep learning approach. *Computers in the Application of Engineering Education*, 29, 572–589. doi:10.1002/cae.22253
- Pant, H.V., Lohani, C.M., & Pande, J. (2021). Exploring the factors affecting learners' retention in MOOCs: A systematic literature review. *IJICTE*, 17(4), 1-17. doi:http://doi.org/10.4018/IJICTE.20211001.0a3
- Petronzi, D., & Hadi, M. (2016). Exploring the factors associated with MOOCs engagement, retention and the wider benefits for learners. *European Journal of Open, Distance and E-Learning*, 112-129. doi:https://doi.org/10.1515/eurodl-2016-0011
- Said El, R.G. (2017). Understanding how learners use Massive Open Online Courses and why they drop out: Thematic analysis of an interview study in a developing country. *Journal of Educational Computing Research*, 55(5), 724–752. doi:https://doi.org/10.1177/0735633116681302
- Scott, S., & Buzzlogix. (2015). *Sentiment Analysis 101*. <https://www.kdnuggets.com/2015/12/sentiment-analysis-101.html>
- Sraidi, S., Smaili, E., Azzouzi, S., & Charaf, M. (2022). A sentiment analysis based approach to fight MOOCs' drop out. In *Networking, intelligent systems and security. Smart innovation, systems and technologies* (Vol. 237, pp. 509-520). Springer. doi:https://doi.org/10.1007/978-981-16-3637-0_36
- Stone, C., & O'Shea, S. (2019). Older, online and first: Recommendations for retention and success. *Australasian Journal of Educational Technology*, 1, 57–69. doi:DOI:10.14742/ajet.3913
- Tang, C.M., & Chaw, L. (2019). Driving high inclination to complete Massive Open Online Courses (MOOCs): Motivation and engagement factors for learners. *Electronic Journal of e-Learning*, 17, 118-130. doi:https://doi.org/10.34190/JEL.17.2.05
- Uchidiuno, J.O., Ogan, A., Yarzebinski, E., & Hammer, J. (2018). Going global: Understanding English language learners' student motivation in English-language MOOCs. *International Journal of Artificial Intelligence in Education*, 28(4), 528–552. doi:https://doi.org/10.1007/s40593-017-0159-7
- Watted, A., & Barak, M. (2018). Motivating factors of MOOCs completers: Comparing between university-affiliated students and general participants. *The Internet and Higher Education*, 37, 11-20. doi:https://doi.org/10.1016/j.iheduc.2017.12.001
- Xavier, M., & Meneses, J. (2020). *Dropout in online higher education: A scoping review from 2014 to 2018*. eLearn Center, Universitat Oberta de Catalunya. doi:https://doi.org/10.7238/uoc.dropout.factors.2020
- Xiong, Y., Li, H., Kornhaber, M.L., Suen, H. K., P.B., & Goins, D. D. (2015). Examining the relations among student motivation, engagement, and retention in a MOOC: A structural equation modeling approach. *Global Education Review*, 2(3), 23-33.

Authors:

Harsh Vardhan Pant is an Assistant Professor at the Amrapali Institute, Haldwani, India and is pursuing a PhD in Computer Science from Graphic Era Hill University, Bhimtal Campus, India. The key area of his research is in the analytical study in a distance education domain such as MOOCs. Email: pant.vardhan@gmail.com

Manoj Chandra Lohani is a Professor and Head (CSE) and Campus Director, Graphic Era Hill University, Bhimtal, India. As well as a PhD in Computer Science, he holds an MS degree in Cyber Security and Cyber Laws. With more than 20 years in teaching and research, Prof. Manoj Chandra Lohani gives the Bhimtal Campus of GEHU both academic and administrative leadership. Email: getmlohani@gmail.com

Jeetendra Pande is an Associate Professor, Computer Science, Uttarakhand Open University, Haldwani-263139, INDIA. Dr. Pande has published over 18 research papers in international journals, 15 papers in conference proceedings, and three refereed books. His research interests include cyber security, computer forensics, component-based software development, education technology, and Open Educational Resources. He has been

involved in many projects as a Principal Investigator sanctioned by the government of India, government of Uttarakhand and Commonwealth Educational Media Center for Asia. He is a member of several academic and professional bodies in the area of computer science and distance education. Email: jpande@uou.ac.in

Cite this paper as: Pant, H.V., Lohani, M.J., & Pande, J. (2023). Thematic and sentiment analysis of learners' feedback in MOOCs. *Journal of Learning for Development*, 10(1), 38-54.

Notes

ⁱ https://onlinecourses.swayam2.ac.in/nou22_cs09/preview

ⁱⁱ https://onlinecourses.swayam2.ac.in/nou22_cs09/preview

ⁱⁱⁱ https://onlinecourses.swayam2.ac.in/nou22_ge51/preview