On the Need for Fine-Grained Analysis of Gender Versus Commenting Behaviour in MOOCs

Mohammad Alshehri
Computer Science
Warwick University
malshehri@uj.edu.sa

Jonathan Foss
Computer Science
Warwick University
jonny@dcs.warwick.ac.uk

Alexandra I. Cristea
Computer Science
Warwick University
a.i.cristea@warwick.ac.uk

Mizue Kayama
Faculty of Electrical and Computer Engineering
Shinshu University
kayama@cs.shinshu-u.ac.jp

Lei Shi
Centre for Educational Development and Support
Liverpool University
Lei.Shi@liverpool.ac.uk

Ahmed Alamri
Computer Science
Warwick University
a.alamri.1@warwick.ac.uk

Adam Tsakalidis
Computer Science
Warwick University
a.tsakalidis@warwick.ac.uk

ABSTRACT

Stereotyping is the first type of adaptation ever proposed. However, the early systems have never dealt with the numbers of learners that current Massive Open Online Courses (MOOCs) provide. Thus, the umbrella question that this work tackles is if learner characteristics can predict their overall, but also fine-grain behaviour. Earlier results point at differences related to gender or to age. Here, we analyse gender versus commenting behaviour. Our fine-grained analysis shows that the result may further depend on the course topic, or even week. Surprisingly, for instance, women chat less in a Psychology-related course, but more (or similar) on a Computer Science course. These results are analysed in this paper in details, including two different methods of averaging comments, leading to remarkably different results. The outcomes can help in informing future runs, in terms of potential personalised feedback for teachers and students.

CCS Concepts

•Applied computing → Education; E-learning; • Human-centered computing → Human computer interaction (HCI); Interaction paradigms, Web-based interaction • Human-centered computing → HCI design and evaluation methods; User studies; User models.

Keywords

Learner characteristics; stereotypes; MOOCs; FutureLearn; online behaviour prediction.

1. INTRODUCTION

Stereotyping is one of the earliest user modelling approaches to adaptation and recommendation. It was first introduced by Rich in a book recommender system, Grundy [1], which built models for individual users, based on personal information, gathered through interactive dialogues. A stereotype is a collection of physical characteristics or frequently occurring characteristics of individual users, such as gender, age, engagement, performance and so on. Creating stereotypes has become a common approach to user modelling – it uses a small amount of initial information to adopt a large number of default assumptions [2] which may be updated when more information about individuals becomes available [1]. Stereotyping has been criticised as being too simplistic, and then, again, applied, due to its simplicity.

With the advent of the MOOCs, past stereotypes can be evaluated once again at a much larger scale than by preceding research, and confirmed or infirmed. Whilst MOOCs have started being analysed more thoroughly in the literature, few researches, as will be seen, are looking into the temporal, fine-grained analysis of the behaviour, and establishing any relation between the learner behaviour and learner stereotypes.

Our main purpose with this research is to predict the learner overall and fine-grain behaviour based on learner characteristics. In this paper, we specifically focus on the gender stereotype, and its relation to the way learners comment in a MOOC. We base our study on a truly massive FutureLearn course collection of 7 courses delivered via 27 runs between 2012-2016.

ICIEI'18, June 30 – July 2, 2018, London, UK.

DOI: http://dx.doi.org/

2. RELATED RESEARCH

As in educational systems, there are two types of stereotyping: fixed and default [2]. A fixed stereotyping classifies learners based on their performance, into predefined stereotypes, which are determined by, for example, their academic level. In a default stereotyping, a learner is usually stereotyped to default values at the beginning of a learning session; then the settings of the initial stereotype may be gradually altered, as the learning process proceeds and more behavioural data is collected [3].

A large body of research has been conducted to explore whether and how learner characteristics can predict their behaviours. Jeske et al. [4] suggest that self-reported learning characteristics can add an important perspective on why and how different learners have different patterns of performance and behaviour while learning. Packham et al. [5] find that successful learners are female, aged between 31 and 50, regardless of their educational level and employment status. Ke and Kwak [6] report that older learners invest more time in online participation. González-Gómez et al. [7] suggest that males have more positive attitudes towards online learning, due to their higher computer self-efficiency. Many earlier results point at differences of behaviours related to characteristics such as age and gender. Vail et al. [8] show that females and male students benefit differently from adaptive support.

Over the last six years, massive open online courses (MOOCs) have become increasingly popular and their scale and availability enable a diverse set of learners worldwide to take online courses. In the meantime, the amount of learner data collected, including demographic data and behavioural data, has also been increasing. This provides an unprecedented opportunity to further explore the influence of learner characteristics on their behaviours. One approach to understanding learners on MOOCs is by identifying groups of learners with similar behavioural patterns [9] such as clustering learners using engagement factors, including the number quizzes attempted [10], [11]. Chua et al. [12] and Tubman et al. [13] analyse learner commenting behaviours, to explore patterns of discussion that occur in MOOCs.

On the other hand, comments have been studied in many setups, including MOOCs. [14] emphasises the importance of using machine learning methods to analyse MOOCs comments, to detect the emotions of learners and predict the popularity of each course. [15] focused on grouping students based on their preferences, by conducting an online pre-course survey. According to these groups, the relationship between gender showed that females preferred asynchronous text-based posts more. [16] investigated the dropout rate, via analysing two MOOC courses with 176 learner's comments on different objects (video, articles, exercises etc.). The study indicated that learners with no negative comments are likely to drop the course very soon. [17] explored the relationship between sentiment ratio measured based on daily forum posts and the number of learners who drop out each day. The study recommended to use sentiment analysis with caution, while analysing noisy and quantity-limited comments.

Our study examines how basic learner characteristics, such as gender, can influence learning behaviours, such as the patterns of making comments.

3. METHODOLOGY

3.1 Terminology

FutureLearn is a MOOC online education platform that provides courses upon weekly basis. Each weekly learning unit consists of several steps, which can be an article, discussion, video or a quiz. The website also allows learners to comment on any given step.

3.2 Data Collection

When a learner joins FutureLearn for the first time, they are directly prompted to complete a survey about their characteristics. Existing learners are also prompted to complete this data, if missing. All the questions on the survey are optional and they aim to extract certain information about a learner, such as gender, age group and education level. In parallel, the system generates logs "to correlate unique IDs and time stamps to learners", recording learner activities, such as steps visited, completed, comments added or question attempts.

3.3 Dataset

The current study is analysing data extracted from 27 runs of 7 MOOCs courses, on 4 main topics: literature (Literature and Mental Health (LT): 6 Weeks), Shakespeare and his world (SP): 10 Weeks; psychology (The mind is flat (MF): 6 Weeks), Babies in mind (BIM): 4 Weeks; computer science (Big Data (BD): 9 Weeks), and business (Leadership (LS): 6 weeks and Supply chains (SC): 6 Weeks) delivered through FutureLearn, by the University of Warwick. The study covers 19425 female and 6648 male enrolled learners, out of which 11473 female and 3802 male learners have accessed the course material at least once, and out of which 6240 females and 1833 males have commented at least once. The material overall has a total number of 2590 steps.

3.4 Counting Comments

This paper focuses on comments of female and male learners. We have started by looking at overall numbers, such as total number of comments, $NComm$. To understand gender differences, we looked at the total number of comments posted by women: $NComm^F$, or by men: $NComm^M$. However, this was not enough: to obtain fine-grained, temporal results, we had to analyse comments on a weekly basis, i.e., to trace $NComm(wi)$, the number of comments per week wi, or, more precisely, $NComm^F(wi)$ and $NComm^M(wi)$, i.e., the number of comments written by women and man per week wi, respectively. However, the number of women and men in different courses varied – some were subscribed predominantly by women, others by men. Thus, to compare on a fairer basis, we have further averaged the comments of males and females, computed via two versions of formulaes, as below.

3.4.1 Version 1: access average (NFA/NMA)

Version 1 averages behavioural activity (comments) based on the global number of students (female/ male) active in the course, by accessing it. For females, this average is (Eq. 1):

$$NFA(wi) = (\frac{1}{NA^F}) \sum_{k=1}^{N^F(wi)} NComm^F\,(wi) \tag{1}$$

where $N^F$is the total (global) number of females enrolled in the course over all runs; $NA^F$is the total (global) number of females that have accessed the course, for all runs; the rest of the variables have been defined above. For males, the average is (Eq. 2):

$$NMA(wi) = (\frac{1}{NA^M}) \sum_{k=1}^{N^M(wi)} NComm^M\,(wi) \tag{2}$$

where NA^M is the total number of males who have accessed the course, for all runs; the rest of the parameters is as defined above. These formulas already are fairer: they take into account that the gender with most accesses might have posted most comments. However, these numbers still consider many students who may have accessed the course, but have never commented on it. As the goal here is to analyse comments in particular, the next formula deals with this issue.

3.4.2 Version 2: commenting average (NFC/NMC)

Version 2 averages behavioural activity (comments) based on the global number of students (female/ male) active in the course, by commenting (at some point – not necessarily that week). For females, the average is (Eq. 3):

$$NFC(wi) = \left(\frac{1}{NC^F}\right) \sum_{k=1}^{N^F(wi)} NComm^F(wi) \quad (3)$$

where NC^F is the total (global) number of females that have commented the course, for all runs, at some point; the rest of the parameters is as defined above. For males, the average is (Eq. 4):

$$NMC(wi) = \left(\frac{1}{NC^M}\right) \sum_{k=1}^{N^M(wi)} NComm^M(wi) \quad (4)$$

where NC^M is the total (global) number of males that have commented the course, for all runs, at some point; the rest of the parameters is as defined above.

3.4.3 Comparing the two Versions

As can be seen, as for both versions we divide via a constant, the shape of the resulting graphs would be the same (although the overall values would change, depending on the number of women accessed/ or having commented, in general, on the course). As the number of students who access the course is greater than the number of students who comment (as some students are just 'lurking' in the background, without committing), we have (5,6):

$$NA^F > NC^F \quad (5)$$

$$NA^M > NC^M \quad (6)$$

Thus, the following inequations (7, 8) also hold:

$$NFA(wi) < NFC(wi) \quad (7)$$

$$NMA(wi) < NMC(wi) \quad (8)$$

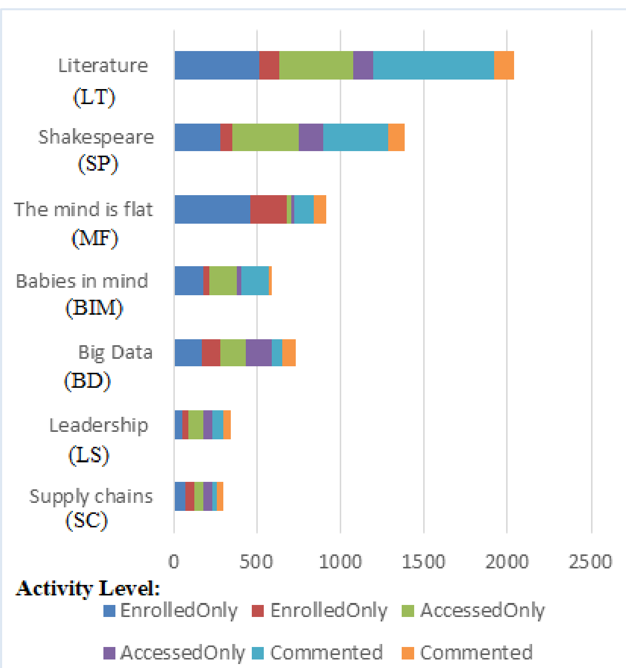

Figure 1. Number of Female and Male students per all runs of each course, split by different levels of activity.

4. RESULTS

4.1 Overall Comments per Gender

Figure 1 shows the numbers of students who were enrolled on average on each course, ordered by number of students enrolled (Ox representing absolute numbers). The most popular courses were clearly on the literature topic. However, of the 6099 students enrolled on the LT course over its 3 runs, only 4214 (69%) students accessed the course at all. Furthermore, only 2513 of those students made any comments. Furthermore, although the Psychology course MF was one of the most popular courses to enroll on, only 26.5% of those enrolled on the course accessed it.

4.2 Average Comments per Learner

Whereas the above results look at the proportion of male and female learners who made comments, the analysis further looks at how many comments were made for each course, at the fine granularity level of the week. This analysis considers the average number of comments made by all learners who commented on the course at least once (solid line; Version 2, Section 3.4.2), and all learners who accessed the course at least once (dotted line; Version 1 in Section 3.4); additionally, male learners are shown with a blue line and female learners are represented by a red line.

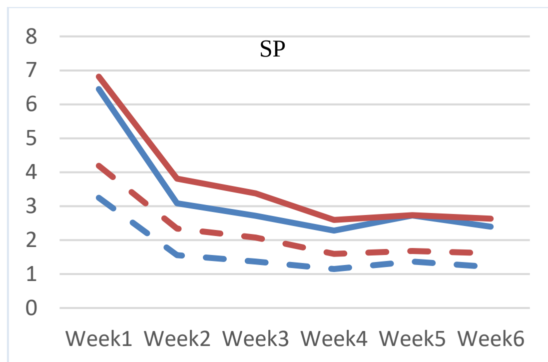

Figure 2. Literature topic (SP: Shakespeare): comments per learner (version 1 -solid & version 2-dotted; female -red/ male -blue).

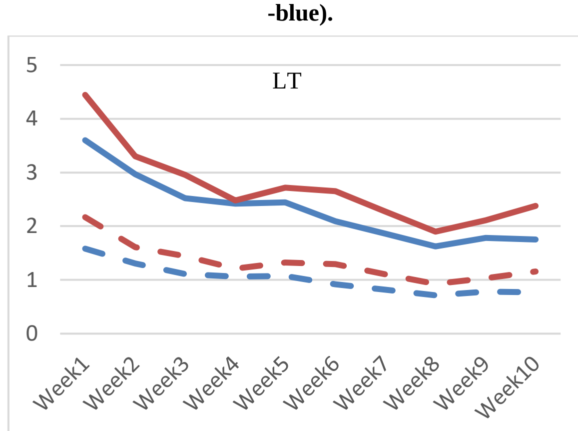

Figure 3. Literature topic (LT: Literature): comments per learner (version 1 -solid & version 2-dotted; female -red/ male -blue).

Figures 2 and 3 show that for the Literature topic, on average, there were more comments made by female learners than male learners. For Version 1, this difference is consistently statistically significant ($p < 0.05$; Wilcoxon signed ranked test, due to non-normal distribution), but for Version 2, the difference is only significant for weeks 2, 3 and 6 (MF) and for weeks 1, 3, 6 and 7 (SP).

Figure 4 shows a close gender balance for the MF course. However, for weeks 3, 5 and 6 there is a statistically significant ($p < 0.05$ for the Wilcoxon signed rank test) difference when considering only the subgroup of learners who made any comment (Version 2). For the BIM course (Figure 5), on average, female learners made more comments than male learners, although not statistically significantly so. However, when considering all learners who accessed the course (Version 1), there is a significant difference for every week ($p < 0.05$ for the Wilcoxon signed rank test).

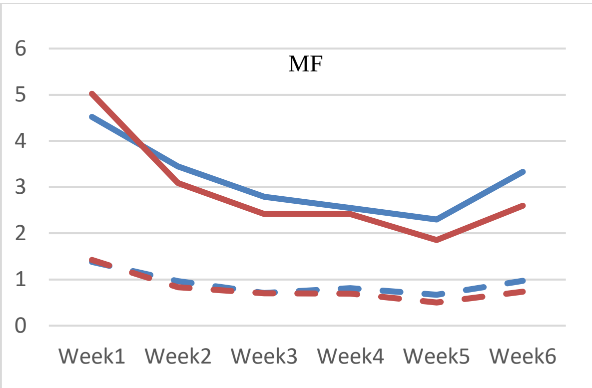

Figure 4. Psychology topic (MF: The mind is flat): comments per learner (version 1-solid& version 2-dotted; female-red/ male-blue).

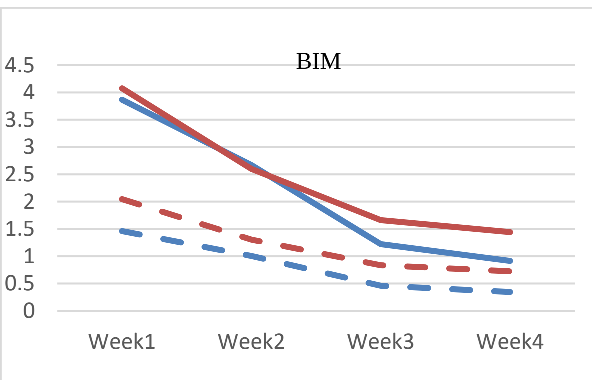

Figure 5. Psychology topic (BIM: Babies in mind): comments per learner (version 1-solid& version 2-dotted; female-red/ male-blue).

Figure 6 shows that male learners of the "Big Data" course made on average more comments than female learners. None of these differences is statistically significant, apart from Week 3 ($p < 0.05$ for the Wilcoxon signed rank test). This significance occurs when considering both subgroups. During week 7, there were more comments made by female learners than male learners, however this is not statistically significant.

Figures 7 and 8 shows that male learners of both business courses made on average more comments than female learners, but none of these differences are statistically significant. The only statistical significance ($p < 0.05$) relates to weeks 2 and 6 for SC, when considering Version 1.

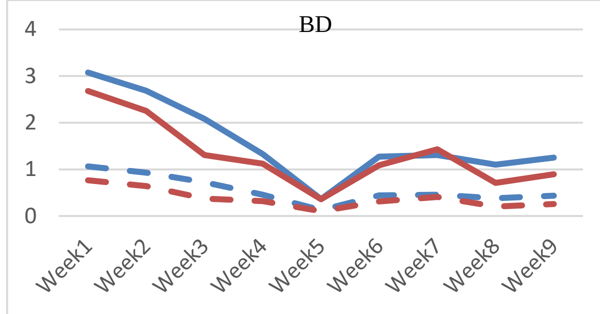

Figure 6. Computer Science topic (BD: Big Data): comments per learner (version 1-solid & version 2-dotted; female-red/ male-blue).

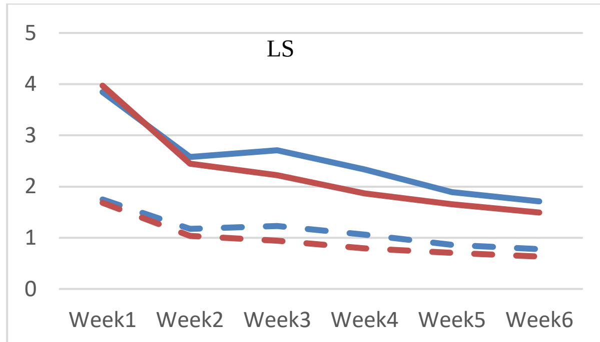

Figure 7. Business topic (LS: Leadership): comments per learner (version 1-solid&2 - dotted; female-red/ male-blue).

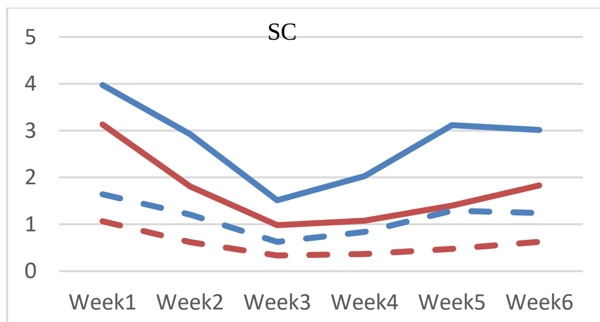

Figure 8. Business topic (SC: Supply chains): comments per learner (version 1-solid&2 - dotted; female-red/ male-blue).

5. DISCUSSION

The analysis in this paper has highlighted a number of issues which may have been predictable, as well as a few surprises. Firstly, overall, in the courses we have analysed, there are generally speaking more females registered than males. We have also been able to make statements with statistical significance, in general, for the larger courses, such as the literature courses, which were the most popular, followed by the Psychology courses. Computer Science courses are only marginally more popular than Business courses, in our selection.

We have shown that grouping the courses per topic made sense, and that results were relatively similar within such groups. The latter may be some special case, or this might need to be revisited, e.g., by a teacher of that subject, to check the appropriateness of the classification, and the match between real and desired outcomes.

Importantly, the way the average of comments per learner is computed influences the significance of the results (and, in some cases, the results themselves). Due to the great differences between learners who access the course, or learners who actually comment, in terms of numbers, the conclusions need to clearly vary, when speaking of one cohort or the other.

Expectations in terms of volume of comments coming from female or male learners clearly vary thus with the topic of the course. Therefore, whilst global statements across courses should best be avoided, it is useful to see how students react to a specific course, and then plan for future runs, accordingly. This would help a teacher better understand how to structure the course in a more gender-neutral way, and be enticing to both genders. Furthermore, learners could be notified of options which are targeted to their respective gender. Specific weeks can be analysed when they are triggering behaviour different from the rest of the course – e.g., week 7 in the Computer Science course (see Figure 4), where more female learners comment; or week 6 on the Business topic (SC; Figure 5).

6. CONCLUSION

This paper is advocating the need for fine-grained analysis of behaviour analysis in general, and, in particular, when analysing how the gender may influence behaviour such as commenting. Our analysis shows that, overall, whilst the participation of females is clearly larger in terms of absolute numbers in the relatively varied MOOC courses we have analysed, in terms of comments produced by the two genders, the topic of the course, the course itself, and often, the week of the course determines which of the genders is commenting more often.

Thus, this study clearly shows that it is not enough to study such data on a global scale, as has been done in past studies - because adding up data over several courses with different topics, and over different weeks, may render deceiving results.

Moreover, this study has found several significant differences in the behaviour of female and male learners, in terms of their commenting frequency, at a very fine granularity level: here, at the level of the week of a course. Hence, further studies should look into how the topic and time-scale together influence the behaviour of female and male learners for other courses – as possibly other interesting patterns may emerge.

Furthermore, here, we only focussed on one stereotype parameter – gender – and one behavioural parameter – commenting. Future research will include a greater variety of such parameters, for extracting a richer picture of how learner characteristics influence learner behaviour in massive online learning environments.

7. REFERENCES

[1] E. Rich, “User Modeling via Stereotypes,” *Cogn. Sci.*, vol. 3, no. 4, pp. 329–354, Oct. 1979.

[2] V. J. Shute and D. Zapata-Rivera, “Adaptive Educational Systems,” *Adapt. Technol. Train. Educ.*, vol. 7, no. 27, pp. 1–35, 2012.

[3] J. Kay, “Stereotypes, Student Models and Scrutability,” in *Intelligent Tutoring Systems*, 2000, 19–30.

[4] D. Jeske, C. Stamov-Roßnagel, and J. Backhaus, “Learner Characteristics Predict Performance and Confidence in e-Learning: An Analysis of User Behaviour and Self-evaluation,” *J. Interact. Learn. Res.*, vol. 25, no. 4, pp. 509–529, Oct. 2014.

[5] G. Packham, P. Jones, C. Miller, and B. Thomas, “E‐learning and Retention: Key Factors Influencing Student Withdrawal,” *Educ. Train.*, vol. 46, no. 6/7, pp. 335–342, Aug. 2004.

[6] F. Ke and D. Kwak, “Online Learning Across Ethnicity and Age: A Study on Learning Interaction Participation, Perception, and Learning Satisfaction,” *Comput. Educ.*, vol. 61, pp. 43–51, Feb. 2013.

[7] Q F. González-Gómez, J. Guardiola, Ó. Martín Rodríguez, and M. Á. Montero Alonso, “Gender Differences in E-learning Satisfaction,” *Comput. Educ.*, vol. 58, no. 1, pp. 283–290, Jan. 2012.

[8] A.Vail, K. Boyer, E. Wiebe & J. Lester “The Mars and Venus Effect: The Influence of User Gender on the Effectiveness of Adaptive Task Support”. Proc. User Modeling, Adaptation and Personalization, UMAP’2015, July, Dublin, 265-276, 2015.

[9] Z. Liu *et al.*, “MOOC Learner Behaviors by Country and Culture; an Exploratory Analysis.,” *EDM*, vol. 16, pp. 127–134, 2016.

[10] R. Ferguson and D. Clow, “Examining Engagement: Analysing Learner Subpopulations in Massive Open Online Courses (MOOCs),” 2015, pp. 51–58.

[11] R. F. Kizilcec, C. Piech, and E. Schneider, “Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses,” in *Proceedings of the third international conference on learning analytics and knowledge*, 2013, pp. 170–179.

[12] S. M. Chua, C. Tagg, M. Sharples, and B. Rienties, “Discussion Analytics: Identifying Conversations and Social Learners in FutureLearn MOOCs,” 2017.

[13] P. Tubman, M. Oztok, and P. Benachour, “Being Social or Social Learning: A Sociocultural Analysis of the FutureLearn MOOC Platform,” in *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, 2016, pp. 1–2.

[14] Z. Liu, W. Zhang, J. Sun, H. N. Cheng, X. Peng, and S. Liu. 2016. Emotion and associated topic detection for course comments in a MOOC platform. In *Educational Innovation through Technology* (EITT), 2016 Int. Conf. on IEEE, 15-19.

[15] Q. Zhang, K.L., Peck, A. Hristova, K.W. Jablokow, V. Hoffman, E. Park, and R.Y. Bayeck. 2016. Exploring the communication preferences of MOOC learners and the value of preference-based groups: Is grouping enough? *Educational Technology Research and Development* 64, 809-837.

[16] N. Dmoshinskaia. 2016. Dropout prediction in MOOCs: using sentiment analysis of users' comments to predict engagement University of Twente.

[17] M. Wen, D. Yang, and C. Rose. 2014. Sentiment Analysis in MOOC Discussion Forums: What does it tell us? In *Educational data mining* 2014, 130-137.