

Learning Behavior of Men and Women in MOOC Discussion Forums – A Case Study

Catrina John, Christoph Meinel
Hasso Plattner Institute, University of Potsdam, Potsdam, Germany
{catrina.john, christoph.meinel}@hpi.de

Abstract—Recent sociotechnical innovations in online communication make user-content interaction, dynamic user-user interaction and user-generated content interesting for content analysis, but raise complex challenges of reliable and valid coding as well as content analytical interpretations. We approach these new issues by analyzing the asynchronous communication and discourses in an exemplary Massive Open Online Course (MOOC) in IT security with 4,203 enrollments and 1,343 posts on 192 topics in the forum by 274 different users. We analyze the user-driven contributions to forum discussions using mixed methods and answer the question of whether men and women raise different questions in the discussion forums and if yes, with what kind of effects. We show that many women prefer closed and consent questions, as well as rhetorical questions over discussion, hypothetical questions, and contact requests. The question type significantly influences the way of answering: Direct and open questions particularly encourage fellow learners to answer in a helpful way and in a short time of two to three hours. Responses to questions asked by women were characterized more often by detailed explanations, a simple reply or expression of appreciation, and less often with vagueness. In other words, no one repeated the initial questions asked by women and none of their questions were met with a reaction of amazement. The qualitative part of our content analysis revealed that women's discussion posts imply more often uncertainty or a slightly lower level of self-efficacy than men's contributions. Forum discussions took place particularly in the afternoon from 2 to 6 p.m. CET, a time period when many women have other obligations, therefore making it less than ideal. We conclude by making recommendations based on our findings for improvements in supporting underrepresented groups in online learning in science, technology, engineering, and mathematics (STEM) and discuss potential limitations of our case study.

Keywords— Massive Open Online Course (MOOC), Five-step model of interactivity, Rich data, Mixed methods

I INTRODUCTION

Massive Open Online Courses (MOOCs) play a key role in contemporary continuing education in the fields of science, technology, engineering and mathematics (STEM) at a cutting-edge level. Such courses open the latest scientific findings and valuable learning content of prestigious educational institutions for everybody at anytime, anywhere, and free of charge. The MOOC forum offers a lively and simultaneous dialogue for questions, discussions, and exchange. Virtual learning groups interact with other participants in a dedicated international learning community, which all fundamentally contribute to the rich learning experience with a sense of community. Prompt feedback from the course instructors and teaching team further supports successful learning processes. Thereby, MOOCs help to stay up-to-date in a playful, targeted, and fast way, despite the rapid developments in information technology.

Accompanied by recent sociotechnical innovations in online learning and communication, user-generated content has become an interesting object for content analysis. User-driven content includes communication behavior, which is not created by the operator of an (online) offering itself [1, 2], but e.g. by non-organized actors, or users, who can be both professionals and non-professionals. User-content-interaction and user-user interaction [3] agglomerate to dynamic communication sequences, which contain a range of fundamental challenges especially concerning reliable and valid coding as well as content analytical interpretations of user-generated content.

We approach these new issues by analyzing the asynchronous course discourses in an exemplary Massive Open Online Course (MOOC) in IT security with 4,203 enrollments (18 valid percent women) and 1,378 posts on 195 topics in the forum by 274 different users (49 women, 213 men, and 12 with unknown gender).

In respect to gender equal participation in society as well as education, it is very important to analyze new media, tools, and educational developments as to their effects on gender.

II RESEARCH QUESTION AND HYPOTHESIS

Based on our interest in social-scientific gender research, specifically in detecting certain patterns that prevent women from fully participating in ICT MOOCs, we attempt to answer the question as to if there are differences in the communication behavior (between questions, answers, and comments posted by men and women) in an open forum of a MOOC in ICT security. To answer this general research question, we will address the following sub-questions:

- Are there unequal circumstances of specific user groups, such as men and women, affecting their participation in the course forum? Do the forum discussions take place in specific time periods, which, for example, might exclude women due to their other obligations?
- Do men and women raise different types of questions in the discussion forum? For example, do men ask leading, rhetorical or counter-questions more often than women? Do women ask personal or discussion questions more frequently than men?

Of particular interest in this context were the following questions:

- Do questions posted by women imply a higher level of uncertainty or lower self-efficacy than questions posted by men? Are men more likely to use direct language in the course forum, whereas do women favor indirect speech patterns?

- Do fellow learners react differently to posts by men and women, and if so in what way(s)?

III RELATED WORK

Socialization and social constructionist arguments can help to explain gender differences in language use [4]. According to the socialization approach, while girls and women tend to be more indirect and subtle, men express statements in a more challenging way and respond especially by way of side remarks. Furthermore, men are “more likely than women to dominate social interaction through higher rates of self-assertive speech” [4]]. Following the social constructionist perspective, context seems to influence the communication behavior of men and women, too. For example, task-oriented settings foster the use of assertive language [4]].

To analyze interactions in digitally supported teaching and learning, many studies use Salmon’s model of interactivity and effective online education. This model contains the following five steps [5]:

1. Access and motivation: The anonymity of discussions via a keyboard may reduce prejudices, such as those based on gender. “If technology challenges roles, then it is because social changes have allowed those roles to be challenged. Technology is, in fact, often employed to reinforce traditional roles” [6].
2. Online socialization: formal and informal social interactions [7], the tone and interventions engender good behavior [6]
3. Information exchange: self-explanatory online sharing of information [6]
4. Knowledge construction and development: active learning through widening personal viewpoints and appreciating differing perspectives, critical thinking via relevant exercises, self-evaluation, discussion of subject area knowledge, participants’ experiences and opinions [6]
5. Development of critical thinkers: constructivist approach of using the information provided for personal interpretation based on past experiences and knowledge [7]

Based on this theoretical model, researchers found out that students and instructors rate the following aspects of online learning as important: interactions in MOOCs’ online discussions including two-way communications along with feeling “welcomed,” enjoying interaction with coursemates, increasing the personal learning network as well as creating a sense of community. The satisfaction of many learners with the level of interaction in MOOCs leaves something to be desired. This is due to, for example, the unstructured discussion forums, which make it difficult to find students with similar questions, a lack of opportunity to lead online discussions and few interactions with the teacher him- or herself. Compared to the learners, most instructors are satisfied with the level of interactivity in their MOOCs. They feel very “close” to their online students through regular discussion posts even though some instructors report a lack of time to have “substantive” interaction in MOOCs. Interestingly, students and instructors rate the following items of interaction as less important: regular monitoring and

the provision of instructor-directed discussions, posts by the instructor to encourage communication, the provision of student incentives (i.e. points) to put the necessary time and effort into online discussions, and the prevention of discussions being dominated by a few students. Additionally, many instructors neglect the importance of giving advice for developing online skills, offering student access to the permanent records of discussion, and summarizing online discussions at the conclusion of each discussion thread [7] (which consists of the initial question or statement, replies, and comments on the conversation).

Further substantial research focuses on gender identity on the Internet and questions whether and how gender relations are inherent in web-based communication or whether the Internet opens up opportunities to escape from gender ascriptions [8]. Media research concerning gender-specific usage patterns [9] and communication styles, aspects of digital doing gender [10], and gendered technologies [8] applies theoretical models of gender studies such as gender-related life contexts and socialization. Relevant research studies found increased gendered self-presentation in the new media [11]. Applied media research refers to social-cultural specifics of domestic duties and everyday life, including barely flexible and customizable time budgets, which prove to be a dimension of inequality by gender [8].

Female students appreciate timely flexibility as one important advantage of digitally-supported learning, they also prefer the home environment for learning, and place greater importance on individual anonymity than their male counterparts. Nevertheless, they learn less on the computer per week and perceive disadvantages of digital learning more often than males. To a greater degree than male participants, female learners request more support with information searches, favor written exams with free text problems, prefer longer time spans in between the exams, as well as more personal feedback and motivational support from their advisor during the learning process; they also prefer grades for individual examinations. For female students, overall knowledge about the amount of learning content and their current learning status is more important than for males. On the average, female students learn and comprehend more visually-based than males and find it more important to learn fundamental principles, work methods and clear issues, and to master practical application as well as to solve a concrete problem than male students. Furthermore, for female students tight communication in small groups is decisive [12].

From former analysis, we know that learners who participate actively in a MOOC discussion forum complete the course more successfully than passive learners, whereas women are much more cautious in asking questions in science courses [13]. Men, as well as people in higher positions and with more technical knowhow, are more often opinion leaders in forum discussions than are women in general, people in lower positions, or those with little knowledge of the subject [14]. These opinion leaders use online forums more often. While they are particularly involved in posting comments and answers, they initiate fewer new discussion threads than other participants. Their main motivation is to help others [15].

Beside these findings, there seems to have been a lack of particularly qualitative empirical research on behavioral patterns of different groups of learners in online discussions

up to now. We are trying to approach this gap with our case study on the communication behavior of men and women in MOOC discussion forums.

IV DATA AND RESEARCH METHODOLOGY

Based on the theoretical reflections and related research mentioned above, we analyze the user-driven contributions to the asynchronous forum discussions in one exemplary MOOC about IT security, which was held in German by a male facilitator. Seventeen percent of the 4,203 enrollments were female, which represents our classic clientele in programming MOOCs very well.

We processed and coded the collection of user-generated content, filtered out unnecessary communication (such as FAQs) and set up a mixed method approach [16] to analyze this rich data on different levels of analysis (complete discussion courses, related interactions, user level, and single posts). We include the following variables in the upstream quantitative part of our discussion forum analysis with IBM

SPSS Statistics 25 [17]: usage frequency and intensity of use, number and length of topics, posts, and answers per user [4]], number of replies per thread, sentiments, user ratings such as aggregated scope (view counts) as well as counters (votes). These metrics helped to identify opinion leaders and other usage patterns. Our focus is on similarities and differences concerning gender and other socio-demographic characteristics, such as age.

In the main qualitative part of our standardized discourse analysis [18], we study deductively the question techniques of men and women in course discussion forums and add an inductively created code system (i.e., to categorize discussion topics). In this context, we deal with the contextualization of the forum conversations, communication processes concerning openness, relevance, reactions, signals, and feedback, as well as the effects on other participants and their further involvement.

A. Use Case and Description of Data

“Lively discussion forums and virtual learning groups

TABLE I. EXTRACT FROM OUR CODE SYSTEM

Variable	Codes	Definition	Example
1	Semantic codes	Topics	
1 1	Organizational	Course structure, deadlines, etc.	“Is it permitted to use parts of the course material in educational institutions without infringing on any rights of use or copyrights?”
1 2	Technical	Technical problems	“Maybe it's just a problem with my devices and not a general one, in which case I ask you to ignore this post, but I am not able to download the course content. Could you help me with this problem?”
1 3	Assessment-related	Assessment-related topics	“Has this task been taken out of the evaluation? In my case it is no longer shown on the progress page and the points have also been removed.”
1 4	Understanding-based	Understanding the content of a case, contrary to memorization	“What is the exact technical process of encryption and e-mail transmission with several recipients?”
1 5	Application-oriented	Consolidate theoretical findings in practice, application of what was learned, learning by doing	“I have a question: how it is possible to send encrypted e-mails simultaneously on the computer and on the smartphone, if the company uses a smart card for e-mail encryption?”
...
2	Question codes	Questioning technique/typology	
2 1	Open	Questions starting with “(for) what, who, where(by), why, when, how”, etc.	“Hello, the last step of the test failed. Where can I find information about the reason for the failure?”
2 2	Closed	Questions, which only allow the answers “yes” or “no”	“Dear fellow forumites, does anyone know if there is anything similar to the tools presented in the mobile sector?”
...
2 28	Negated	Questions, which contain a negative	“Wouldn't it be enough after the TLS handshake to only accept TLS-enabled servers for further transmissions, especially since SMTPS would fail otherwise?”
2 29	Hypothetical	Questions based on assumptions	“In the examples, we always assume a 1:1 relationship for encryption. How is it then with several recipients? What is the exact technical procedure of the encryption and e-mail sending? Assuming I have four recipients - is the e-mail sent 4 times?”
3	Reaction codes	Reactions of the learning community	
...
3 8	Explanation	Detailed clarification	“The actual message content is encrypted with a symmetric key. This key is generated once and is only used for this message. The symmetric key is then encrypted with the public keys of the recipient. The finished encrypted message thus consists in principle of: 1. unencrypted header data, 2. symmetrically encrypted message content, 3. asymmetrically encrypted session key (for each recipient and for the sender one each).”
3 9	Astonishment	Reactions, which express wondering	“I don't quite understand the point of the question. A REGEXP can only prove syntactical correctness. With a dig on the host part, you get ahead and it's easier. Then all you need is an expression that detects the presence of @ and the absence of illegal characters.”
...

support exchange and collaborative learning” [19] in each of our MOOCs. “Depending on prior knowledge, per course week three to six hours should generally be planned for the serious study of course material. This time is devoted to the learning videos, reading materials, and checking one’s progress with the provided self-tests, as well as active participation in the discussion forums” [19]. Via the forum, participants can asynchronously engage in discussion online at any time. The discussion is available on the course page and open for all learners. Most discussions take place in the open forum. They differ from those in learning groups as they involve a larger number of discussion participants and consequently produce more lively and varied conversations. Therefore, we excluded the learning groups from our analyses.

Up to now, our discussion forum has been rather unstructured. Nevertheless, users can vote for every initial forum post, rate the replies up or down, and thereby help fellow learners to evaluate forum posts. Moreover, learners can filter by tags, for example course sections or key words, sort by most recent activity, latest questions, or best voted first, search for their own free text, and follow chosen discussion threads. During the runtime of a course, learners can start a new topic, reply, and comment on other posts. After a certain time upon conclusion of a course, the forum can be closed. Indeed, learners can continue to read all the posts, but are no longer able to start a new topic, nor add comments or replies. This is necessary because the teaching team will no longer maintain new posts.

At this early stage of our research on forum discussions in MOOCs, we focus our analysis on the initial question(s) of a discussion thread and their first or dominate replies as well as comments.

B. Our Code System

In qualitative and mixed methods approaches, coding allows categorizing “texts for analysis in order to develop new theories or test existing ones” [20]. Concentrated on our research interests, our code system possesses the heuristic function of a “container” [21] and consists of three main parts (see Table I). The top-level structural units comprise semantic codes, which describe the forum threads’ type with respect to the content of a question (1), question codes that define the questioning technique (2), and reaction codes to classify the answers of the learning community (3).

C. Research Methods

As our first lexical search did not convey the expected meaningful insights, and we intended to focus on context information to structure knowledge as well as define relationships in MOOC discussion forums, we scanned the initial questions of the discussion threads and the most important answers as well as comments individually. Our numeric coding of the qualitative data extracted from the discussion forum allowed us to conduct first descriptive and inferential statistics.

V RESULTS AND FINDINGS

In the following paragraphs, we will present our key quantitative as well as qualitative findings about communication behavior in an exemplary MOOC forum

focusing on similarities and differences of the participants by gender.

A. Quantitative Outcomes of Participants’ Behavior in a MOOC Discussion Forum

In our exemplary MOOC, 46 percent of our male learners and significantly fewer women (36 percent) completed the course. Forty-nine women, which are 16 percent of all female enrollments, authored 11 percent of 1,378 posts and initiated 39 of all 195 topics in the discussion forum of this course (20 percent). The majority of those female learners (34 women) reveal their gender via their display name or signature of their post(s). At least one woman contributed in more than one third of the threads (70 of 195). Nevertheless, in only 13 of them (about 20 percent of the discussion threads with female participation) did more than one woman participate actively. Women began 23 percent of 111 discussion threads, which started with an initial question.

Furthermore, we identified 23 of the most active learners in the forum discussions (8.4% of all active discussion participants and 0.5% of all course enrollments). Each of them posted between ten and 77 questions, answers, or comments. Twenty of these most active users were men and only three were women.

B. Time-Based Analysis

Most female as well as male learners made discussion contributions in the afternoon from 2 to 6 p.m. CET (most frequently around 3:00 p.m.), which made up almost one third of all forum posts. The second most common time of the day for discussions in MOOCs was in the morning between 6 and 10 a.m. CET with a peak at 10:00 a.m. (almost one quarter of all contributions to the discussion, independent of gender) followed during the evening (6:00 p.m. to midnight CET, with 20 percent of all posts). More than 15 percent of the discussion questions, answers and comments came around midday (12 noon to 2:00 p.m. CET). Less than ten percent of participants posted between 10:00 a.m. and 12 noon and around three percent between midnight and 6:00 a.m. CET. Even though women were overrepresented participating at night (see Fig. 1), the correlation is statistically not significant.

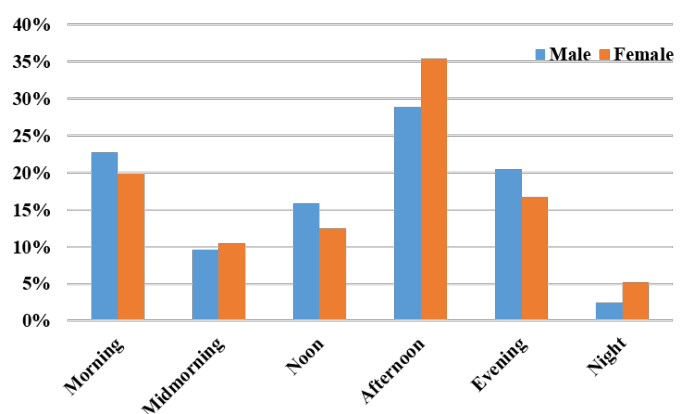


Fig. 1. Forum posts by time of day and gender (N=823).

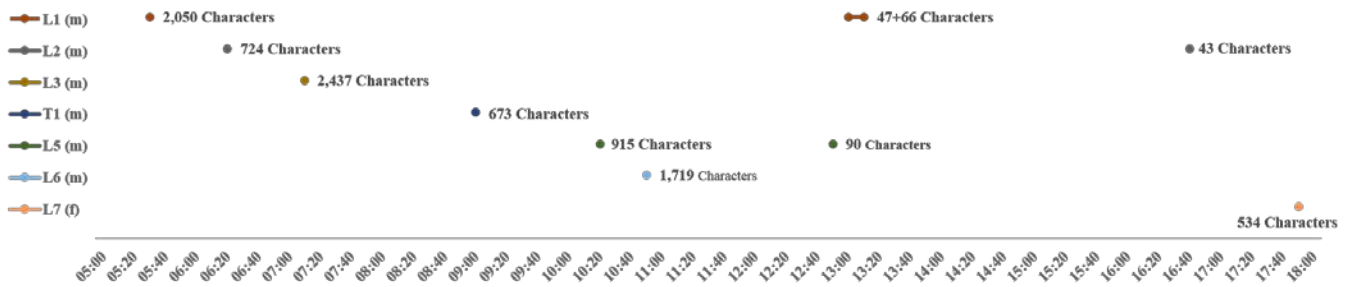


Fig. 2 Course of discussion – a typical example by time of day, gender and characters (with spaces).

C. Course of Discussion

On average, a discussion runs for two days with up to 32 participants. In the first part of a new discussion thread, typically a group of about four learners (L) participated and half of them, especially the initiator of the discussion, contributed more than once during the course of the discussion. We identified different types of behavior patterns in discussants. Some learners mostly participate in the early stage of a discussion thread, whereas others preferred to get in late. Women very often start or close a discussion thread. The teaching team (T) tried to stay in the background (see Fig. 2) and only stepped in to a misdirected discussion, to inspire participants to take part, or to express appreciation for learners' contributions.

The length of contributions in our MOOC discussion forum ranged from three to more than 4,200 characters (with spaces). On average, discussion posts encompass about 400 characters. Interestingly, men write significantly longer posts than women (453 vs. 302 characters), which corresponds to Leaper's findings about gender variations in adult language use [4]. In many cases, the initial post and one or two answers or comments in a discussion thread are comparatively long, while other posts are much shorter. Assuming that online learners type approximately 200 characters per minute, which is equivalent to a slow to medium typing rate, they invest about two minutes on the average to post a question, answer or comment in the discussion forum.

D. Semantic Annotation

Due to the workshop format of the analyzed MOOC with practical exercises on the subject of sending encrypted and digitally signed e-mails, 54 percent of the initial questions of a discussion thread relate to application-oriented subjects (see Table I). In line with cognitive-constructivist learning, 21 percent of the discussion threads covered comprehension questions. Organizational questions dominated less than ten percent of the discussion threads, and feedback was shared (9 and 7%). Less than five percent of the initial questions in a discussion thread dealt with assessment issues (4%), technical problems (3%), or personal validation concerning

the understanding of subject matter (2%), each posted by men. One discussion thread was a constituent part of a task in the MOOC and one initial question posted by a man implies deliberate provocation or stumble. Women were marginally underrepresented in asking practical questions. On the other hand, they were slightly overrepresented in asking initial organizational questions. Nevertheless, the topics are statistically independent of gender.

E. Linguistically-Related Outcomes

We classified the type of initial questions in the discussion threads into 14 different categories. Our learners used simple open and closed questions most frequently (32% and 23%). By formulating closed questions, many learners, especially women (and first time attendees with little experience in formulating forum posts), seem to positively suggest their engagement in the learning material to investigate finding the answer by themselves (statistically not significant). Instead, those questions may appear suggestively, limiting or even hindering. As specific types of closed questions, 20 percent posted alternative or multiple-choice questions, which limit the solution possibilities to two (or rarely more real) options, and thus help to avoid dismissal. Six percent, especially inexperienced online learners (statistically not significant), formulated complex negative questions, which functioned like rhetorical questions, four percent posted contact requests, and three percent asked polite indirect questions to open a new discussion thread by writing in the subjunctive form. Less than two percent used control questions to make sure of their learnings (mostly unneeded) or hypothetical questions to attract the attention as well as engage the learning community, and to open up new perspectives as well as solution approaches [22]. The few rhetorical questions posted seemed to arouse curiosity, signal confidence, but not to expect an answer with consequences, and therefore reduced the activation of the other discussion participants [23]. In line with our forum rules [19], pushy or rude questions intended to provoke another participant, such as certain types of counter-questions or boomerang questions, were rare.

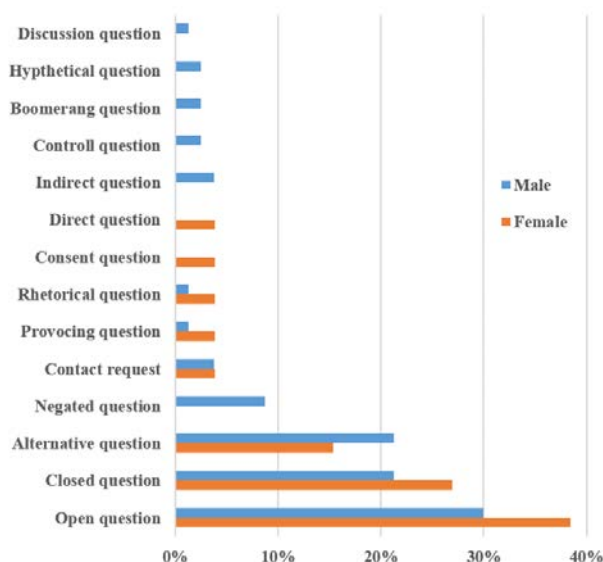


Fig. 3 Questioning typology by gender (N=106).

Women seem to prefer closed questions, which ameliorate the dissemination of knowledge. Furthermore, they were overrepresented in asking rhetorical, and provoking questions, and contact requests, as well as direct questions and consent questions to obtain approval [23]. On the other hand, female learners were underrepresented in posting alternative and discussion questions to reach the learning objectives as well as posting neither control, hypothetical, negative, nor boomerang questions (see Fig. 3). In any case, the small relationship between questioning technique and gender is statistically not significant.

Fellow learners (or the teaching team) answered almost every initial question of a thread (95 percent), especially questions asked by women. In discussion threads opened by men, participants significantly addressed each other directly, which seems to oppose Hirschman's observation that "females use a greater number of 'personal' references" than males [24]. Apart from this, statistically independent of the question technique and gender of the asking person the large majority of answers (85%) were of high quality, underpinned by well-researched links and lengthy explanations (with a small exception for rhetorical and closed questions). Thus, most answers (83%) were useful in approaching the problem, with minor validity for provoking and indirect questions.

Forty-one percent of the answers, especially to low-threshold and activating closed questions, were simple and posted within two to three hours (less than two hours concerning men's questions and more than three hours on women's questions—not a statistically significant difference). Less than 20 percent of the initial questions (especially open questions) were answered by giving a hint or detailed explanation. Regarding the evaluation of the initial question, about five percent repeated closed questions in their own words to agree and emphasize their need for assistance or they expressed their appreciation for the topic in an open, alternative or rhetorical question.

On the other hand, the same number of co-learners made a request or even posed a counter-question, especially when answered indirect questions, or they directly contradicted an open as well as a negated question. One open question

provoked an excuse. The learning community supplemented an indirect, and a boomerang question. One negated question induced a personal evaluation and one alternative question caused amazement. Women received more often both detailed explanations and a simple reply or appreciation and less often a hint. Contrary to the expectation, no one repeated initial questions asked by women, phrasing them with other words or by adding something. Furthermore, nobody responded with a query, approval, opposition, excuse, or personal valuation to women's initial questions and none of their questions caused amazement. The differences by gender are, however, not at a statistically significant level.

F. Qualitative Insights

A further qualitative area of our content analysis revealed that women's discussion posts imply more often uncertainty or lower self-efficacy than men's contributions. For example, women use the word "uncertain/unsure" more often in an "I" statement than men, a finding that corresponds to prior research indicating that self-effacing speech is stereotypical for women [25]. In contrast, men generalize the context in which they use a word, for example: "Large uncertainty predominates this issue," or "so that one can feel certain".

VI DISCUSSION, IMPLICATIONS AND CONCLUSION

As it turned out, the developed code system is a valid framework for a mixed methods analysis of MOOC forum data. On this basis, we have shown that MOOCs are an attractive learning opportunity for underrepresented groups, such as women in IT, and that women and men use discussion forums very successfully to facilitate and enrich their online learning experience. According to our case study, we found the following similarities and differences of learning behavior in MOOCs by gender. Women are well represented among the contributors in the MOOC discussion forum (16% compared to 18 validated percent of female enrollments in total). Even though women initiated comparatively many discussion topics (20%), posts by female learners were underrepresented in the discussion forum with 11%. Especially the most frequent discussants were almost entirely men and male learners write significantly longer posts than women do statistically-speaking. Posts have proven to be most prevalent from 2 to 6 p.m. in the afternoon, when most forum discussions took place and it is at this time that many women have other obligations. Likewise, stereotypically, women were slightly underrepresented in asking practical questions on the one hand and were marginally overrepresented in asking organizational questions.

Even though, many women preferred closed questions, which ameliorate the dissemination of knowledge, and consent questions, as well as rhetorical questions over discussion and hypothetical questions and contact requests, the small relationship between questioning technique and gender turned out not to be statistically significant in our case study. Nevertheless, the question type significantly influenced the way of answering. Direct and open questions particularly encourage fellow learners to answer in a convenient way and in a short time (less than two hours concerning men's questions and more than three hours on women's questions). Based on this finding, we recommend that participants, as well as the teaching team, formulate direct and open questions as well as gender neutral posts to

foster effective forum discussions. Thus we can encourage minority groups, such as women in STEM, to participate successfully in MOOC discussions and reach the learning objectives.

The qualitative part of our content analysis revealed that women's discussion posts imply more often uncertainty or lower self-efficacy than men's contributions [see also 4]. This finding runs in the same direction as prior research indicating that self-deprecation is more common of women than men [26]. Women received from the other forum members more often detailed explanations, a simple reply or appreciation and less often a vague answer. Contrary to expectations, no one repeated initial questions asked by women phrasing the questions using other words. Furthermore, nobody made a query, an approval, an objection, an excuse, or a personal valuation to women's initial questions and none of their questions caused amazement, although the differences by gender are not at a statistically significant level in our case study.

The paper at hand encompasses the findings of our early analysis of forum discussions in a German programming MOOC and contains course-specific findings. Therefore, the interesting insights can only be transferred to other MOOCs with caution for the time being. Furthermore, user behavior can be mixed or influenced by platform mechanisms (e.g., via algorithmic and organizational curation of contributions to the forum discussions) [27]. At this time, we are still at the beginning of determining opportunities to correct biases, which can arise at different points during the collection of user-generated content and may influence the results.

VII FUTURE WORK

On the basis of the limitations of our case study defined above, we will further develop the code system in our future work (e.g., by applying multiple tags of questioning techniques to the same text snippet [28]), and prove its validity to analyze discussion forums of different kinds of MOOCs. We will extend our analysis focusing on additional socio-demographic characteristics of the responding learning community (e.g. highest degree, professional life, career status, and position) to recognize potential additional influencing variables on the learning behavior of women and men in MOOC discussion forums.

In regard to potential group composition effects [22], such as gender specific communication behavior occurring particularly in mixed-gender interactions when gender is most salient [29], we will compare discussion threads with half male and half female participation with men-only, or women-only discussions, as well as conversations with less than 50 percent, but at least two women, and with only one woman. Our interest lies on the effect of these participant groups on the atmosphere in MOOC discussion forums. Depending on the results of these evaluations, we will experiment with slight modifications to our discussion forum with respect to smaller learning groups (e.g. with regard to single-sex educational settings) to promote contributions of underrepresented groups such as women in STEM and evaluate those adjustments in different MOOCs as well as accompanying surveys. Concurrently, we will improve our computer-assisted learning analytics tools, such as our implemented sentiment analysis, meaningfully visualize the extracted information (e.g. in knowledge graphs, to gain additional and deeper insights). In the context of our applied

educational research, we will use all the findings to improve gender-neutral teaching and learning in future MOOCs.

It could also be interesting to investigate the discussions specifically in relation to the dramaturgy of the course (e.g. end of a course week and exam period) as well as in respect to particular posts (introduction, main part, conclusion, summary, outlook, acknowledgements, goodbye), and their impacts on the communication behavior of men and women. Another research question may focus on the long-term effects of different progresses of discussion on learners' further participation in MOOC forums by gender.

REFERENCES

- [1] C. Figallo and N. Rhine, "Tapping the grapevine. User-generated content," in *E-Content: the magazine of electronic research & resources*, 24(3), 2001, pp. 38–43.
- [2] W. Schweiger and O. Quiring, "User-Generated Content auf massenmedialen Websites. Eine Spielart der Interaktivität oder etwas völlig anderes?," in M. Friedrichsen, W. Mühl-Benninghaus, and W. Schweiger (Eds.), "Neue Technik, neue Medien, neue Gesellschaft? Ökonomische Herausforderungen der Onlinekommunikation," Munich: Fischer, 2007, pp. 97–120.
- [3] T. B. Ksiazek, L. Peer, and K. Lessard, "User engagement with online news: Conceptualizing interactivity and exploring the relationship between online news videos and user comments," in *New Media & Society*, 18, 2016, pp. 502–520.
- [4] C. Leaper and M. M. Ayres, "A meta-analytic review of gender variations in adults' language use: talkativeness, affiliative speech, and assertive speech," in *Personality and Social Psychology Review*, 11(4), 2007, pp. 328–363.
- [5] G. Salmon, "E-moderating: the key to teaching and learning online," 2005.
- [6] N. A. J. Moore, "Review of e-moderating—The key to teaching and learning online," in *Language Learning & Technology*, vol. 6, no. 3, 2002, pp. 21–24.
- [7] H. Khalil and M. Ebner, "How satisfied are you with your MOOC?—A research study about interaction in huge online courses," in *Journalism and Mass Communication*, vol. 5, no. 12, 2015, pp. 629–639.
- [8] M. Lünenborg and T. Maier, "Gender Media Studies, Eine Einführung," Constance and Munich: UVK Verlagsgesellschaft, 2013.
- [9] B. Kampmann, B. Keller, M. Knippelmeyer, and F. Wagner (Eds.), "Die Frauen und das Netz. Angebote und Nutzung aus Genderperspektive," Wiesbaden: Springer Gabler, 2013.
- [10] C. Funken, "Digital doing gender," in: S. Munker and A. Roesler (Eds.), "Praxis Internet. Kulturtechniken der vernetzten Welt," 2002, pp. 158–181.
- [11] D. Nübling and H. Kothoff, "Genderlinguistik: eine Einführung in Sprache, Gespräch und Geschlecht," 2018.
- [12] J. Arrenberg and S. Kowalski, "Lernen Frauen und Männer unterschiedlich? Eine Studie über das Lernverhalten von Studierenden," Cologne University of Applied Sciences, 2007.
- [13] J. Qiu, J. Tang, T. X. Liu, J. Gong, C. Zhang, Q. Zhang, and Y. Xue, "Modeling and predicting learning behavior in MOOCs," 2016. [9th ACM International Conference on Web Search and Data Mining USA, 2016].
- [14] Burson-Marsteller, "E-fluentials: The Power of Online Influencers," 2010.
- [15] T. Smith, J. R. Coyle, E. Lightfoot, and A. Scott, "Reconsidering Models of Influence: The Relationship between Consumer Social Networks and Word-of-Mouth Effectiveness," in *Journal of Advertising Research* December, 2007, pp. 387–397.
- [16] U. Kuckartz, "Mixed Methods. Methodologie, Forschungsdesigns und Analyseverfahren," Wiesbaden: Springer Fachmedien, 2014.
- [17] IBM, "IBM SPSS Statistics," 2019. Retrieved 24 October, 2019 on <https://www.ibm.com/products/spss-statistics>.
- [18] P. Rösser, "Inhaltsanalyse. 3. Auflage," Constance and Munich: UVK Verlagsgesellschaft mbH, 2017.

- [19] openHPI, "Forum Guidelines," 2018. Retrieved 3 March, 2020 on https://open.hpi.de/pages/forum_rules.
- [20] T. Tietz, "Linked Data Supported Content Analysis for Sociology," in M. Acosta et al. (Eds.), "SEMANTiCS 2019, LNCS 11702, pp. 34–49.
- [21] T. Richards and L. Richards, "Using hierarchical categories in qualitative data analysis", in U. Kelle (Ed.), "Computer-aided qualitative data analysis," London: Sage, 1995, pp.80–95.
- [22] D. N. Maltz and R. A. Borker, "A cultural approach to male–female miscommunication," in J. J. Gumperz "Language and social identity. Studies in interactional sociolinguistics 2," Cambridge University Press 1982.
- [23] D. Zillmann, "Psychologie der rhetorischen Frage," in *Semiosis Internationale Zeitschrift für Semiotik und Ästhetik*, 17. Jahrgang, Heft 1–4, 1992, pp. 235–243.
- [24] L. Hirschman, "Female-male differences in conversational interaction," in *Language in Society* 23, 1994, pp. 427–442.
- [25] L. Rudman, "Self-Promotion as a Risk Factor for Women: The Costs and Benefits of Counterstereotypical Impression Management," in *Journal of Personality and Social Psychology* March 1998, Vol. 74, No. 3, 629–645.
- [26] J. Wolfe, "Communication Styles in Engineering and Other Male-Dominated Fields," in B. Bogue and E. Cady (Eds.) "Applying Research to Practice (ARP) Resources," 2012.
- [27] N. Couldry and A. Hepp, "The mediated construction of reality," Cambridge, MA: Polity, 2017.
- [28] T. Staubitz and Ch. Meinel, "Peer Assessment in MOOCs–User Feedback," 2020. [IEEE EDUCON Global Engineering Education Conference Portugal], unpublished.
- [29] L. Carli, "Gender Differences in Interaction Sytle and Influence," in *Journal of Personality and Social Psychology*, Vol. 56, No. 4, 1989, pp. 565–576.