Evaluation of Student Feedback Within a MOOC Using Sentiment Analysis and Target Groups

Karsten Ø. Lundqvist[1], Tharindu Liyanagunawardena[2], and Louise Starkey[3]
[1]School of Engineering and Computer Science, Victoria University of Wellington, [2]Department of Online Education, University College of Estate Management, [3]School of Education, Victoria University of Wellington

Abstract

Many course designers trying to evaluate the experience of participants in a MOOC will find it difficult to track and analyse the online actions and interactions of students because there may be thousands of learners enrolled in courses that sometimes last only a few weeks. This study explores the use of automated sentiment analysis in assessing student experience in a beginner computer programming MOOC. A dataset of more than 25,000 online posts made by participants during the course was analysed and compared to student feedback. The results were further analysed by grouping participants according to their prior knowledge of the subject: *beginner*, *experienced*, and *unknown*. In this study, the average sentiment expressed through online posts reflected the feedback statements. Beginners, the target group for the MOOC, were more positive about the course than experienced participants, largely due to the extra assistance they received. Many experienced participants had expected to learn about topics that were beyond the scope of the MOOC. The results suggest that MOOC designers should consider using sentiment analysis to evaluate student feedback and inform MOOC design.

Keywords: MOOC, teaching programming, sentiment analysis, target group, feedback, learner analytics

Introduction

Since 2011, technological development has enabled the growth of online learning with free courses known as Massive Open Online Courses (MOOCs) attracting thousands of learners. The success of MOOCs depends on the active involvement of large numbers of learners who, through dynamic engagement, self-organise into learning communities where they share skills, objectives, knowledge, and interests, by commenting within the learning system and using other social networking tools (McAuley, Stewart, Siemens, & Cormier, 2010).

A challenge when running a MOOC is gaining an accurate understanding of learner experience because the number of participants makes it impossible to follow all posts and interactions. Participant comments and actions can provide an impression of the sentiments and concerns of learners within a course. For example, some disgruntled participants may leave a course, trolling by engaging in fruitless argumentation (Donath, 1999), while other participants, struggling with course material, may vent frustration. Analysis of individual learner experience is an important aspect of course evaluation but difficult to undertake when there are thousands of participants. Without analytical tools to understand overall sentiments and how they may vary across different groups of learners, it is easy to disproportionately focus on negative posts.

This paper presents a method for evaluating learner group experiences within a MOOC using automated sentiment analysis of the posts and comments made by participants within the course and compares this to statements given in a dedicated feedback section within the MOOC. The purpose is to provide insight into learner groups' experiences during the course that is not limited to survey responses at the conclusion of the course.

Studies that investigate individual MOOCs from a learner's perspective have drawn data from learner experience surveys, participant demographics, and learner progression through courses, such as the number of videos viewed or tests taken (e.g., Kop, Fournier, & Mak, 2011), participant size and completion rate (e.g., Adamopoulos, 2013), or from the behaviour, motivation, and communication patterns of online students (e.g., Swinnerton, Hotchkiss, & Morris, 2017). These metrics mirror attendance and completion data used to evaluate formal higher education. However, retention alone does not reflect the quality of a MOOC (Downes, 2015). Applying the same measures used in formal learning is problematic as MOOCs are free and usually stand-alone courses and, as such, there are limited consequences for learners who choose to not engage in aspects of a course. Those enrolled may stop participating due to other time commitments, course design, or the content being too challenging or too easy. Instructional designers should consider why groups of learners disengage from a course in order to better inform pedagogical decisions (Liyanagunawardena, Parslow, & Williams, 2017).

While MOOCs have high levels of organization and a range of course material presented, they can be poorly designed (Margaryan, Bianco, & Littlejohn, 2015). Good instructional design follows constructivist principles; it promotes learning through problem-solving, encourages the use of prior knowledge to form understanding, and includes reflection on learning through discussion and critique (Merrill, 2002). However, the size of MOOCs makes it difficult to both provide personalized learning relevant to the diverse backgrounds and prior experiences of the participants and offer individualized feedback and support. The digital nature of a MOOC can provide aggregated data from target groups, allowing the opportunity to

investigate whether the design meets the needs of the intended learners. An aspect of constructivist theory is that adult online learners need to be self-directed and motivated to enable learning from their course experience (Huang, 2002). Negative sentiments about a course can impact motivation and ultimately the learning experiences of participants in a MOOC.

MOOC datasets are complex to analyse due to the variety of types of data available. MOOCs produce large amounts of interaction data between learners including discussion statements, likes, and follows, and individual interactions with the system, such as timestamps of actions, videos watched, test results, and logins, which provide evidence of participant engagement and experience. MOOC learners can be encouraged to engage in social learning; therefore, textual data in the form of comments and forum posts can be a valuable source for understanding participants' sentiments within a MOOC.

Sentiment analysis was first used as a term in 2003 and is therefore a relatively new area of study within natural language processing (Liu, 2012). With the growing popularity of social networking, sentiment analysis techniques have been used with social networks, especially text-rich sources such as blogs and Twitter (e.g., Hong & Skiena, 2010; Miller, Sathi, Wiesenthal, Leskovec, & Potts, 2011; Tumasjan, Sprenger, Sandner, & Welpe, 2010). Advances have enabled data-mining techniques and artificial analysis to be applied to MOOCs (e.g., Crossley et al., 2015; Wen, Yang, & Rose, 2014) and, more recently, analysis has begun to be used to identify sentiment from within MOOCs (Moreno-Marcos et al, 2018; Pérez, Jurado, & Villen, 2019).

The purpose of this study was to develop a nuanced understanding of the sentiments of MOOC participants. From a constructivist perspective, how participants feel about a course will influence their experiences and engagement. The study was guided by two research questions:

- How does feedback about a course align with the general sentiment expressed in online posts in a course?

- How does sentiment vary between different groups of MOOC learners?

Context

This study examines "Begin Programming: Build Your First Mobile Game" a MOOC designed to introduce computer programming to beginners. According to Perkins, Hancock, Hobbs, Martin, and Simmons (1986), there are two different types of novice programming learners: *stoppers* and *movers*. Stoppers stop when they encounter a problem, whereas movers experiment with the code and use feedback from the system combined with what they know to try to solve the problem. The course provided small steps to support stoppers and a large code base as a safe sandpit to experience and explore the taught concepts using mover tactics. The MOOC participants developed a mobile game from the provided code base, learning from the examples while playing with the code to modify the behaviour of the code. Anecdotally, this appeared to work well to avoid stopper behaviour and provide a constructivist learning experience for aspiring programmers. The approach aligns with the socio-constructivist aspirations behind the FutureLearn platform, where the MOOC was hosted (Sharples, 2013).

The course ran for seven weeks, with one new topic per week (see Table 1). Each week was structured into steps, which contained textual content, video, tests, and/or assignments. All steps had a commenting facility similar to a Facebook wall to allow real-time and in-context discussions. The aim of the first week was to set up the development environment. If learners could not set up the environment, they were unable to participate in the rest of the MOOC. The next three weeks introduced data type, conditional statements, arrays, and looping. The last three weeks introduced algorithms, problem solving, and functions, with the final week having a test and steps for reflection.

Table 1

Weekly Topics

Week	Topic
1	Welcome
	Set up development environment
2	Data types
	Variables
3	Conditional statements
4	Arrays
	Looping
5	Algorithms
	Problem solving
6	Functions
7	Student evaluation
	Student reflections

The course ran eight times (sessions), each time with improvements, but with three main iterations. The initial iteration was designed for the first standard Android programming environment integrated within the Eclipse development environment. During sessions one to four, the course only had minor updates, because most of the critical feedback from learners in the early sessions was about Eclipse or the lack of functionalities on the FutureLearn platform.

Shortly after the fourth offering of the MOOC, Google changed the development environment to Android Studio and, therefore, all material needed to be updated. This was the second iteration. The third iteration was developed for the final session after the original academic lead had left. At that time, some of the content and support provision were changed by the new academic lead (Table 2).

Table 2

Different Sessions of the MOOC

Session	Start date	Type
1	October 28, 2013	Eclipse
2	February 24, 2014	Eclipse and minor updates
3	October 20, 2014	Eclipse and new minor updates
4	February 16, 2015	Eclipse and new minor updates
5	June 01, 2015	Android Studio and major revision
6	September 12, 2015	Android Studio
7	April 04, 2016	Android Studio
8	October 31, 2016	Android Studio with new teaching team

Available Data

The FutureLearn platform provides comma-separated values (CSV) files with engagement data to individual MOOC developers. These include enrolment, learner activity, statements, and test data. The statement file has each comment made in the course and includes a statement identifier, learner identifier, the step number, time it was made, time it was modified (if it was), number of likes, and, if it is a reply to another comment, the identifier of the "parent" statement.

It was difficult to gauge learner perspectives from within the MOOC and respond with changes to the teaching material due to the size of the cohort and the nature of the online platform. As the course began, the focus was on emerging issues and interventions to support the more than 10,000 learners. The focus and reflection while running the new MOOC became distorted because of negative critique by participants. This potential disconnect between the teacher's experience and the overall feedback given in the course was the first main motivator for investigating sentiment.

Across the sessions, retention rates increased yet engagement levels decreased and interaction through posting and commenting also decreased. To understand why, participants were asked to provide feedback from within the MOOC. From those statements, it appeared that participants with prior experience of learning programming were critical of how the course was structured around a large code base without having introduced smaller examples first. This observation provided a reason to explore learner groups within this study. The patterns among different types of learners were anecdotal, with no scientific validation. Therefore, it seemed pertinent to perform an in-depth learner-group-based analysis to understand if the pedagogical design of the MOOC was working for students new to programming.

Method and Research Design

The information collected from comments and posts in sessions one to four was focused on external issues rather than sentiment about the course. To investigate learner experience, we decided to perform an in-

depth investigation of sessions five to seven, comparing sentiment with participants' programming experience. These later sessions were the most homogeneous, with minor changes in content and teaching staff. Differences between the sessions would have been unlikely to have contributed to the results through bias caused by changes.

Sentiment Analysis

Sentiment analysis was employed to investigate the online comments of participants in the MOOC. The VADER (Valence Aware Dictionary for sEntiment Reasoning) sentiment algorithm was used (Hutto & Gilbert, 2014). VADER was designed using sentiment ratings from more than 90,000 English statements originating from social media. Using VADER does not require any prior training, and it has been found to be more consistent than human investigators on large English-text datasets from online sources. The algorithm was benchmarked against seven other automated sentiment algorithms, and VADER both outperformed the others (Hutto & Gilbert, 2014) and was found to be the most accurate algorithm that did not require training (Gonçalves, Dalip, Costa, Gonçalves, & Benevenuto, 2016). Because there are no training corpora available from MOOCs, VADER was used in this study. VADER produces four sentiment scores: positivity, negativity, neutrality, and a compound score. The compound score is not an average of the other three, but a reflection of the overall sentiment of the provided text. The compound score was used in this study.

Feedback Groups

Starting in the fourth week of the course, participants could provide feedback. There was a free text comment box where, under the title, "The good, the bad, and the ugly," they could respond to, "Please post a comment below with one good thing and one bad thing about the course." Statements made during this step (n=337), including the replies, were analysed, and the individuals (n=264) who made the statements and replies were categorised into types:

- *Positive*: No negative point or an insignificant negative point. For instance, statements about anonymous learner functionalities or service were disregarded because the course developers did not influence these aspects.

- *Structure*: Critiquing the teaching/pedagogical approach, i.e., using a large code base instead of small examples. Could include positive points. If a statement was mainly positive but also critiqued the structure, it would be categorised here.

- *Negative*: Negative about the course. Could include structure critique as well, but no positive points about the course.

- *Irrelevant*: Comments without positive or negative value, such as questions. These learners were subsequently added to the *other* group of participants who did not provide feedback.

- *Other*: Not making a feedback statement. This group (merged with *irrelevant*) was used as a baseline for statistical analysis.

Staff statements were not included in the analysis. Statements and replies made by the same person were merged, so that the evaluators would have the full picture of each individual's feedback. Many comments were both positive and negative, and, of these, most commonly critiqued the pedagogy, teaching, structure, or methods used in the course.

Two researchers, the academic lead and a research assistant, independently categorised the statements. The first categorisation had a Krippendorff alpha of -0.106. This indicates disagreement or systematic difference. Out of the 337 total feedback statements, there were 39 disputes. To rectify this, the evaluators met to discuss and reach consensus on the statements they had assigned to different categories. The relatively high number of inconsistencies mostly originated from grammar and language, including the common abbreviations and slang of textspeak, which was not known to one of the evaluators.

To explore whether the feedback provided in the course reflects sentiment within the online posts, the following hypotheses were then tested:

- $Hypothesis_{A1}$: Participants who were positive in the feedback were more positive overall than participants who gave no feedback or irrelevant feedback.

- $Hypothesis_{A2}$: Participants who were negative in the feedback were more negative overall than participants who gave no feedback or irrelevant feedback.

- $Hypothesis_{A3}$: Participants who criticised the pedagogy in the feedback were more negative overall than participants who gave no feedback or irrelevant feedback.

- $Hypothesis_{A4}$: Participants who were positive in the feedback were more positive overall than participants who gave negative feedback.

- $Hypothesis_{A5}$: Participants who were positive in the feedback were more positive overall than participants who criticised the pedagogy in the feedback.

- $Hypothesis_{A6}$: Participants who were negative in the feedback were more negative overall than participants who criticised the pedagogy in the feedback.

Data Analysis

The VADER score was calculated using Algorithmia.com. This study used the compound score resulting from running the algorithm on each full comment made by participants. This included the programming code, which only rarely has a sentiment. A test was run on ten representative posts containing ordinary text with embedded programming code and it was found that the positive and negative sentiment of the ordinary text was not affected by the embedded code.

Each post made in the MOOC was analysed at an individual learner level. Then, the scores from all posts made by an individual were averaged. The comments were saved separately in a CSV file using the original database files, and then encoded into JSON (JavaScript Object Notation) files as required by the Algorithmia.com service.

An initial calculation was run on all posts from the three Android Studio-based sessions. Using the statistical programming language R to run the Shapiro-Wilk normality test (Shapiro & Wilk, 1965), it was found that the *p*-value was lower than 2.2e-16; thus, using an alpha value of 0.05, the data could not be considered normally distributed. The Wilcoxon rank sum test with continuity correction (using zero method Pratt) was selected to evaluate the statistical significance of the sentiment data because it does not assume normality and incorporates tied values in the ranking procedure (Pratt, 1959).

Target Groups

A selection of participant were manually evaluated to establish the level of prior learning. A selection from other (150 out of 7,562) and positive (150 out of 239) were randomly assigned a floating-point number between zero and one. These participants were then sorted according to this random number and used in this order. In addition, all participants from negative (n = 2) and structure (n = 38) were used. Evaluating all of their statements these participants were then grouped by the two researchers using the following experience categories:

- *Prior*: Indication of prior programming teaching and learning experience.

- *Beginner*: No indication of prior programming learning experience.

- *Unknown*: No indication that supports membership of *prior* or *beginner*.

All posts and feedback made within the course by the selected participants were used in the categorisation. However, most evidence came from the learner introductions, help provided to other participants, or in the feedback statements.

For *other* and *positive*, margin of error calculations were used to find confidence intervals to evaluate significant statistical differences of experience between the various categories of opinion (Calder, 1953). The other two groups had full categorisation of all participants, so no similar evaluation was needed.

Results

Of the 3,531 participants who made at least one post in sessions five to seven, 264 (7.5%) individuals also wrote something in week four's feedback step. The final distribution among the feedback groups was:

- *Positive*: 218

- *Structure*: 28

- *Negative*: 2

- *Irrelevant*: 16 (these participants were added to other, i.e., treated as participants who did not provide feedback.)

After running the VADER algorithm on the posts made by all learners from each group (see Table 3), it was discovered that *positive*, *negative*, and *other* all had relative averages following the hypothesis. However, the statements made by *structure* had a higher score than the *other* group, indicating a more positive attitude in their statements. This was a surprise and therefore, without further analysis, Hypothesis$_{A3}$ was rejected.

Table 3

Average and Standard Deviation of Compound Sentiment Scores by Groupings

Category	Average	SD	n
Positive	0.2897	0.4524	4,746
Negative	0.1100	0.4416	36
Structure	0.2783	0.4487	417
Other	0.2770	0.4232	30,985

The results of the Wilcox Pratt test are presented in Table 4. There is statistical significance to support that the *positive* participants were more positive than the *negative* and *other* participants in their general statements. *Negative* were more negative than *structure* and *other*, with statistical significance. Therefore, Hypothesis$_{A1}$, Hypothesis$_{A2}$, Hypothesis$_{A4}$, and Hypothesis$_{A6}$ were accepted.

Hypothesis testing was carried out to explore the first research question. The results show that the voluntary feedback does reflect the general sentiment within the rest of the posts. However, participants who critiqued the teaching method and underpinning pedagogy were not negative in their other comments in the MOOC.

Table 4

Statistical Tests (Wilcox Pratt) of Hypotheses

Hypothesis	p-value	Effect	Support ($p<0.05$)
Hypothesis$_{A1}$ *Positive > Other*	0.0007	-4.3e-05	Accepted
Hypothesis$_{A2}$ *Negative < Other*	0.0367	0.166	Accepted
Hypothesis$_{A3}$ *Structure < Other*	NA	NA	Rejected
Hypothesis$_{A4}$ *Positive > Negative*	0.0165	-0.200	Accepted
Hypothesis$_{A5}$ *Positive > Structure*	0.5061	NA	Rejected
Hypothesis$_{A6}$ *Negative < Structure*	0.0344	-0.180	Accepted

Target Groups

To explore the second research question, the different target groups were identified and analysed against feedback sentiment. The results of assigning target groups for learners in the feedback groups are presented in Table 5. With a 98% confidence level, those criticising the structure were more likely to have had prior experience and were less likely to have been beginners to programming than the participants giving positive or no feedback (*positive* and *other* groups). They were also more likely to self-identify their skill level. A large proportion of participants who did not provide feedback said very little in the comments, and thus, their experience level is not known. The *positive* group disclosed their experience level less often than the *structure* group. The *negative* group was too small to make any meaningful comparisons with other groups. However, all disclosed prior experience.

Table 5

Target Group Results

	Prior	Beginner	Unknown	Sample	Total
Other	37.3% (56)	25.3% (38)	37.3% (56)	150	7,562
98% conf.	31.3%-43.4%	19.9%-30.7%	31.3%-43.4%		
Positive	48.0% (72)	24.7% (37)	27.3% (41)	150	239
98% conf.	43.9%-52.1%	21.3%-28.0%	23.9%-31.2%		
Structure	62.3% (18)	14.3% (4)	21.43% (6)	28	28
Negative	100% (2)	0%	0%	2	2

Discussion

Through the use of feedback and target groupings, this study evaluated the sentiments of learners and compared participants new to programming with those who had prior experience. The participants who had prior experience criticized the MOOC structure but remained positive overall: 78.3% were positive while only 21.7% were negative or criticised the structure.

People with prior experience appeared to be more likely to disclose their level of expertise. For example, the study showed that in the feedback groupings, the lower the number of experienced learners, the higher the number of participants with an unknown background. This is a very human result that could be expected given that revealing oneself as a beginner creates vulnerability, especially if disclosures are made when providing feedback. Many participants used their prior experience as an argument to validate their feedback. For example, a *structure* participant with prior experience said: "I think that I would prefer a bit more of a 'hello world' approach, as I did with BBC basic [sic] years ago." This participant criticises the choice to use a large code base instead of use the traditional approach to start with no code. Participants without prior knowledge could not critique course design in the same way as they would have had no comparative experience.

 No matter what feedback group they belonged to, participants with prior experience were helpful in their posts. Many contributed 20 or more posts meant to help other participants. These helpers supported learners who found activities in the course challenging. For example, the first week was especially demanding as it required participants to set up on their computer a development environment that was needed for subsequent activities. Even with extensive online guidance in the MOOC, the activity was error prone, and, for beginners, this was problematic and frustrating. While those with experience tried to aid the beginners, they also acknowledged the frustrations and offered solutions: "I think the course should be split in to two: 'Begin Programming' and 'Build Your First Mobile Game'. I think the current compromise

is too abstract for beginners and doesn't have enough progressions…" (*structure* participant with prior experience who had helped many)

All beginners who criticised the structure of the MOOC mentioned the difficulty of working with something that one does not fully understand. For example, a beginner who commented on the structure said: "I guess the disadvantage amongst many great things of this course is that one can barely understand how to develop a game from scratch…[sic]"

This research suggests that the course design may have been effective in shifting beginning participants' ways of learning. Perkins et al. (1986) described the differences between stoppers and movers, and this MOOC was developed with these modalities in mind. Data, especially those gathered from the positive beginner participants, seem to suggest that some stoppers turned to mover learning. None of them mentioned having difficulties working with a larger code base and most instead focused on what they enjoyed about the course. For example, one beginner, in responding to the request for one good and one bad thing about the course, said, "Good: I think the entry level is appropriate and almost everyone should be able to follow it. Bad (no [sic] too good): I would have liked to have more questionaries [sic] during the weeks.", therefore indicating that they enjoyed the tests and would have liked more of them. Further study is needed to investigate and validate the design approach.

The *negative* group was small and consisted mainly of participants with prior experience. Much of their feedback seemed to be founded on the fact they had expectations that were based on their backgrounds. For example, one *negative* participant said:

> To be honest, I'm very disappointed so far. Nothing of the original promise seems to be delivered, and instead, the reader gets a slow paced, very-very crude and underquality [sic] 'programming basics' course, which is demoed on the Android platform, via something that remotely resembles an app that was intended to look like a game, but that could actually be just about anything else with no effect on the course itself. I'd gladly mention positives as well, but frankly, there was nothing so far I liked about the course.

Such strong sentiments may generate an emotional response in course designers. However, these should be put in perspective; it was a very small minority who reported such views, and many participants who were in the *negative* group expected an advanced course on game development, which was never the intention of this MOOC.

Impact of Learners with Prior Experience

Promotional information for the course stated that it was meant for beginners, yet it attracted learners with prior programming experience. The impact of having these experienced participants in the course was investigated. Although some of these learners gave negative feedback, many engaged with other students with positive sentiment. Therefore, their contribution to the course was classified as positive, which meant that their general sentiment was not significantly different from learners in the *other* or *positive* categories.

However, it should also be noted that the *negative* learners were so few overall that the impact of their comments should be negligible. A participant with prior experience is disproportionately more likely to give

feedback. This is a problem because they were not the target group for the course and their views therefore are less relevant.

Beginners who provided feedback were almost 10 times more often positive than negative when compared to the participants with prior experience. The participants with prior experience were only five times more often positive than negative. Therefore, the constructivist approach adopted in course design, that is, to engage with beginners, appears to have been successful.

Participant Sentiment Indicated by Feedback and in Posts

There was a connection between the views that participants expressed when asked for feedback from within the course and the sentiments articulated in their posts. Positive and negative feedback were mirrored in the two sets of data. Likewise, the sentiments expressed by both beginners and more experienced learners followed the same pattern regardless of the source of the data.

There is a concern however with the results generated by the *other* group, which had a higher number of beginners and yet expressed more negative sentiments overall. The first observation is that on average these beginner participants were generally positive in their statements; they were much closer to the *other* group than the *negative* group. Their statements expressed the frustration of setting up the development environment. Having a milestone at the beginning of a course that must be completed in order to follow the rest of the course skews the sentiment scores negatively. The course has been improved continuously in this aspect, through videos, text content, and FAQs, but this remains an issue which the design team continues to address.

MOOC Target Groups Versus Other Participants

While the intention of this research was not to predict behaviour, sentiment analysis has been a valuable tool to analyse and better understand learners. Before this study was undertaken, course evaluation was primarily based on anecdotal evidence and speculative and skewed observations from a teacher's standpoint, and the students' perspectives had not been given enough attention. This study has shifted the perspective and helped identify areas of concern.

Two studies have found a significant relationship between sentiment expressed in comments from within a MOOC and the dropout rate of participants (Adamopoulos, 2013; Wen et al., 2014). However, neither study examined the reasons for the sentiment. This paper has expanded these findings, showing there is a connection between the sentiment expressed in direct feedback and posts made throughout the course. Furthermore, participants who are not part of the target group of a course can show tendencies that appear to be contradictory. People with prior experience may provide critical feedback on the structure or content of a course, while their other sentiments, expressed as support for participants without prior learning experience, are positive. The implication is that the open nature of a MOOC can cause dissonance, and that MOOC designers have to judge whether this is a concern.

Recommendations

Any MOOC which offers learners the opportunity to interact by text can use automated sentiment analysis to measure general sentiment within that course. Evaluating these results can be used to investigate how

participants experience learning and their attitudes towards a course. Designers and instructors can confidently make appropriate modifications without having to depend on the results of traditional questionnaire-type evaluations carried out at the conclusion of a course.

Limitations

A limitation in this study is the low response rate, which could bias the sample. Only 264 individuals provided feedback during the fourth week of the course in the free text comment box, which is a reflection of the number of participants commenting in MOOCs in general. In the very first MOOC run on the edX platform, only 3% of all active participants made one or more comments (Breslow et al., 2013). In this course, 38.4% of all active participants made one or more comments, with 2.3% of those giving feedback. If there were no self-selection present, i.e. negative participants were not more likely to provide feedback, then with a confidence level of 95% there would be up to 1.7% of negative participants in the MOOC.

Conclusion

Using sentiment analysis on text data from a MOOC has helped the teaching team make evidence-based observations and conclusions that otherwise might have been overshadowed by anecdotal evidence from teaching experiences. The study found that there is a relationship between general sentiment of posts and the feedback given about the MOOC, and that the general sentiment within this course was positive. A few learners were both positive in their general comments and critical in their feedback about the practical experimental learning method used in the course. Grouping the learners by level of prior experience identified that the negative statements were made by participants with prior learning experience, which appeared to influence their views of how the subject should be taught and learnt. However, the course was designed for beginners; therefore, the levels of prior experience needed to be considered when analysing data. Sentiment analysis has enabled a nuanced evaluation of learner experience by learner group and has aided the course team to make design decisions informed by research, thereby improving the MOOC for future learners.

References

Adamopoulos, P. (2013). What makes a great MOOC? An interdisciplinary analysis of student retention in online courses. *Proceedings of the 34th International Conference on Information Systems*, 1–21. Retrieved from https://aisel.aisnet.org/icis2013/proceedings/BreakthroughIdeas/13/

Breslow, L., Pritchard, D.E., DeBoer, J., Stump, G.S., Ho, A.D., & Seaton, D.T. (2013). Studying learning in the worldwide classroom research into edX's first MOOC. *Research & Practice in Assessment*, *8*, 13–25. Retrieved from https://eric.ed.gov/?id=EJ1062850

Calder, K. (1953). *Statistical inference*. New York: Holt.

Crossley, S., McNamara, D.S., Baker, R., Wang, Y., Paquette, L., Barnes, T., & Bergner, Y. (2015). Language to completion: Success in an educational data mining massive open online class. *Proceedings of the 8th International Educational Data Mining Society*, 388–391. Retrieved from https://eric.ed.gov/?id=ED560771

Donath, J.S. (1999). Identity and deception in the virtual community. In M.H. Smith & P. Kollock (Eds.), *Communities in Cyberspace* (pp. 29–59). London: Routledge.

Downes, S. (2015). The quality of massive open online courses. In B.H. Khan & M. Ally (Eds.), *International handbook of e-learning, volume 1: Theoretical perspectives and research* (pp. 65–77). Abingdon: Routledge.

Gonçalves, P., Dalip, D.H., Costa, H., Gonçalves, M.A., & Benevenuto, F. (2016). On the combination of "off-the-shelf" sentiment analysis methods. *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, 1158–1165. doi: 10.1145/2851613.2851820

Hong, Y., & Skiena, S. (2010). The wisdom of bookies? Sentiment analysis vs. the NFL point spread. *Proceedings of the International Conference on Weblogs and Social Media (IcWSm-2010)*, 251–254. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewPaper/1527

Huang, H. (2002). Toward constructivism for adult learners in online environments. *British Journal of Educational Technology (BJET)*, *33*, 27-37. doi: 10.1111/1467-8535.00236

Hutto, C.J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Eighth International AAAI Conference on Weblogs and Social Media*, 216–225. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109

Kop, R., Fournier, H., & Mak, J.S.F. (2011). A pedagogy of abundance or a pedagogy to support human beings? Participant support on massive open online courses. *The International Review of Research in Open and Distributed Learning*, *12*(7), 74–93. doi: 10.19173/irrodl.v12i7.1041

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167. doi: 10.2200/S00416ED1V01Y201204HLT016

Liyanagunawardena, T. R., Parslow, P., & Williams, S. A. (2017). Exploring "success" in MOOCs: Participants' perspective. *Massive Open Online Courses and Higher Education: Where to Next?*, 92–108. Retrieved from http://centaur.reading.ac.uk/68956/1/BookChapter%207.pdf

Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of massive open online courses (MOOCs). *Computers & Education*, *80*, 77–83. doi: 10.1016/j.compedu.2014.08.005

McAuley, A., Stewart, B., Siemens, G., & Cormier, D. (2010). *The MOOC model for digital practice*. Retrieved from http://www.davecormier.com/edblog/wp-content/uploads/MOOC_Final.pdf

Merrill, M.D. (2002). First principles of instruction. *Educational Technology Research and Development*, *50*(3), 43–59. doi: 10.1007/BF02505024

Miller, M., Sathi, C., Wiesenthal, D., Leskovec, J., & Potts, C. (2011). Sentiment flow through hyperlink networks. *Proceedings of the Fifth International Conference on Weblogs and Social Media*, 550–553. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewPaper/2883

Moreno-Marcos, P.M., Alario-Hoyos, C., Merino, P.J., Estévez-Ayres, I., & Kloos, C.D. (2018). Sentiment analysis in MOOCs: A case study. *IEEE Global Engineering Education Conference (EDUCON)*, 1489-1496. Retrieved from https://ieeexplore.ieee.org/abstract/document/8363409

Pérez, R.C., Jurado, F., & Villen, A. (2019). Moods in MOOCs: Analyzing emotions in the content of online courses with edX-CAS. *IEEE Global Engineering Education Conference (EDUCON)*, 1467-1474. Retrieved from https://ieeexplore.ieee.org/abstract/document/8725107

Perkins, D.N., Hancock, C., Hobbs, R., Martin, F., & Simmons, R. (1986). Conditions of learning in novice programmers. *Journal of Educational Computing Research*, *2*(1), 37–55. doi: 10.2190/GUJT-JCBJ-Q6QU-Q9PL

Pratt, J.W. (1959). Remarks on zeros and ties in the Wilcoxon signed rank procedures. *Journal of the American Statistical Association*, *54*(287), 655–667. doi: 10.1080/01621459.1959.10501526

Shapiro, S.S., & Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611. doi: 10.2307/2333709

Sharples, M. (2013). *Social learning and large scale online learning*. Retrieved from https://about.futurelearn.com/blog/massive-scale-social-learning

Swinnerton, B., Hotchkiss, S., & Morris, N. (2017). Comments in MOOCs: Who is doing the talking and does it help? *Journal of Computer Assisted Learning*, *33*(1), 51–64. doi: 10.1111/jcal.12165

Tumasjan, A., Sprenger, T.O., Sandner, P.G., & Welpe, I.M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Proceedings of the Fourth International Conference on Weblogs and Social Media*, 178–185. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewPaper/1441

Wen, M., Yang, D., & Rosé, C. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? *Proceedings of Educational Data Mining*, 130–137. Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/long%20papers/130_EDM-2014-Full.pdf