M-CAFE 1.0: Motivating and Prioritizing Ongoing Student Feedback During MOOCs and Large on-Campus Courses using Collaborative Filtering

Mo Zhou[1] Alison Cliff[1] Sanjay Krishnan[2] Brandie Nonnecke[3]

Camille Crittenden[3] Kanji Uchino[4] Ken Goldberg[1,2]

[1]UC Berkeley
IEOR Department
{mzhou, alisoncliff, goldberg}@berkeley.edu

[2]UC Berkeley
EECS Department
sanjay@eecs.berkeley.edu

[3]UC Berkeley
CITRIS Connected Communities Initiative
{nonnecke, ccrittenden}@berkeley.edu

[4]Fujitsu Laboratories of America
kanji@us.fujitsu.com

ABSTRACT

During MOOCs and large on-campus courses with limited face-to-face interaction between students and instructors, assessing and improving teaching effectiveness is challenging. In a 2014 study on course-monitoring methods for MOOCs [30], qualitative (textual) input was found to be the most useful. Two challenges in collecting such input for ongoing course evaluation are insuring student confidentiality and developing a platform that incentivizes and manages input from many students. To collect and manage ongoing ("just-in-time") student feedback while maintaining student confidentiality, we designed the MOOC Collaborative Assessment and Feedback Engine (M-CAFE 1.0). This mobile-friendly platform encourages students to check in weekly to numerically assess their own performance, provide textual ideas about how the course might be improved, and rate ideas suggested by other students. For instructors, M-CAFE 1.0 displays ongoing trends and highlights potentially valuable ideas based on collaborative filtering. We describe case studies with two EdX MOOCs and one on-campus undergraduate course. This report summarizes data and system performance on over 500 textual ideas with over 8000 ratings. Details at http://m-cafe.org.

Author Keywords

Education; Course evaluation; Collaborative filtering; MOOCs

ACM Classification Keywords

K.3.1 **[Computers Users in Education]**: Distance learning.
H.5.3 **[Information Interfaces and Presentation]**: Group and Organization Interfaces – Web-based interaction

SIGITE'15, September 30-October 03, 2015, Chicago, IL, USA

ACM 978-1-4503-3835-6/15/09···$15.00
DOI: http://dx.doi.org/10.1145/2808006.2808020

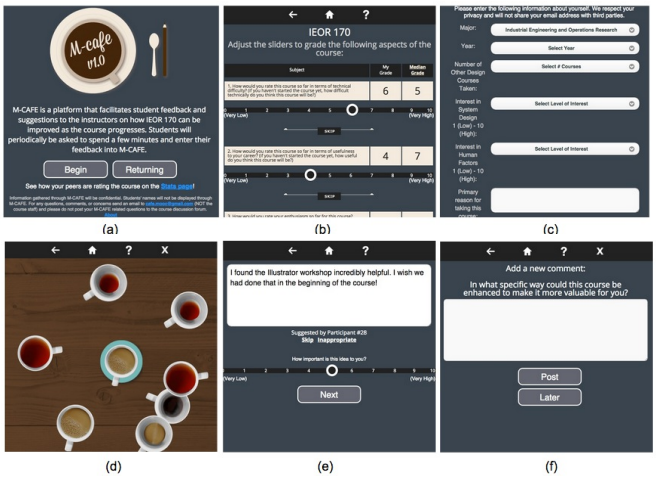

Figure 1: Screenshots of M-CAFE 1.0 Interface.

1. INTRODUCTION

A widely recognized indicator of teaching effectiveness is student interaction with instructors [2, 10, 18]. However, for MOOCs and large on-campus courses it is difficult to provide such interaction [5, 14, 20, 26]. Although most MOOCs and large on-campus courses include some form of end-of-course evaluation, new methodologies offer potential to facilitate student interaction and provide more frequent feedback to instructors.

Most instructors and administrators rely heavily on quantitative data from end-of-course student evaluations to assess teaching performance post-hoc. However, quantitative data cannot capture open-ended observations and suggestions, and is prone to bias [4, 15, 28]. Qualitative data can be a valuable supplement, but the volume and unstructured nature of the data pose an analysis challenge. Stephens-Martinez et al. found that most instructors found little value in transcripts from course chat rooms because they are too voluminous [30]. Furthermore, Piazza and other existing discussion forums focus on Q&As about specific course content and are not designed to provide instructor feedback on how to improve the course in general. New tools are needed to encourage students to speak frankly, motivate regular participation, and manage input effectively so it is not overwhelming for instructors. The tools would create a dynamic feedback-reaction system between students and instructors while insuring student confidentiality. One approach to manage input is Collaborative Filtering, a popular approach based on peer-to-peer evaluation on

items that is widely used to help users find potential interests at Amazon, Netflix, and many other applications.

In this paper, we present the MOOC Collaborative Assessment and Feedback Engine (M-CAFE 1.0), a mobile and web-based platform designed to encourage students to check in weekly to quantitatively assess the course, their own performance, provide qualitative ideas about how the course might be improved, and rate ideas from other students. This platform builds on Opinion Space [21] and extends a preliminary report earlier this year on M-CAFE 1.0 [35] with more data from an on-campus course and an assessment of system performance. M-CAFE 1.0 uses a combination of importance sampling and statistical analysis to quickly and collaboratively identify valuable ideas. M-CAFE 1.0 is complementary to existing platforms, such as Piazza and stackExchange and allows students to step back and consider their own performance and the performance of their instructors, filling the gap between voluminous transcripts from existing platforms and a one-time-only, end-of-course evaluation. M-CAFE 1.0 is also independent of the registrar database to maintain student confidentiality.

2. RELATED WORK

2.1 Course Evaluations

End-of-course evaluations are an almost universal resource for instructors and administrators to obtain direct assessment on teaching [3, 6, 16, 33]. It consists primarily of quantitative assessment on course aspects, including clarity of lectures, usefulness of homework assignments and difficulty of course concepts. Some instructors collect unofficial mid-term student evaluations to obtain ongoing feedback on instruction effectiveness. Since Fall 2010, MIT started using Online Subject Evaluation for both mid-term and end-of-term evaluations across the entire campus, facilitating adjustments of instruction during the course. The lack of efficient ongoing course evaluation platforms and the outdated paper-based evaluations motivated us to explore a more efficient and frequent feedback platform.

Several studies are critical of purely quantitative course evaluations [4, 15, 23]. Stark and Freishtat [29] argue that quantitative rating data from course evaluations are subject to significant self-selection biases. Other subjective factors such as perceived grading leniency [15] have also been implicated in introducing bias. Braga et al. [4] found a negative correlation between student evaluation of course effectiveness and performance in follow-on courses. A related critique [22] is that "apples-to-oranges" aggregations of student ratings on a fixed numerical scale cannot inherently represent subjectivity. Studies agree that qualitative (textual) feedback can be valuable [17, 30] but difficult to analyze for large courses.

We designed M-CAFE 1.0 to collect both qualitative and quantitative ongoing feedback throughout the course, offering the opportunity for students to provide weekly input and instructors to respond to feedback in a timely manner. We observe changes in quantitative rating data from week to week for relative analysis and utilize peer-to-peer collaborative filtering to enable rapid identification of insightful ideas.

2.2 Student Satisfaction & Perceived Learning

Student satisfaction is a significant predictor of learning outcomes. Studies find that clarity of design, interaction with instructors and active discussion among course participants significantly influence students' satisfaction in online learning [11, 31]. Richardson and Swan further recognized that students with higher perceptions of social presence in the courses had significantly higher scores in perceived learning and perceived satisfaction with the instructor [27]. Therefore, most MOOCs and large courses include some version of discussion forums to collect qualitative input and to facilitate peer interaction and instructor engagement [7, 8, 9]. However, existing forums such as Piazza, stackExchange and Internet Relay Chat can be intimidating to students and instructors when the quantity of text is overwhelming. M-CAFE 1.0 facilitates student interaction with instructors and among course participants, and provides timely feedback to the instructor, making in-time course design modifications possible. Furthermore, M-CAFE 1.0 encourages students to assess and track their own motivation, enthusiasm, and performance over the duration of the course.

2.3 Natural Language Processing (NLP)

One way to address the scale issue of qualitative data is to use Natural Language Processing (NLP). Adamopoulos applied two opinion mining tools [24], an orientation analysis mechanism and a sentiment analysis mechanism to course reviews to identify what course features affected the retention rate [1]. Reich et al. conducted text analysis using a variant of Latent Dirichlet Allocation called a Structural Text Model (STM), which identifies popular topics in text, relationships between text and correlation patterns between topics to find insightful patterns in discussion forum text [32]. In the M-CAFE 1.0 setting, however, insightful ideas may be rare and could vary greatly in content from week to week, making it difficult to infer from word-document structure. In course evaluations, NLP may identify popular topics, but the proposed insights are more important.

2.4 Collaborative Filtering (CF)

We explore an alternative approach for qualitative analysis called Collaborative Filtering (CF) [12, 13, 19, 25]. CF is rating-based and relies on the crowd to find insightful items in a large dataset. This is in contrast to prior approaches, which are content-based and rely on mathematical representations of items. CF has commonly been used to recommend books and movies. The idea is to combine subjective evaluations provided by humans to assign a numerical reputation to each item. Often, the reputation values are computed based on a local neighborhood for customized recommendations but similar techniques can also assign global reputations in a "peer-to-peer" approach [28]. Recent studies also demonstrate that social network information can improve accuracy of CF-based recommender systems [34]. For M-CAFE 1.0, we utilize peer ratings on each textual idea and combine standard error to solicit feedback with the Wilson metric to compute reputation values and rank ideas.

Due to the self-selection nature of user rating, most CF systems suffer from sparse ratings matrix with many null values [34]. The popular usage of a list-based presentation of items to users can be responsible for the problem, in which case, highly rated items are shown on top of the list, enjoying greater exposure. In M-CAFE 1.0, we balance the exposure of items by simultaneously providing a subset of items (normally 6) with mixed rankings, permitting the system to collect feedback on all items.

3. M-CAFE 1.0 User Interface Summary

Upon entering M-CAFE 1.0 (Figure 1a), students are required to register by email and are given the option to provide their age, gender, home country, years of college-level education, and the primary reason for taking the course (Figure 1b). Then they rate five quantitative assessment topics (QAT) on a scale of 1 to 10: Course Difficulty, Course Usefulness, Self-enthusiasm, Self-performance and Homework Effectiveness (Figure 1c). Students click on mugs (Figure 1d) to view their peers' ideas, evaluate how valuable the ideas are on a scale of 1-10 (Figure 1e) and suggest new ideas (Figure 1f).

4. Case Studies and Analysis

M-CAFE 1.0 has so far been used in three courses: two MOOCs on edX: CS 169.2x and CS 169.1x, and a face-to-face classroom-based undergraduate course - IEOR 170 (taught by co-author Prof. Ken Goldberg), all offered through UC Berkeley. Students were invited to participate at the beginning of the course, and email reminders were sent on a weekly basis. Table 1 summarizes M-CAFE 1.0 participation statistics for the courses. M-CAFE 1.0 is fully confidential and no individual identity is revealed on the platform or to the instructors. Since participation in M-CAFE 1.0 is voluntary, it inherently suffers from self-selection bias, i.e., students who are more actively involved in the course tend to participate more often in M-CAFE 1.0. However, considering that M-CAFE 1.0 aims to collect valuable feedback for the instructors, we expect the active students to provide more insightful ideas because they are more invested in the course and its outcomes.

Table 1: Participation statistics in different stages of M-CAFE 1.0 for the three courses.

	CS 169.2x	CS 169.1x	IEOR 170
Student count	348	253	96
QAT set rating count	741	312	424
Idea count	167	82	270
Peer-to-peer rating count	4,000	1,715	2,483
Date range of the course	Jun - Jul, 2014	Oct - Dec, 2014	Jan - May, 2015
Term length of the course	6-week	8-week	16-week

4.1 Quantitative Analysis Topics

4.1.1 Graph Visualization of QAT Rating Changes

For each quantitative analysis topic, an average score and the associated standard error is computed each week. The changes in ratings over the weeks are then plotted to provide a straightforward visualization of the QATs to instructors. For example, Figure 2 is a plot of the course difficulty ratings over ten weeks, generated from M-CAFE 1.0 data collected in IEOR 170.

Figure 2 provides a visualization of the changes in course difficulty over time. By viewing the plots, instructors can quickly identify the average changes from past weeks. The error bars indicate two standard errors (SE) above and below average, revealing the significance of changes at a 5% significance level. As we can see immediately from the plot, the course difficulty level increased gradually from the beginning of the semester to the latter half. It reached its peak in week 7 after the middle of the term, and at the time, the course was significantly more difficult than when it began.

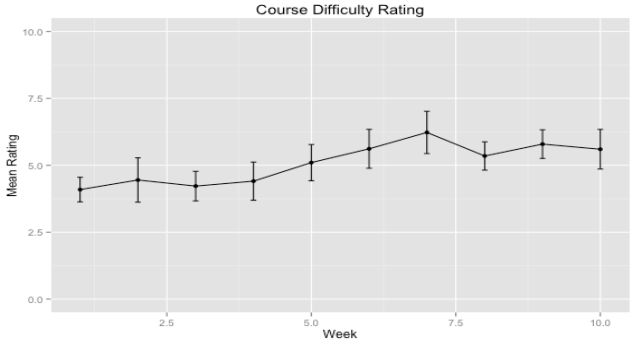

Figure 2: The average and two-standard errors of ratings on course difficulty over ten weeks of IEOR 170.

Revealing the relative changes in average QAT ratings provides instructors insights on the impacts of courses event changes, for example, how does this homework/exam affect students' perceptions of the course aspects? The visualizations quantify the effectiveness of individual course events and yield a decomposed evaluation on detailed course activities within fixed time intervals.

4.1.2 Understanding the Relationships Between QAT Rating Changes – a Validity Check

The quantitative feedback feature of M-CAFE 1.0 also provides the possibility of assessing the relationships between weekly average QAT rating changes. In turn, the agreement of the relationships between QAT rating changes and common beliefs could further demonstrate the reliability of the quantitative feedback from M-CAFE 1.0. Stark et al. [29] points out that quantitative scores in course evaluations suffer from validity concerns and may not be informative. Thus, we believe this analysis would be a valuable consistency check for the quantitative feedback obtained in M-CAFE 1.0.

In all three courses that implemented M-CAFE 1.0, we observe a negative correlation between course difficulty and self-enthusiasm, suggesting that difficult course materials lead to higher anxiety and thus reduced student enthusiasm. Course usefulness is positively correlated with homework effectiveness and self-enthusiasm. We speculate that as students become more enthusiastic and the homework assignments become more effective, the usefulness of the course would be rated more highly. The other relationships are not consistent among the three courses and are possibly dependent on the different student bodies and course materials involved.

The agreement of the relationships between QAT rating changes and common beliefs is encouraging and is the building block of further analysis on the QAT ratings.

4.1.3 Social Influence Bias and Student Confidence

The visualizations (see Figure 2) for the QATs are available not only to instructors but can also be made available to students. Students can evaluate their ratings against the class average to get a better idea of where they stand among the peers. For example, if one student in CS 169.2x is finding the homework in week 4 particularly challenging and considers dropping the course, he might feel less stressed and gain some confidence if he learns that most of his classmates are in the same situation, i.e. the course difficulty rating is much higher in week 4 than the previous weeks. M-CAFE 1.0 tries to summarize information in minimal volume and at the same time, provide a representative indicator of the course aspects that can be valuable to both instructors and students.

Furthermore, M-CAFE 1.0 displays the median grade on each QAT after the student provides a rating. This feature and the QAT visualizations can potentially lead to less bias by reducing apple-to-orange comparisons in numerical scoring. One major concern about quantitative feedback on course evaluations is the varied scales among students, i.e., two students who feel the same difficulty level may provide different ratings because they don't have a common scale to refer to. M-CAFE 1.0 reduces the scale variability among students by giving the median, the average and the standard error of the ratings, allowing students to acquire knowledge of a "middle ground" rating on course aspects. However, unlike traditional paper evaluations, interactive systems are susceptible to social-influence bias, where students can change their ratings after seeing the

median or rate the course close to the average rating. It would be interesting to investigate but it is beyond the scope of this paper.

4.2 Qualitative Feedback with CF

4.2.1 Identifying the Most Valuable Ideas for Instructors

A shortcoming of qualitative data is its lack of structure. Natural Language Processing (NLP), although has been an active research field for years, is not effective in selecting a subset of insightful ideas from M-CAFE 1.0-generated qualitative data. Current text analysis of qualitative data hints at the important words or phrases, whereas the underlying sentence structure and word meanings are mostly ignored. As an alternative approach to NLP, Collaborative Filtering (CF) has gained popularity for ranking and recommending items in fields using peer-to-peer ratings.

4.2.2 The Ranking Metric

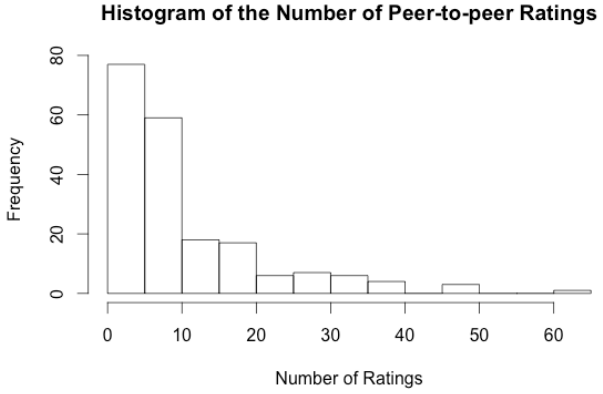

Figure 3: Histogram of the number of peer-to-peer CF ratings for ideas in IEOR 170.

As can be seen in Figure 3, few ideas received more than 20 peer-to-peer CF ratings. Ranking ideas by their mean or median rating would not be reliable due to the small sample size and the variation in rating differences. Instead, we compute the Wilson score for each idea using the lower bound of the binomial proportion confidence interval. This incorporates the variance in ratings as follows. We took the mean grade **g** and then calculated the 95% confidence interval of **g** using standard error: **g +/- 1.96*SE(g)**. We then rank the ideas by the lower bound **g - 1.96*SE(g)**.

4.2.3 M-CAFE 1.0 Customized CF Implementation

In M-CAFE 1.0, we adopt a 10-point scale for rating qualitative ideas. One major drawback of the CF approach is the bias in item exposure resulting from ranked-list presentation. For example, an item that receives low ratings from the first several users would be at the bottom of the list, reducing its future exposure to the new users who might rate it more highly. This is particularly relevant to the CF implementation in M-CAFE 1.0 because students are likely to hold varying opinions and low initial ratings may not reflect the attitude of the class. To overcome this problem, instead of a ranked-list design, we adopt an interface that displays a subset of ideas simultaneously as seen in Figure 1d.

4.2.4 CF Performance Assessment

A popular method to assess CF performance in a recommender system is to predict the rating a user will give to a new item and compare it to the actual rating that user provided. However, in the M-CAFE 1.0 setting, our focus is not to predict student ratings but to provide the instructor with a CF subset, a subset of potentially valuable ideas. Thus we assess the performance of CF in M-CAFE 1.0 from three perspectives:

1. Does the CF subset cover a broad range of topics?
2. Is the CF subset similar to a subset selected manually?
3. Is the ranking of the CF subset similar to the instructor's ranking?

4.2.4.1 Topic Coverage

Many ideas suggested are similar. We evaluate results from CS 169.2x, the course with the most peer-to-peer ratings, to assess CF on topic coverage.

As topics emerged, we manually tagged the ideas with keywords to separate them into different categories. The following topics were most popular for CS 169.2x:

1. Chat forum: students expressed their concerns with lack of activity of peers and TAs in the chat forum, stackExchange, which inadequately facilitated discussion and assistance.
2. Basics: this category included any idea that requested extra help on the basic fundamentals of Ruby, Rails or Rspec. Specific ideas to aid students, including extra exercises, more basic exercises, a review lecture, etc.
3. Javascript: this category included requests for more lecture time and homework devoted to Javascript.
4. Additional time: additional time was requested between homework releases and due dates, as well as between the end of CS169.1x and the beginning of CS 169.2x.
5. Additional exercises: this category included ideas that requested additional exercises on various topics, excluding Javascript and the basics of Rails and Rspec.
6. Security: these ideas requested that security related materials be covered in more depth.
7. Update technology: these ideas requested the course to use the most up-to-date version of Rails (i.e., Rails 4 instead of Rails 3) and other platforms or tools.

Next, we picked the top 20 ideas from the CF ranking and investigated their topics. Figure 4 illustrates the distribution of the topics of these ideas. They cover a wide range of topics with few replications, demonstrating the effectiveness of CF in selecting valuable ideas with broad topic coverage and fewer repeats. Even though results show a wide variety of topics among the top ideas, we could improve its efficiency further with topic tagging.

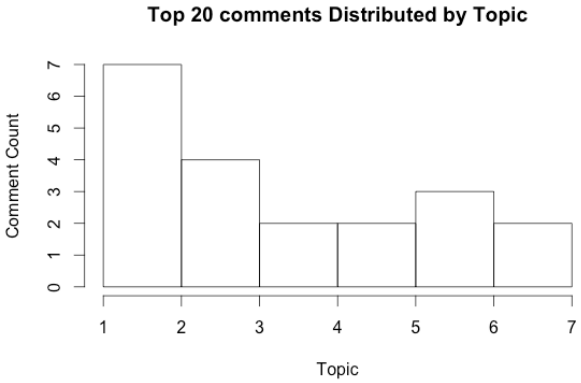

Figure 4: The number of ideas of each topic among the top 20 ideas in CS 169.2x.

4.2.4.2 Value of Ideas in the CF subset

For the data from CS 169.1x and CS 169.2x, we manually rated the quality of ideas. Each idea is given a quality score between 1 and 5, where:

1: Not readable
2: Readable but not relevant
3: Not a constructive suggestion
4: Includes a constructive criticism
5: Includes constructive criticism, rationale, and potential solution

For example, an idea with a quality score of 1 is:

Devise + Omniauth !!!

And an idea with a quality score of 5 is:

Design patterns are hard to grasp without getting your hands dirty in a messy problem. I think using a quiz for that week instead of a challenging homework assignment was a mistake. I understand the concepts as abstract entities but would still have a hard time figuring out when and how to use them. I felt the same way about the Javascript week as well. A homework assignment doing JS and AJAX on the rotten potatoes example would have been ideal.

We observed that the quality scores follow a somewhat normal distribution. We evaluate the relationship between the Wilson scores of the ideas and their quality scores by fitting a linear regression. Taking the Wilson scores as the dependent variable and the quality scores as the regressor, we observe that the coefficient of the quality variable is significantly positive with a p-value of 0.0165 for the data from CS 169.2x and 0.0172 for the data from CS 169.1x, suggesting that the CF subset scores are correlated with manually provided scores.

4.2.4.3 CF Ranking vs. Instructor Ranking

The instructor of IEOR 170 (Prof. Goldberg) manually ranked what he felt were the top 30 of the 270 ideas suggested. But this ranking was not consistent with the ranking performed with CF. We speculate that this could be due to the low sample size (many ideas had only 4 ratings), repetition among idea topics, and the framing of the student rating question which was "How important is this idea to you?" In the next version we will modify this to: "How valuable is this idea for the instructor to consider?" To address the other two concerns, we would experiment other ways of presenting ideas to encourage student participation and add a topic-tagging feature to reduce topic repetition of ideas.

5. INSTRUCTOR INTERFACE

To improve the effectiveness of M-CAFE 1.0, we developed a weekly update for instructors, which include participation statistics, visualizations of the week-to-week QAT ratings with error bars and a list of the top-rated ideas from the previous week. Instructors and students can view M-CAFE 1.0 statistics on the statistics page. M-CAFE 1.0 is a flexible platform in that the QATs and the discussion question can be easily modified through a user-friendly admin panel. Instructors can collect feedback according to personal interest and develop customized discussion topics. In addition, instructors can flag users or mark ideas as resolved.

6. CONCLUSION

In this paper, we introduce a new course evaluation platform, M-CAFE 1.0, which aims to collect ongoing student evaluation on course issues to generate timely feedback. M-CAFE 1.0 provides visualizations and peer-to-peer collaborative filtering to highlight insightful ideas.

7. FUTURE WORK

In future work, we will refine the graphics and text in the M-CAFE interface, build a new code base and add topic tagging to organize suggested ideas, reduce repetition, and encourage input and evaluation of diverse ideas. We will also explore how sorting and presenting ideas based on factors such as time or novelty will affect participation.

8. ACKNOWLEDGEMENTS

This work is supported in part by a grant from Fujitsu Laboratories, the Blum Center for Developing Economies and the Development Impact Lab (USAID Cooperative Agreement AID-OAA-A-12-00011) as part of the USAID Higher Education Solutions Network, UC Berkeley's Algorithms, Machines, and People (AMP) Lab, and the UC CITRIS Connected Communities Initiative. Thanks to Armando Fox, Dan Garcia, Animesh Garg, Björn Hartmann, Marti Hearst, Allen Huang, Sam Joseph, Steve McKinley and Kristen Stephens-Martinez. Thanks to all students who participated and colleagues and reviewers who provided feedback on earlier drafts.

9. REFERENCES

[1] Adamopoulos, P. What makes a great MOOC? An interdisciplinary analysis of student retention in online courses. *Thirty Fourth International Conference on Information Systems.* (Milan, Italy, 2013).

[2] Angelo, T.A. Cross, P.K. Classroom assessment techniques: A handbook for faculty. *National Center for Research to Improve Teaching and Learning 1993.* (Ann Arbor, MI, 1993).

[3] Airasian, P. W., and Russell M.K. Classroom assessment: Concepts and applications. McGraw-Hill, Boston, 2001.

[4] Braga, M., Paccagnella, M., and Pellizzari, M. Evaluating students' evaluations of professors. *Economics of Education Review, 41*, (2014), 71-88.

[5] Breslow, L. Studying Learning in the Worldwide Classroom Research into edX's First MOOC. *Research and practice in Assessment 8,* (2013) 13-25.

[6] Clayson, D.E. Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education 31.1,* (2009), 16-30.

[7] Clow, D. MOOCs and the funnel of participation. *Third Conference on Learning and Analytics Knowledge (LAK) 2013 Conference.* (Leuven, Belgium. April 20, 2013)

[8] Coetzee, D., Fox, A.,Hearst, M.A., and Hartmann, B. Chatrooms in MOOCs: All Talk and No Action. *Proceedings of the First ACM conference on Learning @ Scale Conference.* (2014)

[9] Coetzee, D., Fox, A., Hearst, M.A. and Hartmann, B. 20 Should Your MOOC Forum Use a Reputation System? In *Proceedings of the 17th ACM Conference on Computer-Supported Collaborative Work* (Baltimore, MD, 2014)

[10] Cotton, K. Monitoring Student Learning in the Classroom. *Northwest Regional Educational Laboratory.* (1988).

[11] Eom, S. B., Wen, H. J., & Ashill, N. The Determinants of Students' Perceived Learning Outcomes and Satisfaction in University Online Education: An Empirical Investigation. *Decision Sciences Journal of Innovative Education*,*4*(2). (2006), 215-235.

[12] Goldberg, D., Nichols, D., Oki, B. M. and Terry, D. Using Collaborative Filtering to weave an Information Tapestry. *Communications of the ACM, 35(12).* (1992), 61-70.

[13] Goldberg, K., Roeder, T.,Gupta, D., and Perkins, C. Eigentaste: A Constant Time Collaborative Filtering Algorithm. *Information Retrieval Journal*, 4(2). (2001), 133-151.

[14] Graham, C., Cagiltay, K., Lim, B., Craner, J., & Duffy, T. M. Seven principles of effective teaching: A practical lens for evaluating online courses. *The Technology Source, 30(5),*(2011).

[15] Greenwald, A. G. and Gillmore, M.G. Grading leniency is a removable contaminant of student ratings. *American psychologist 52, 11. (*1997), 1209-1217.

[16] Harvey, L., and Askling, B. Quality in higher education. In *The dialogue between higher education research and practice*. Springer Netherlands. (2003), 69-83.

[17] Hill, P. Emerging Student Patterns in MOOCs: A (Revised) Graphical View http://mfeldstein.com/

[18] Kizilcec, R. F., Piech, C., and Schneider, E. Deconstructing disengagement: Analyzing learner subpopulations in Massive Open Online Courses categories and subject descriptors. *In Proceedings of the Third International Conference on Learning Analytics and Knowledge*. (2013), 170–179.

[19] Konstan, J.A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., and Riedl, J. Group Lens applying collaborative filtering to Usenet news. *Communications of the ACM, 40(3),* (1997), 77-87.

[20] Koller, D., Ng, A., Do, C., Chen, Z. Retention and Intention in Massive Open Online Courses: In Depth. *Educause Review*. (June, 2013) http://www.education .edu/

[21] Krishnan, S., Goldberg, K., Uchino, K., and Okubo, Y., Using a Social Media Platform to Explore How Social Media Can Enhance Primary and Secondary Learning. *Learning International Networks Consortium (LINC) 2013 Conference.* (MIT, Cambridge, MA. 2013).

[22] Marsh, H.W., and Roche, L.A. Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist, 52(11).* (1997)

[23] McKeachie, W.J. Student ratings: The validity of use. *American Psychologist, 52(11).* (1997), 1218-1225.

[24] Pang, B and Lee, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2(1- 2)*. (2008), 1-135.

[25] Pearson, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 2.* (1901), 559–572.

[26] Quillen, I. Why do students enroll in (but don't complete) MOOC courses? http://blogs.kqed.org/mindshift

[27] Richardson, J. C., Swan, K. Examing social presence in online courses in relation to students' perceived learning and satisfaction. (2003).

[28] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. Item-based Collaborative Filtering recommendation algorithms. *In Proceedings of the 10th International Conference on World Wide Web.* ACM Press (2001), 285-295.

[29] Stark, P. B. and Freishtat, R. An evaluation of course evaluations. *UC Berkeley Technical Report; Center for Teaching and Learning.* (2014).

[30] Stephens-Martinez, K., Hearst, M.A., and Fox, A. Monitoring MOOCs: which information sources do instructors value? *Proceedings of the First ACM conference on Learning @ Scale Conference.* (2014), 79-88.

[31] Swan, K. Virtual interaction: Design factors affecting student satisfaction and perceived learning in asynchronous online courses, Distance Education, 22:2. (2001), 306-331.

[32] Reich, J., Tingley D., Leder-Luis, J., Roberts, M.E., Steward, B.M, Computer-assisted reading and discovery for student generated text in Massive Open Online Courses. (2014).

[33] Weinberg, B.A., Fleisher, B.M. and Hashimoto, M. Evaluating methods for evaluating instruction: The case of higher education. *National Bureau of Economic Research*. (2007).

[34] Yang, X., Guo, Y., Liu Y. and Steck, H. A survey of collaborative filtering based social recommender systems. *Computer Communications Journal, Volume 41*, (2014), 1-10.

[35] Zhou, M., Cliff, A., Huang, A., Krishnan, S., Nonnecke, B., Uchino, K., Joseph S., Fox A. and Goldberg, K. M-CAFE 1.0: Managing MOOC Student Feedback with Collaborative Filtering. *Proceedings of the Second ACM Conference on Learning@ Scale*. (March, 2015), 309-312.