CrowdLearn: Open Sourcing Learner Feedback to Enable Open-Ended Assessments in MOOCs

John Liu
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA U.S.A.
johnhliu@mit.edu

Ewoud de Kok
FeedbackFruits
Amsterdam, Holland
ewoud@feedbackfruits.com

***Abstract*—Open-ended assessments are crucial for robust learning especially common in upper-level university disciplines. This paper describes the iterative development of the *CrowdLearn* tool to enable open-ended, interactive formative and summative assessments in massive open online courses (MOOCs). The tool crowdsources learner feedback to enable assessments such as case studies, essays, and project reports. By giving learners the opportunity to discuss submissions and feedback from all other learners, the tool enables learners to learn from each other. This tool has now been implemented in eight courses hosted on the edX platform that span science, engineering, and social science.**

***Index Terms*—MOOCs, crowdsource feedback, open-ended assessment, collaborative learning, scalable**

I. Introduction

Science and engineering courses often culminate in a capstone project. Management courses tend to teach using the case-study method, where students learn through open-ended analysis and writing, then deepen and sharpen their thinking through in-class debate and reflection. The central assessment of social science and humanities classes is typically a research paper or writing project, developed through multiple rounds of feedback and iterations.

Because of the open-ended nature of these assessments, upper-level courses are difficult to capture fully in the MOOC environment. Most assessments common to MOOC platforms have an intended correct answer. These assessment types, such as multiple choice or numerical response, are limited in their ability to assess learning when there is a range of potentially good answers or learning products. Most commercially-available versions of open-ended assessment at scale require either self- or peer-review, which are limited in ability to evaluate open-ended, interactive assessments at scale.

In this paper, we describe the iterative development of the *CrowdLearn* tool with course designers to enable open-ended assessments in MOOCs at scale. We demonstrate the tool's implementation in courses that span science, engineering, and social science. We present three case examples of how *CrowdLearn* is being used to achieve scalable grading, collaboration, and open-ended assessment.

II. Related Work

Self-assessment has been associated with increases in self-regulated learning [7] and self-efficacy [24]. Because the quality of this type of assessment is limited by bias, its usage should generally be limited to formative as opposed to summative assessments [1]. This is especially true for classes with credential weight. For example, schools that will give graduate program credit for a student who has earned an edX MicroMasters program credential expect that the assessments in an MicroMasters edX course [23] is comparable to its counterpart on campus.

Peer assessment has been shown to improve final grades [17], improve metacognition [7], and peer-review in MOOCs has been shown to increase completion rate [13]. Substantial work has been carried out to study approaches to assignment submission [17], [26], [28], user interfaces [15], [17], types of peer reviewers [14], [17], [29], develop algorithms to calculate grades or assign reviewers [6], [16], [19], [26], [30], and approaches to provide scores and feedback [5], [17], [27].

One current limitation is that MOOC peer assessment is often an isolating process. While assessment feedback is a "social practice" [7], limited work exists in embedding a social process for peer assessment [12]. Open spaces can create social interactions and yield better learning [3], yet most peer assessment types do not integrate with discussions [10], [11].

For example, some major MOOC platforms, such as edX, only offer peer grading and feedback that is one-way and anonymous [21]. The main source of social interaction is over MOOC forums, yet forum activity is mixed in quality, and challenging to assess the quality of a learner's input or its impact on other learners in a scalable way [2], [10]. We further assessed that pre-existing web-based tools that host collaborative learning were not focused on open-ended assessment, could not offer it scalably, and did not readily integrate with the edX platform [4], [20], [31]. These challenges lead to two potential limitations: i) no readily-evaluable discussion or debate among peers [10]; and ii) anonymity removes authentic social pressure that could promote quality of feedback [9].

III. Iterative Design

We used learner reflections and MOOC designer feedback to iteratively design the *CrowdLearn* tool (Figure 1).

2023 IEEE Learning with MOOCS (LWMOOCS) | 979-8-3503-1559-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/LWMOOCS58322.2023.10306150

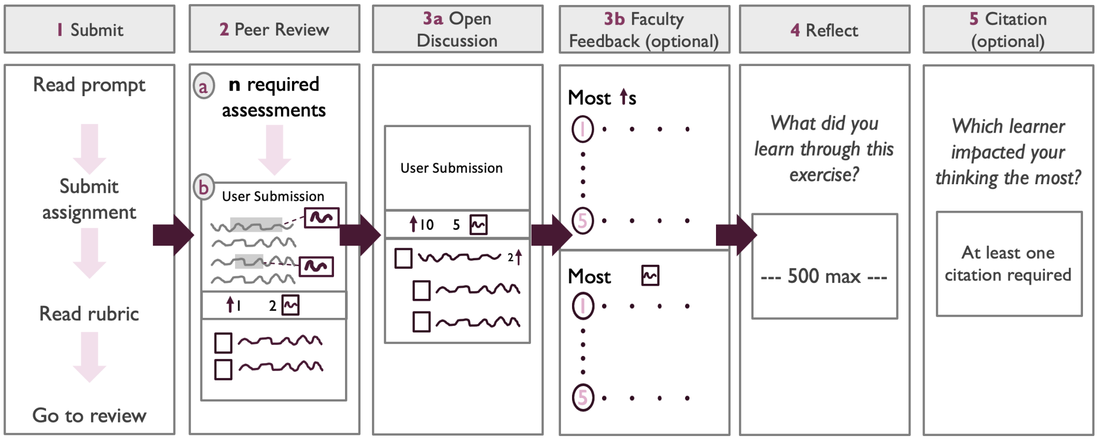

Fig. 1. The design of the *CrowdLearn* tool: submission, peer review, forum-style open discussion and upvoting, faculty comment (optional), reflection, and citation (optional). The fifth section "Citation" was added in Version 2.

A. Version 1

1) Design: The previously described limitations to open-ended assessment were particularly relevant in our specific context, when the first author was developing an engineering management MOOC as part of an edX MicroMasters program [22]. In the management classroom, a student's performance in a case study is usually based on how compelling an argument is in written and verbal form, how well grounded it is in proper analysis, and how students debate different conclusions and tradeoffs in class discussion. Scalably grading learner dynamics – particularly peer debate and discussion – is not available in the edX platform, but could be important to capture when the course has credential weight. We initially set out to develop a tool to enable case study learning in the MOOC environment. We wanted to provide our learners with the rich learner interaction that characterized the management classroom and its robust assessments.

Our initial design focused on assessing both the open-ended case study and learner interactions at scale. We derived our inspiration from online reviews. Instead of hiring tens of thousands of professional reviewers, platforms that offer reviews for products (Amazon), movies (Netflix), books (Goodreads), recipes (Allrecipes) source their own users. These platforms then give the community the option of commenting and upvoting those reviews. This also often allows the most helpful reviews to "float" to the top for subsequent users to see. We designed *CrowdLearn* Version 1 to engage learners through four stages:

1) **Submit**: Learners read the prompt, then write and submit their assignment.
2) **Peer Review**: Learners study the rubric. They are then required to review a set number of peer assessments, where they comment and upvote based on the given rubric.
3) **Open Discussion**: a) Learners' discussion: When learners have finished the formal assessments, they can participate in an open discussion where learners can view and upvote all submissions and peer comments on these submissions. b) Instructor comment (optional): instructors can engage in the open discussion by commenting on the top submissions and discussions. Instructors can also participate where there is most activity, review high-profile submissions, and display their expert thinking, thereby also enabling the class to more accurately interpret the rubric.
4) **Reflect**: Learners bullet-point their major take-aways.

Using this design, the EdTech company FeedbackFruits [8] developed this tool with Learning Tools Interoperability (LTI) integration into the edX platform [18].

2) User-testing: *CrowdLearn* Version 1 was released for case-study instruction into a two-MOOC series, *Management in Engineering*. At this point in the development, we discovered that this tool could also be useful to facilitate project-based learning in an advanced engineering MOOC that we were developing in the same MicroMasters program: *Manufacturing Process Control II*. We used this tool to host two assignments: 1) a critique of four process control papers and 2) a process control "mini-proposal."

These learning activities were made available to the 100-200 learners in each course who had paid for the option to earn an edX certificate, as opposed to the thousands of audit learners. To keep the stakes low for learners while we piloted the tool, we kept both open discussion and reflection ungraded. Because the majority of submitted learner reflections (52 of the total 68 reflections) came from *Manufacturing Process Control II*, we chose to focus our preliminary analysis on this set of reflections. To examine the tool's usage and effect, we classified the comments from reflections into broad categories including impact on learner thinking (corrected, enriched, reinforced) and complaints (Figure 2).

Our preliminary analysis of learner reflections revealed that the discussion and learner feedback enabled by the tool was perceived to be valuable and constructive to learning. The impact that most learners reported was classified as enriching

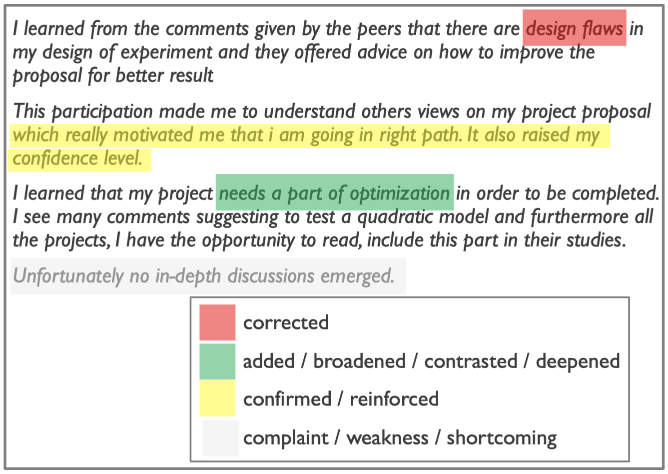

Fig. 2. Representative classifications of comments from reflections into broad categories of impact on learner thinking (corrected, enriched, reinforced), or complaints on the tool.

learner thinking (Figure 3). Learners felt it was valuable to be able to read both their peers' submissions and comments on other submissions, a kind of "open-source" learning. There was surprisingly little commentary on the LTI itself, which suggests that the learners were more focused on the learning activity than the tool. Our analysis additionally identified that learners perceived that the environment enhanced their learning process through collaboration and identifiable interaction in assessments (as opposed to anonymous). Learner reflections also indicated that the tool contributed to their sense of online community while deepening learner engagement with the project work, simultaneously increasing their confidence level and understanding of the course material.

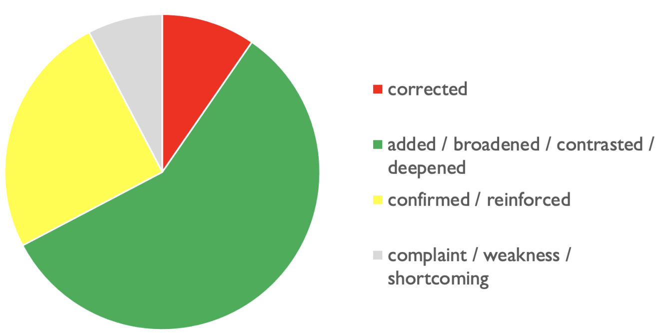

Fig. 3. Breakdown of learner reflections by type of impact on learning (corrected, enriched, confirmed) or complaint on the tool. Analysis of learner reflections reveal that the assignments enabled by the *CrowdLearn* tool enabled a wide range of positive impact on learner thinking.

Given its basic design, Version 1 of the tool demonstrated a few areas of limitation, particularly:

1) Peer review was limited to upvoting. Because of this lack of granularity, staff still had to manually grade the assignments, which meant this version was not yet able to offer scalable assessment.
2) Staff grades had to be submitted to the FeedbackFruits team to manually push to the edX gradebook.
3) Learner navigation of the user interface and staff functionality could be clumsy. This may have challenged further development of a learning community.

B. Version 2

The validation of Version 1's ability to effectively host open-ended assessments and to drive impactful learning motivated further development and adoption. At this point in the development, other MOOC designers at the lead author's institution expressed interest in using the CrowdLearn tool. We gathered a group of eight MOOC designers to brainstorm and rank design features for the proposed Version 2. The MOOC designers first brainstormed the following design features:

1) **Custom grader**: Instructors can set a grading scheme that uses metrics associated with various aspects of the entire learning activity to automatically assign a grade to each learner. These metrics include completion of steps, points given to submission, upvotes on comments, number of comments given or received, and citations (see feature #6).
2) **Grading schemes**: Instructors have a set selection of different grading schemes that they can choose from. These schemes could range from one that primarily weights peer review grades to one that primarily weights peer feedback and interaction.
3) **Gamification**: To de-incentivize spamming, learners start out with a limited number of comments to give. To incentivize positive behavior, every comment that is replied or upvoted they receive more comments to give. Learners have set limits of points to give out during peer review and open discussion. Instructors would be able to turn this option on or off, and also set limits on comments and points.
4) **Leaderboard**: In a board at the top of the discussion, learners can see the top three submissions (by upvotes or points), and submissions that are ranked just ahead of their own. Instructors would be able to turn this on or off.
5) **Flagging**: During peer review or open discussion, learners can privately flag submissions or comments that have a glaring problem, such as an inappropriate comment or an assignment that they suspect exhibits plagiarism.
6) **Citation**: Students are asked after reflection to cite the learner(s) who most positively impacted their thinking. Instructors could use this feature to incentivize learners to write useful submissions and give useful feedback over the duration of the whole activity.
7) **Plagiarism report**: After the reflection step, students are given the option to report any submissions they suspect exhibit plagiarism.
8) **Staff view submissions**: Instructors can cycle through submissions without having to go back to the first page to browse, select, and view another submission.

9) **Special cases**: Instructors can upload, remove, or change submissions for learners who have trouble uploading an assignment, submit the wrong assignment, or want to turn in assignments late.
10) **Staff overwrite**: Instructors can overwrite individual grades based on their own assessment or in response to a learner appeal.
11) **Learner search**: The open discussion stage has a search function to look for usernames or keywords in comments.
12) **Comment navigation**: Instructors and learners in open discussion will be pointed directly to new comments, or an activity feed that gives an overview of all comments and replies.
13) **LaTeX**: Functionality to type comments in LaTeX.
14) **Cohorts**: Open discussion can be carried out in smaller cohorts, the assignment submission and discussion can be set up to handle group work.
15) **Notification control**: Learners can control options on how *CrowdLearn* notifies relevant activity (e.g. emails them when another learner reviews their assignment or responds to their comment).
16) **Auto-search plagiarism**: Instructors can search all submitted assignments for similarities between assignments and works on the web.
17) **Analytics+**: Instructors have access to in-depth learner engagement analytics, such as data visualization of interactions between learners, or interactions in open discussion.
18) **Default edX ORA**: Instructors can choose an option that automatically changes the design and structure of Stage 1 and 2 to be exactly patterned after edX's open-response assessment.

Features were scored by adding "must have" and "will use" votes, with "must have" votes weighted double. Those that scored five and above were included in Version 2's design (Table 1). Due to budgetary constraints, auto-search plagiarism was not included in the final development.

As a result of this collaborative process, the proposed design features focused on scalable and automatic grading, learner interaction and navigation, and instructor functionality. The custom grader feature was chosen over grading schemes because MOOC designers felt they were unable to predict how they would eventually want to grade their assignments and wanted maximum flexibility. Features that enhanced learner interactions such as gamification and leaderboard were considered by many MOOC designers to be interesting but not at the core of *CrowdLearn's* added value. Most also felt other features like learner notifications, default edX, analytics+ were nice to have but not necessary for them to adopt the new tool in their course. Overall, the chosen and developed features enabled *CrowdLearn* Version 2 to address limitations in Version 1, enable greater scalability, and ensure its usefulness to host assessments in a range of different disciplines.

TABLE I
BRAINSTORMED AND SCORED DESIGN FEATURES OF *CrowdLearn* V2

Type	No.	Name	Must-have	Will use	Score
Scalable Auto Grading	1	Custom grader*	6	1	13
	2	Grading schemes	2	0	4
Learner Interaction	3	Gamification	0	2	2
	4	Leaderboard	0	2	2
	5	Flagging*	2	5	9
	6	Citation*	2	5	9
	7	Plagiarism report	0	2	2
Instructor Functionality	8	Staff view sub.*	4	3	11
	9	Special cases*	7	1	15
	10	Staff overwrite*	8	1	17
Other Features	11	Learner search*	0	6	6
	12	Comment navigation*	0	6	6
	13	LaTeX*	4	0	8
	14	Cohorts*	1	3	5
	15	Notification control	0	3	3
	16	Auto-search plag.	1	5	7
	17	Analytics+	0	2	2
	18	Default edX ORA	0	1	1

*Features scored 5 and above were chosen.

IV. Course Implementation

Version 2 was implemented into the two MOOCs previously described and six additional MOOCs in manufacturing, data science, material science, and social science. Across these eight courses, thousands of learners have used the *CrowdLearn* tool to complete assessments ranging from case studies, essays, projects, proposals, and critiques (Table 2). Assessments offered in *CrowdLearn* occupy a significant portion of the overall grade in these courses, ranging from 20% to 100%. These courses enroll learners in numbers ranging from scores to hundreds.

We present three examples of how *CrowdLearn* is being used in courses of different disciplines to achieve scalable grading, collaboration, and open-ended assessment.

A. Policy for Science, Technology and Innovation

This course teaches innovation systems to develop science and technology. Each week, learners study the underlying

TABLE II
COURSE IMPLEMENTATION

Course name	Level	Assess type	# of learners assessed	% grade
Policy for Science, Technology and Innovation	Undergrad.	Case study	200-300	50%
Tools for Academic Engagement in Public Policy	Grad.	Essay	200-300	100%
Global Shakespeares: Re-Creating the Merchant of Venice	Undergrad.	Video of a self-produced scene	30-50	40%
Data Analysis: Statistical Modeling and Computation in Applications	Grad.	Projects	300-500	50%
Structural Materials: Selection and Economics	Undergrad.	Capstone project	100-150	50%
Fund. of Mfg. Processes	Undergrad.	Report	50-150	20%
Mfg. Process Control II	Grad.	Critique, proposal	100-150	20%
Mgmt. in Engineering: Strategy and Leadership	Grad.	Case studies	100-150	60%

fundamentals and factors that drive STEM innovation, such as the history of technology policy or the roles of key stakeholders in the innovation system. While the content is US-centric, the learner population is global. At the end of the course, learners are asked to integrate all of the pieces they have learned. Learners are asked to pick a technology they would like their government to support and craft a technology policy recommendation in the context of their own country. Using *CrowdLearn*, learners submit this policy recommendation, receive feedback on their own while discussing each other's policy recommendations. One learner commented, *It is humbling to receive feedback and read about differing thoughts on various submissions from like-minded peers. It also gave me insights that I may not have previously come across or thought about.* The trove of hundreds of policy recommendations acts as a fresh library of case studies for learners to further enrich their own learning. One learner commented: *Reading my peers' work also opened my eyes to the geographical differences in technology maturity and potential technological solutions.*

B. Data Analysis: Statistical Modeling and Computation in Applications

This course is part of the graduate-level *MITx MicroMasters program in Statistics and Data Science*. Students learn how to model, form hypotheses, and perform statistical analysis on real data. The course is anchored in four data analysis projects. In each project, students analyze data sets in the four major different domain areas focused on the course: high-dimensional, networks, time series, Gaussian processes. In Part 1 of the project, students complete assessments to receive guidance on set-up, such as choice of plotting or code packages. Because the problems are highly-scaffolded and close-ended, the assessments can be offered using traditional edX assessments such as multiple choice or checkboxes. In Part 2, instructors release a Jupyter notebook containing basic sample code. Learners use it to carry out preliminary data analysis on a simple data set and answer close-ended questions on the edX platform. In Part 3, learners are tasked with analyzing a data set that is larger or more complex than in part 2. Using *CrowdLearn*, they submit their own project report (code, analysis, and justification) and discuss each other's projects. By having access to all submitted projects, learners are given the option to learn from a range of good or bad examples of analysis on the same data set. They can mentally benchmark the quality of their own analysis or justification when they see other reports of their peers.

C. Structural Materials: Selection and Economics

In this course, learners are exposed to various aspects of materials selection, namely: fundamental material properties, technical analysis, and broader economic and societal considerations. The course's capstone project is the primary opportunity for learners to apply their learnings to select a material. Learners must research materials for an application of their own interest and choosing, then downselect and justify material choice using an analysis of material properties and economic external factors. When learners submit their report to the *CrowdLearn* platform, they have access not only to the analyses of different materials, but also diverse approaches to integrate the many aspects of material choice. One learner commented: *I loved the proposed dynamic, as it was possible to analyze the point of view of other authors on their particular materials. These spaces are essential to discuss this type of subject, promoting a productive debate for our careers.*

V. RESULTS AND DISCUSSION

In summary, we have iteratively designed a tool that offers scalable, open-ended assessments in the edX environment. The tool also gives a MOOC instructor the ability to assess potentially thousands of learner contributions to discussion. Our work suggests that the tool fulfilled the objectives to 1) handle open-ended assessments at scale, and 2) create an environment for collaborative learning through "open-source" learning and identifiable interaction. The tool has now been successfully implemented into eight MOOCs ranging from engineering, science, and social sciences.

In each of the courses, instructors include the peer review grading as a major component of the overall grade of the assessment. However, instructors of one course, *Data Analysis: Statistical Modeling and Computation in Applications*, have developed a grading protocol that is more involved. Instructors will manually grade reports that receive widely varying peer grades (max-min > 4/10) or ones that receive too few peers grades. The final grade is a combination of instructor and peer grade. For the report in question, they will also check the peer reviews themselves. If the review quality is poor, they will deduct from the project grade of that reviewer. This leads to instructors grading an estimated 15% of the total submissions, which incentivizes quality grading, but also detracts from the assessment's ability to cost-effectively scale. Because this course is part of a MicroMasters program and has credential weight, instructors in this course may have needed to adopt another level of grading rigor. Instructors from other courses, (including the two courses from another MicroMasters program) are comparatively "hands off" and only intervene when there is a learner complaint. Educators who use the *CrowdLearn* tool in the future will need to balance the tradeoffs between grading rigor and scalability.

We did not include this in our analysis, but were surprised by instructor commentary and learner reflections that referred to a sense of a strengthened community because of the dynamics within *CrowdLearn*. There is also anecdotal evidence of increased reports from learners on plagiarism by other learners. We are interested in investigating in the future whether identifiable learner interactions and the open nature of submissions in this environment is giving MOOC learners the ability, agency, and motivation to grow into a stronger and even self-policing community.

In the future, we are interested in investigating the new types of MOOC learning the *CrowdLearn* environment may enable. More methodical analysis of instructor feedback and learner reflection may help us understand how this tool enables impactful and collaborative learning. Perdue and Sandland recently showed that open, transparent learning MOOC environments lead to an improvement in learner feedback [25]. We are interested in investigating how factors such as peer citation or grading schemes can incentivize positive learner dynamics and thoughtful feedback. Future tool development may include more flexibility and granularity in grading, advanced search functionality, improved communication between the tool and the edX platform, and streamlined grading for instructors.

Acknowledgment

We would like to thank the Digital Learning Lab and staff at MITx for their feedback on the design and use of the *CrowdLearn* tool. We thank Dr. Jessica Sandland, Meghan Perdue, and Nhi Nguyen for useful discussions.

References

[1] W. Admiraal, B. Huisman, and M. Van de Ven, "Self-and peer assessment in massive open online courses," *Int. J. High. Educ.*, vol. 3, no. 3, pp. 119-128, 2014.

[2] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Engaging with massive online courses," in *Proc. 23rd Int. Conf. World Wide Web*, April 2014, pp. 687-698, http://doi.org/10.1145/2566486.2568042.

[3] K. Appiah-Kubi and D. Rowland, "Peer support in MOOCs: The role of social presence," in *Proc. 3rd ACM Conf. Learning @ Scale*, 2016, pp. 237-240, https://doi.org/10.1145/2876034.

[4] B. Chen. "Designing for networked collaborative discourse: An UnLMS approach." *TechTrends*, vol. 63, no. 2, pp. 194-201, 2019, http://doi.org/10.1007/s11528-018-0284-7.

[5] K. Cooper and H. Khosravi, "Graph-based visual topic dependency models," in *Proc. 8th Int. Conf. Learning Analytics Knowledge,* 2018, pp. 11-15, https://doi.org/10.1145/3170358.3170418.

[6] J. Díez Peláez *et al.*, "Peer assessment in MOOCs using preference learning via matrix factorization," *In NIPS Workshop on Data Driven Education*, 2013.

[7] C. Evans, "Making sense of assessment feedback in higher education," *Rev. Ed. Res.*, vol. 83, no. 1, pp. 70-120, 2013, https://doi.org/10.3102/0034654312474350.

[8] "FeedbackFruits," https://feedbackfruits.com/ (accessed April 18, 2023).

[9] D. Gamage, M.E. Whiting, T. Rajapakshe, H. Thilakarathne, I. Perera, and S. Fernando, "Improving assessment on MOOCs through peer identification and aligned incentives," in *Proc. 4th ACM Conf. Learning @ Scale*, 2017, pp. 315-318, https://doi.org/10.1145/3051457.3054013.

[10] D. Gamage, I. Perera, and S. Fernando, "Increasing interactivity and collaborativeness in MOOCs using facilitated groups: A pedagogical solution to meet 21st century goals," in *Proc. 2018 IEEE Global Eng. Ed. Conf.*, 2018, pp. 354-367, https://doi.org/10.1109/EDUCON.2018.8363324.

[11] D. Gamage, M.E. Whiting, I. Perera, and S. Femando, "Improving feedback and discussion in MOOC peer assessment using introduced peers," in *Proc. 2018 IEEE Int. Conf. Teach., Assessment, Learning Eng.* 2019, pp. 357-364, https://doi.org/10.1109/TALE.2018.8615307.

[12] D. Gamage, T. Staubitz, and M. Whiting, "Peer assessment in MOOCs: Systematic literature review," *Distance Ed.*, vol. 42, no. 2, pp. 268-289, 2021, https://doi.org/10.1080/01587919.2021.1911626.

[13] J. Jiao, Y. Yang, H. Zhong, and G. Ren. "Improving Learning in MOOCs Through Peer Feedback: How Is Learning Improved by Providing and Receiving Feedback?," in *Learning and Knowledge Analytics in Open Education*, 2017, pp. 69-87, http://doi.org/10.1007/978-3-319-38956-1_6.

[14] D.A. Joyner *et al.*, "Graders as Meta-Reviewers: Simultaneously Scaling and Improving Expert Evaluation for Large Online Classrooms," in *Proc. 3rd ACM Conf. Learning @ Scale*, April 2016, pp. 399-408, http://doi.org/10.1145/2876034.2876044.

[15] Y. Kotturi *et al.*, "Structure and messaging techniques for online peer learning systems that increase stickiness," in *Proc. 2nd ACM Conf. Learning @ Scale*, 2015, pp. 31-38, https://doi.org/10.1145/2724660.2724676.

[16] C. Kulkarni *et al.*, "Scaling short-answer grading by combining peer assessment with algorithmic scoring," in *Proc. 1st ACM Conf. Learning @ Scale*, 2014, pp. 99–108, https://doi.org/10.1145/2556325.2566238.

[17] C. Kulkarni, M.S. Bernstein, and S. Klemmer, "PeerStudio: Rapid peer feedback emphasizes revision and improves performance," in *Proc. 2nd ACM Conf. Learning @ Scale*, 2015, pp. 75-84, https://doi.org/10.1145/2724660.2724670.

[18] "10.20. LTI Component." edX.org. https://edx.readthedocs.io/projects/edx-partner-course-staff/en/latest/exercises_tools/lti_component.html (accessed July 10, 2023)

[19] O. Luaces *et al.*, "A factorization approach to evaluate open-response assignments in MOOCs using preference learning on peer assessments," *Knowledge-Based Systems*, vol. 85, pp. 322-328, 2015, http://doi.org/10.1016/j.knosys.2015.05.019.

[20] K. Lyons, "Using a design-based research approach to develop and study a web-based tool to support collaborative learning," *Comp. Ed.* vol. 161, 2021, http://doi.org/10.1016/j.compedu.2020.104064.

[21] "10.26.1. Introduction to Open Response Assessments." edX.org. https://edx.readthedocs.io/projects/edx-partner-course-staff/en/latest/exercises_tools/open_response_assessments/OpenResponseAssessments.html#self-assessment-scoring (accessed April 18, 2023)

[22] "MicroMasters Program in Principles of Manufacturing" mit.edu. https://micromasters.mit.edu/pom/ (accessed July 11, 2023).

[23] "MicroMasters® Programs" edX.org. https://www.edx.org/micromasters (accessed July 11, 2023).

[24] E. Panadero, A. Jonsson, and J. Botella, "Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses," *Ed. Res. Rev.*, vol. 22, pp. 74-98, 2017, http://doi.org/10.1016/j.edurev.2017.08.004.

[25] M. Perdue and J. Sandland, "Evaluating the impact of transparency on peer review quality in MOOCs," *2022 IEEE LWMOOCS*, 2022, pp. 101-107, https://doi.org/10.1109/LWMOOCS53067.2022.9928020.

[26] T. Staubitz *et al.*, "Improving the peer assessment experience on MOOC platforms," in *Proc. 3rd ACM Conf. Learning @ Scale*, April 2016, pp. 389-398., http://doi.org/10.1145/2876034.2876043.
[27] T. Staubitz and C. Meinel, "Team based assignments in MOOCs: Results and observations," in *Proc. 5th ACM Conf. Learning @ Scale*, June 2018, pp. 1–4, https://doi.org/10.1145/3231644.3231705.
[28] T. Staubitz and C. Meinel, "Graded team assignments in MOOCs: Effects of team composition and further factors on team dropout rates and performance, in *Proc. 6th ACM Conf. Learning @ Scale*, 2019, pp. 1-10, https://doi.org/10.1145/3330430.3333619.
[29] J. Tritz *et al.*, "Peer evaluation of student generated content," in *Proc. 4th Int. Conf. Learning Analytics Knowledge*, March 2014, pp. 277-278, https://doi.org/10.1145/2567574.2567598.
[30] W. Wu *et al.*, "Game theory based peer grading mechanisms for MOOCs," in *Proc. 2nd ACM Conf. Learning @ Scale*, 2015, pp. 281–286, https://doi.org/10.1145/2724660.2728676.
[31] X. Zhu, *et al.*, "Reading and connecting: using social annotation in online classes," *Inf. Learning Sciences,* vol. 121, no. 5-6, pp. 261-271, 2020, http://doi.org/10.1108/ILS-04-2020-0117.