Learnersourcing: Student-generated Content @ Scale

Steven Moore
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
stevenjamesmoore@gmail.com

John Stamper
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
john@stamper.org

Christopher Brooks
University of Michigan
Ann Arbor, Michigan, USA
brooksch@umich.edu

Paul Denny
The University of Auckland
Auckland, New Zealand
paul@cs.auckland.ac.nz

Hassan Khosravi
The University of Queensland
Brisbane, QLD, Australia
h.khosravi@uq.edu.au

ABSTRACT

The first annual workshop on Learnersourcing: Student-generated Content @ Scale is taking place at Learning @ Scale 2022. This hybrid workshop will expose attendees to the ample opportunities in the learnersourcing space, including instructors, researchers, learning engineers, and many other roles. We believe participants from a wide range of backgrounds and prior knowledge on learnersourcing can both benefit and contribute to this workshop, as learnersourcing draws on work from education, crowdsourcing, learning analytics, data mining, ML/NLP, and many more fields. Additionally, as the learnersourcing process involves many stakeholders (students, instructors, researchers, instructional designers, etc.), multiple viewpoints can help to inform what future student-generated content might be useful, new and better ways to assess the quality of the content and spark potential collaboration efforts between attendees. We ultimately want to show how everyone can make use of learnersourcing and have participants gain hands-on experience using existing tools, create their own learnersourcing activities using them or their own platforms, and take part in discussing the next challenges and opportunities in the learnersourcing space. Our hope is to attract attendees interested in scaling the generation of instructional and assessment content and those interested in the use of online learning platforms.

CCS CONCEPTS

•Human-centered computing ~ Collaborative and social computing ~ Collaborative and social computing theory, concepts and paradigms ~ Computer supported cooperative work •Applied computing ~ Education ~ Interactive learning environments

KEYWORDS: Learnersourcing, student-generated content, question creation, assessment, multiple-choice question, student learning

L@S'22, June 1–3, 2022, New York City, New York, USA

ACM ISBN 978-1-4503-9518-0/22/06...$15.00.
https://doi.org/10.1145/3491140.3528286

ACM Reference format:

Steven Moore, John Stamper, Christopher Brooks, Paul Denny, & Hassan Khosravi. 2022. Learnersourcing: Student-generated Content @ Scale. In *Proceedings of the Ninth ACM Conference on Learning @ Scale (L@S, '22), June 1–3, 2022, New York City, NY, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3491140.3528286

1 Background

Involving students in the process of creating questions, hints, examples, and other instructional content has been shown to benefit student learning [1]. This is known as a form of *learnersourcing*, where students complete activities in online courses that produce content that can be leveraged by future learners [2]. Learnersourcing offers a domain agnostic way to help scale the creation of high-quality assessments, while also helping students learn the course content. Involving students in this process, instead of crowdworkers or other accessible parties, is essential due to the domain knowledge students are often required to possess by being enrolled in a particular course. It also offers the unique perspective of someone more novice that might have unique insights due to actively learning the material, overcoming potential expert blind spots that instructors typically have when creating assessments [3].

Several learnersourcing systems have been developed to support student participation in a variety of activities, including content review and problem explanation [4]. Two popular learnersourcing systems, PeerWise [5] and RiPPLE [6], were developed to assist students in the process of generating and reviewing multiple-choice questions (MCQs). Using these two tools, tens of thousands of students have generated and reviewed millions of MCQs. This includes the students' providing assessments of the question quality and feedback on how to potentially improve them for the authors and others to utilize. These tools are domain agnostic and can be incorporated into any existing course, allowing for ample research and educational opportunities. Previous research has shown that the use of these tools enables students to generate high-quality assessments that can be leveraged by the students as practice opportunities, all while being effective for their learning [7,8,9]. Learnersourcing activities can also be directly implemented into existing courseware, such as MOOC platforms like Coursera or edX, learning management systems like Canvas

or Blackboard, or other online courseware such as the Open Learning Initiative [10], without the use of external tools. Recent learnersourcing research has demonstrated that students can effectively generate quality questions, even when the learnersourcing activities have limited scaffolding and student participation with them is optional [11,12].

The past five years of L@S has included papers on learnersourcing, as the field continues to grow in popularity, and new techniques are developed to improve the quality of student-generated content. We want to share this concept of learnersourcing with the broader audience of the conference, to show them how they can leverage these existing tools or create activities within their own courses and research platforms to enable students to generate such content. These systems and learnersourcing activities, along with several datasets collected from their use by thousands of students, are available for instructors and researchers to use [13]. Learnersourcing provides invaluable data that can be used to generate question banks, help answer questions related to student learning, and provides ample opportunities for machine learning and natural language processing work.

The main goal of this workshop is to expose attendees to the ample opportunities in the learnersourcing space, including instructors, researchers, learning engineers, and many other roles. We believe participants from a wide range of backgrounds and prior knowledge on learnersourcing can both benefit and contribute to this workshop. As the learnersourcing process involves many stakeholders (students, instructors, researchers, instructional designers, etc.), multiple viewpoints can help to inform what future student-generated content might be useful, new and better ways to assess the quality of the content, and spark potential collaboration efforts between attendees. We ultimately want to show how everyone can make use of learnersourcing and have participants gain hands on experience using these tools, creating their own learnersourcing activities using them or their own platforms, and take part in an exploration of the learnersourcing data we share out. Our hope is to attract attendees interested in scaling the generation of instructional and assessment content and those interested in the use of online learning platforms.

2 Challenges and Opportunities

Learnersourcing presents several challenges when it comes to the process of creating, evaluating, and utilizing the student-generated content. Participation rates with learnersourcing activities can be low when they are not required and even then, a majority of the contributions may be made by a smaller portion of the students [10,1, 14]. To elicit higher quality contributions from students, previous work has also investigated training the students to both create effective MCQs and properly evaluate them from a pedagogical standpoint [15]. The evaluation process of having students review and revise other student-generated questions presents a challenge regarding how we can assist students in optimally acting on the provided feedback. It is an open question on how we can incorporate student evaluation of the materials into the learning process, such as incorporating it into learner models used to power learning analytics [16]. While research indicates the learning benefits of students generating questions, oftentimes the quality of the student-generated questions can be improved. Recent work demonstrated that MCQs authored by students performed as well as those authored by academics, but further work remains to investigate how we might assist students in making consistent high-quality learnersourced contributions [17].

Among these challenges with learnersourcing lie many opportunities to answer these outstanding questions and advance the field, making it more accessible and beneficial to student learning. A clear opportunity regarding the creation of student-generated content is the different ways we can encourage students to make a high-quality contribution, such as leveraging self-regulated learning interventions [18]. While much of the existing learnersourcing research involves the creation of MCQs, there are limitless activity types that can be created and evaluated using learnersourcing techniques. Previous work has leveraged natural language processing [19,20,21], trust-based networks [22], and deep learning methods [23] to assist students in the evaluation of learnersourced content. While human input remains critical in this evaluation process, future work may look at using artificial intelligence to further support students as they evaluate learnersourcing content [24]. Finally, the use of learnersourcing has led to the creation of millions of questions across a plethora of educational domains. These questions can be leveraged in low- and high-stakes environments, to personalize learning, and provide students with more practice opportunities. However, there remains an opportunity to improve the sharing of these questions so that they can be used by students and instructors alike to support the learning process.

3 PeerWise

PeerWise is a system developed in 2008 that allows students to create and share formative practice questions in a convenient manner [25, 26]. It was one of the first learnersourcing systems developed and has since been used across thousands of courses and hundreds of thousands of students to create and evaluate MCQs. Within the system, students take part in the creation, sharing, and answering of MCQs as a supplement to their other instructional materials. When a student answers a question in PeerWise, they are presented with the opportunity to rate the question based on its difficulty and quality, in addition to providing general comments about it. These ratings and comments facilitate identification of the most useful and pedagogically sound items, while flagging those that are incorrect or ambiguous. A key goal of PeerWise is to guide subsequent student learning by revealing areas of student knowledge that are incomplete and providing ample practice opportunities.

4 RiPPLE

RiPPLE is an online adaptive learning platform that employs learner-centered and pedagogically supported approaches to engage students in authentic learning experiences [6]. The platform further aims to harness the creativity and evaluation

power of students as experts-in-training to develop a repository of high-quality learning resources. RiPPLE takes the learnersourcing approach of partnering with students to create a repository of learning resources. In the current version of the platform, students have the ability to contribute different types of resources including multiple-choice questions, multi-answer questions, matching questions, worked examples, and open-ended notes. Unique to the RiPPLE platform is the ability to allow for students and staff to be the joint co-creators of any given subject. Additionally, RiPPLE utilizes an evaluation process that allows students to review other student-generated content using a set of rubric criteria that can be customized depending on the course and resource type [22,24].

5 Pre-Workshop Plans

Prior to the workshop, we will provide access to the two core systems being discussed at the beginning of the session. Participants will be able to investigate the activities and affordances each system offers, the review process for questions, and much more to both learn about and even participate in learnersourcing themselves. Our hope is that during the workshop, they can then ask any questions and gain a better understanding of the types of learnersourcing data, if they wish to do so, before we meet. Finally, we will post a brief survey to collect the backgrounds and interests of the participants to help tailor our discussions and activities.

6 Workshop Structure

This will be a half-day workshop. The workshop focus will be on examining the tools, processes, and content that is both used and generated through *learnersourcing*. We will begin with introductions and an overview of the learnersourcing landscape, to bring all participants, regardless of background, up to speed on the concept and latest trends. Two presentations and demos will then be run to highlight different learnersourcing tools, with an emphasis on how the student-generated content can be used by instructors and researchers. We will then have participant presentations, where accepted submissions will be presented for roughly five minutes each. Following that, we will have a ten-minute break that will include coffee and light snacks. From there, we will then demonstrate how participants can add learnersourcing activities of their own to practically any piece of educational technology (MOOCs, LMSs, etc.). Participants will then engage in a discussion around the challenges, opportunities, and future of learnersourcing, including how we can incentivize quality student-generated content, while also empowering the instructors and learners with insights. The workshop will conclude with a summary of the day's events, core challenges and opportunities we addressed in the discussions, and an emphasis on future collaborations.

7 Post-Workshop Plans

We will publish the accepted one-page papers as part of a compendium, to expand the work on learnersourcing. Additionally, we plan for the interactions during the workshop to result in the adoption of learnersourcing for many of the participants, whether that be using one of the tools, the discussed datasets, or creating learnersourcing activities in their own platforms and courses. Upon completion of the workshop, we will offer participants the chance to join a Slack channel and mailing list dedicated to sharing out advances in learnersourcing. Through these channels, we will continue to share datasets collected from these and other learnersourcing systems. We envision that these datasets can be leveraged by participants for future studies and potentially be the focus of a future workshop or competition at L@S. Ultimately, we want to keep the participants involved and promote collaboration between attendees. We expect this workshop to be repeated and become part of the basis for a community of researchers who are interested in learnersourcing and assessment generation at scale.

8 Call for Participation

In this workshop, we will discuss two of the most popular learnersourcing tools that can be readily utilized in any course. We will also work hands on with participants to create activities in their own systems/courses and begin exploring several learnersourcing datasets made available to attendees. This workshop will bring together participants to discuss new learnersourcing activity types, methods for assessing the quality of student-generated content, and ways we might better scale learnersourcing efforts to create usable instructional and assessment materials. Participants of all backgrounds are welcome, as learnersourcing involves many stakeholders including students, instructors, researchers, practitioners, and many more!

While no submission is required to participate in the workshop, we encourage submissions of various types as stated above. We expanded our submissions to include artifacts such as videos and commentary to express your perspectives on learnersourcing. However, the core submission format is a research, work-in-progression, or position paper, targeting roughly 1 to 2 pages.

- Here are some questions and ideas applicants may want to considering addressing in their submissions:
- Incentivizing student participation with learnersourcing activities
- Exploring novel formats of learnersourcing content
- Assessing student-generated content
- Incentivizing high-quality student contributions
- Providing actionable and explainable insights to students and teachers
- Supporting multi-institutional sharing and collaboration with learnersourced content
- Training students to develop high-quality resources
- Encouraging student participation and engagement with learnersourcing

9 Organizers

The workshop will be organized by a team of learnersourcing researchers and practitioners, all who are involved with the development of a learnersourcing systems or data repositories. Additionally, they all have prior experience conducting learnersourcing research and past workshop organization.

Steven Moore is a PhD student in Human-Computer Interaction at Carnegie Mellon University and is advised by Dr. John Stamper. His research is focused on engaging students in the learnersourcing process and finding ways to assess the quality of their contributions.

John Stamper is an Associate Professor at the Human-Computer Interaction Institute at Carnegie Mellon University and the Technical Director of the Pittsburgh Science of Learning Center DataShop. His work involves leveraging educational data mining techniques and the creation of data tools that can be used with learnersourcing data.

Christopher Brooks is an Assistant Professor at the University of Michigan and is an applied Computer Scientist who builds and studies the effects of educational technologies in higher education and informal learning environments. He has led learnersourcing efforts on the Coursera platform, where he investigated student choice in the generation of multiple-choice questions.

Paul Denny is an Associate Professor in Computer Science at the University of Auckland, New Zealand. He leads the PeerWise project, which hosts more than six million practice questions, with associated solutions and explanations, created by students from 90 countries.

Hassan Khosravi is an Associate Professor in the Institute for Teaching and Learning Innovation and an Affiliate Academic in the School of Information Technology and Electrical Engineering at the University of Queensland. He has conducted extensive learnersourcing research and leads the development and dissemination efforts of the RiPPLE system.

REFERENCES

[1] Khosravi, H., Demartini, G., Sadiq, S., & Gasevic, D. (2021, April). Charting the design and analytics agenda of learnersourcing systems. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 32-42).

[2] Kim, J. (2015, May). Learnersourcing: improving learning with collective learner activity. PhD Dissertation. Massachusetts Institute of Technology

[3] Nathan, M. J., Koedinger, K. R., & Alibali, M. W. (2001, April). Expert blind spot: When content knowledge eclipses pedagogical content knowledge. In *Proceedings of the third international conference on cognitive science* (Vol. 644648).

[4] Williams, J. J., Kim, J., Rafferty, A., Maldonado, S., Gajos, K. Z., Lasecki, W. S., & Heffernan, N. (2016, April). Axis: Generating explanations at scale with learnersourcing and machine learning. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale* (pp. 379-388).

[5] Denny, P., Hamer, J., Luxton-Reilly, A., & Purchase, H. (2008, September). PeerWise: students sharing their multiple-choice questions. In *Proceedings of the fourth international workshop on computing education research* (pp. 51-58).

[6] Khosravi, H., Kitto, K., & Williams, J. J. (2019). RiPPLE: A Crowdsourced Adaptive Platform for Recommendation of Learning Activities. *Journal of Learning Analytics*, *6*(3), 91–105.

[7] Duret, D., Christley, R., Denny, P., & Senior, A. (2018). Collaborative learning with PeerWise. *Research in Learning Technology*, *26*, 1-13.

[8] Abdi, S., Khosravi, H., Sadiq, S., & Demartini, G. (2021). Evaluating the quality of learning resources: A learnersourcing approach. IEEE Transactions on Learning Technologies, 14(1), 81-92.

[9] Dale Hancock, Nicole Hare, Paul Denny and Gareth Denyer. 2018. Improving large class performance and engagement through student-generated question banks. Biochemistry and Molecular Biology Education, 46(4), pages 306-317.

[10] Bier, N., Moore, S., & Van Velsen, M. (2019). Instrumenting courseware and leveraging data with the Open Learning Initiative (OLI). In *Companion Proceedings 9th International Learning Analytics & Knowledge Conference, Tempe, AZ.*

[11] Moore, S., Nguyen, H. A., & Stamper, J. (2021, June). Examining the Effects of Student Participation and Performance on the Quality of Learnersourcing Multiple-Choice Questions. In *Proceedings of the Eighth ACM Conference on Learning@ Scale* (pp. 209-220).

[12] Singh, A., Brooks, C., Lin, Y., & Li, W. (2021, June). What's In It for the Learners? Evidence from a Randomized Field Experiment on Learnersourcing Questions in a MOOC. In *Proceedings of the Eighth ACM Conference on Learning@ Scale* (pp. 221-233).

[13] Wang, Z., Lamb, A., Saveliev, E., Cameron, P., Zaykov, J., Hernandez-Lobato, J. M., ... & Zhang, C. (2021, August). Results and Insights from Diagnostic Questions: The NeurIPS 2020 Education Challenge. In NeurIPS 2020 Competition and Demonstration Track (pp. 191-205). PMLR.

[14] Denny, P., McDonald, F., Empson, R., Kelly, P., & Petersen, A. (2018, April). Empirical support for a causal relationship between gamification and learning outcomes. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1-13).

[15] Denny, P. (2015, February). Generating practice questions as a preparation strategy for introductory programming exams. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education* (pp. 278-283).

[16] Abdi, S., Khosravi, H., & Sadiq, S. (2020, July). Modelling learners in crowdsourcing educational systems. In *International Conference on Artificial Intelligence in Education* (pp. 3-9). Springer, Cham.

[17] Huang, A., Hancock, D., Clemson, M., Yeo, G., Harney, D., Denny, P., & Denyer, G. (2021). Selecting student-authored questions for summative assessments. *Research in Learning Technology, 29.*

[18] Lahza, H., Khosravi, H., Demartini, G., & Gasevic, D. (2022, March). Effects of Technological Interventions for Self-regulation: A Control Experiment in Learnersourcing. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 542-548).

[19] Moore, S., Nguyen, H. A., & Stamper, J. (2020, July). Evaluating Crowdsourcing and Topic Modeling in Generating Knowledge Components from Explanations. In *International Conference on Artificial Intelligence in Education* (pp. 398-410). Springer, Cham.

[20] Moore, S., Nguyen, H., & Stamper, J. (2020). Evaluating Crowdsourcing and Topic Modeling in Generating Knowledge Components from Explanations. *Artificial Intelligence in Education, 12163*, 398 - 410.

[21] Darvishi A, Khosravi K, Abdi S, Sadiq S, and Gašević D. (2022) Incorporating Training, Self-monitoring and AI-Assistance to Improve Peer Feedback Quality. In Proceedings of the Ninth ACM Conference on Learning @ Scale

[22] Darvishi, A., Khosravi, H., & Sadiq, S. (2021, June). Employing peer review to evaluate the quality of student generated content at scale: A trust propagation approach. In Proceedings of the Eighth ACM Conference on Learning@ Scale (pp. 139-150).

[23] Ni, L., Bao, Q., Li, X., Qi, Q., Denny, P., Warren, J., Witbrock, M., Liu, J. (2022). DeepQR: Neural-based Quality Ratings for Learnersourced Multiple-Choice Questions. *The 12th Symposium on Educational Advances in Artificial Intelligence* (pp. 1–9).

[24] Gyamfi, G., Hanna, B., & Khosravi, H. (2021). Supporting peer evaluation of student-generated content: a study of three approaches. *Assessment & Evaluation in Higher Education*, 1-19.

[25] Denny, P., Hamer, J., Luxton-Reilly, A., & Purchase, H. (2008, September). PeerWise: students sharing their multiple choice questions. In *Proceedings of the fourth international workshop on computing education research* (pp. 51-58).

[26] Denny, P., Luxton-Reilly, A., & Hamer, J. (2008, January). The PeerWise system of student contributed assessment questions. In *Proceedings of the tenth conference on Australasian computing education-Volume 78* (pp. 69-74).