Learnersourcing: Student-generated Content @ Scale

2nd Annual Workshop

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

Anjali Singh
University of Michigan
Ann Arbor, Michigan, USA
singhanj@umich.edu

Xinyi Lu
University of Michigan
Ann Arbor, Michigan, USA
lwlxy@umich.edu

Hyoungwook Jin
School of Computing, KAIST
Daejeon, South Korea
jinhw@kaist.ac.kr

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

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

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

Xu Wang
University of Michigan
Ann Arbor, Michigan, USA
xwanghci@umich.edu

Juho Kim
School of Computing, KAIST
Daejeon, South Korea
juhokim@kaist.ac.kr

John Stamper
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
jstamper@cs.cmu.edu

ABSTRACT

We propose the second annual workshop on Learnersourcing: Student-generated Content @ Scale. This full-day workshop is designed to explore the vast potential of learnersourcing, which combines the efforts of humans, AI, and other data sources to create and assess educational materials. This presents an opportunity for instructors, researchers, learning engineers, and professionals from various fields to discover how learnersourcing can enhance education. This workshop is open to individuals with diverse backgrounds and levels of experience with learnersourcing. By drawing on principles from education, crowdsourcing, learning analytics, data mining, machine learning (ML), and natural language processing (NLP), we aim to foster an environment where all participants can learn from and contribute to the conversation. Learnersourcing involves a wide range of stakeholders, including students, instructors, researchers, and instructional designers. Through bringing together these different perspectives, we hope to uncover what future learnersourced content could look like, identify innovative methods to evaluate its quality, and ignite collaborative projects among participants. Our goal is for attendees to leave the workshop with a practical understanding of how to engage with learnersourcing. Participants will get hands-on experience with current tools, develop their own learnersourcing activities, and join discussions about the future challenges and opportunities in this field.

L@S '24, July 18–20, 2024, Atlanta, GA, USA

ACM 979-8-4007-0633-2/24/07.
https://doi.org/10.1145/3657604.3664643

CCS CONCEPTS

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

KEYWORDS

Learnersourcing; Student-generated content; Question creation; Educational content creation

ACM Reference format:

Steven Moore, Anjali Singh, Xinyi Lu, Hyoungwook Jin, Hassan Khosravi, Paul Denny, Christopher Brooks, Xu Wang, Juho Kim, and John Stamper. 2024. Learnersourcing: Student-generated Content @ Scale: 2nd Annual Workshop. In *Proceedings of Eleventh ACM Conference on Learning @ Scale (L@S, '24), July 18–20, 2024, Atlanta, GA, USA.* ACM, NY, NY, USA. 4 pages. https://doi.org/10.1145/3657604.3664643

1 BACKGROUND

Learnersourcing involves students in generating and evaluating educational content such as questions, hints, and examples [7]. Previous research has demonstrated that this enhances student

learning, by actively engaging them with the material and invoking critical thinking [15]. This approach utilizes such student-generated content in online courses for the benefit of future learners, offering a scalable method to create high-quality assessments while deepening students' understanding of the material. The unique insights from students, who may spot nuances overlooked by experts, are invaluable due to their direct engagement with the course content. Systems such as PeerWise [2], RiPPLE [5], Upgrade [16], and ReadingQuizMaker [9] have utilized learnersourcing by enabling students to create, assess, and use peer-generated educational materials. This approach has expanded beyond just crafting new questions; it now includes the creation of various forms of assessments, like programming exercises and instructional videos [4, 8, 13].

The interest in learnersourcing at the Learning @ Scale conference has primarily grown over the past eight years, showcasing the development of new techniques to enhance the quality of student-generated content. The advent of large language models (LLMs) offers fresh opportunities for this field and simplifies previous challenges in learnersourcing. This workshop aims to introduce the concept of learnersourcing to a wider audience, demonstrating how to utilize or develop activities that empower students to generate and evaluate educational content. We will share tools, activities, and datasets from thousands of students to help instructors and researchers leverage learnersourcing interventions to improve student learning and enhance courseware. Such initiatives provide rich data for creating question banks, understanding student learning, and supporting machine learning and natural language processing research. Our goal is for participants to discover the potential of learnersourcing, gain practical experience with related tools and activities, and explore the shared learnersourcing data. We aim to attract those interested in expanding the creation of educational content and using online learning platforms.

2 OPPORTUNITIES AND CHALLENGES

Learnersourcing presents a wealth of opportunities to enhance student learning and advance the field through innovative approaches for the creation, evaluation, utilization, and oversight of educational content [6]. These four pillars of the learnersourcing process can be seen in Figure 1. One such opportunity lies in facilitation of high-quality student-generated content, utilizing the recent advances of LLMs like GPT-4 to assist students in the development of educational materials [3]. These advances help overcome previous challenges, such as leveraging LLMs to address the cold start problem in learnersourcing. Although much of the current research focuses on generating multiple-choice questions (MCQs), the potential for creating diverse types of activities is vast. For example, students can collaborate with LLMs to develop and refine assessment questions and learning content explanations, engaging in critical thinking as they assess the model's suggestions. This collaboration not only speeds up content improvement, but also ensures students are actively involved in the learning process.

Previous work has explored using NLP, trust-based networks, and deep learning to assist students in the evaluation of educational content, highlighting the critical role of human input alongside AI [1]. As learnersourcing has already led to the creation of millions of questions across various educational domains, enhancing the sharing and utilization of this content can further support personalized learning and offer extensive practice opportunities. Balancing the innovative use of LLMs with careful human oversight promises to improve learnersourcing techniques and outcomes, making educational content more accessible, diverse, and effective for learners.

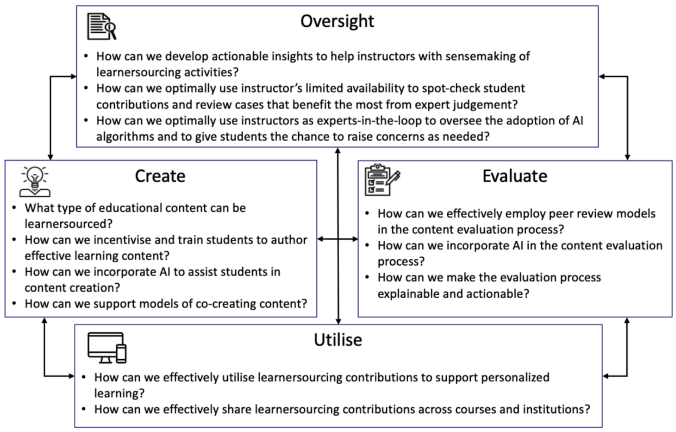

Figure 1: Challenges and opportunities relating to the four key aspects of learnersourcing that involve students, instructors, and AI.

Learnersourcing faces several challenges in the four facets of creating, evaluating, utilizing, and overseeing student-generated content. One significant hurdle is low participation rates in learnersourcing activities, especially when they are optional, with a tendency for a small portion of students to contribute the majority of the content [10]. Additionally, the process of having students and AI review and revise peer-generated questions introduces another layer of complexity. This AI integration into the learnersourcing process introduces additional biases inherent in LLMs. The risk of perpetuating these biases in educational content underscores the importance of a human-in-the-loop approach, where both students and instructors play a crucial role in content evaluation and moderation. Recent findings suggest that student-authored MCQs can match the quality of those created by experts, highlighting the potential of learnersourcing [14]. Yet, the challenge of ensuring consistent high-quality contributions from all students across any domain persists.

3 PRE-WORKSHOP PLANS

Before the workshop, we will grant participants early access to various learnersourcing systems created by our organizing

committee and provide a curated list of significant learnersourcing research papers on the workshop's website. This pre-workshop access is designed to help participants understand the functionalities and evaluation methods of each system and engage with the learnersourcing process. While optional, we wanted to make these resources available, particularly for participants that might not be familiar with this field. Additionally, we will distribute a short survey to gather details about participants' backgrounds and interests. The feedback from this survey will enable us to tailor the workshop content and activities to better match the preferences and needs of our attendees, aiming to create an engaging and relevant experience for all, regardless of prior experience in learnersourcing. Participants are not required to have a laptop, but having a mobile device with internet access will allow them to participate and ask questions asynchronously during the workshop.

4 WORKSHOP STRUCTURE

The inaugural version of our workshop took place at the 2022 Learning @ Scale conference as a half-day event, drawing over 30 participants and featuring 6 accepted paper presentations [11]. Encouraged by this initial success and having since expanded our organization team and community engagement, we anticipate closer to 50 participants in attendance for this year's edition. In response to the positive feedback, we are extending the workshop to a full-day format. This year's workshop will include a mix of brief presentations, interactive activities, and round-table discussions focused on learnersourcing. While we have outlined a provisional schedule below, please note that certain sessions, such as invited speaker talks and paper presentations, will be interspersed throughout the day to maintain a dynamic and engaging agenda:

- **Opening remarks**: Introductions of workshop organizers and participants, and a background to the focus of the workshop.
- **Invited Guest Speakers**: Two leading researchers in the field, not on the organizing committee, will give a presentation throughout the day regarding the use of AI in the learnersourcing process.
- **Short Presentations**: Authors of accepted submissions present their work which would be followed by a Q&A session.
- **Round-table discussion**: Participants will rotate among discussions on topics related to learnersourcing, focusing on creation, evaluation, use, and management of content.
- **Hands-on activity**: Participants will collaborate to brainstorm and develop new learnersourcing activities.
- **Open discussion**: A facilitator will guide a collective discussion among all participants, summarizing insights from the round-table discussions, building consensus through the collaborative creation of shared notes.
- **Concluding remarks**: Closing remarks on the workshop will be made with future steps and opportunities for collaboration between participants, to maintain community engagement.

The primary aim of this workshop is to investigate the collaborative potential of students, instructors, and AI in generating, evaluating, and utilizing educational content through learnersourcing. We welcome participants from diverse fields, not limited to the usual learning sciences, machine learning, natural language processing, and learning analytics, believing that a variety of backgrounds enriches our exploration. With learnersourcing being a multifaceted domain involving students, instructors, researchers, and instructional designers, diverse perspectives are crucial for identifying valuable future content, improving content quality assessment, and fostering collaboration among attendees.

5 POST-WORKSHOP PLANS

After the workshop, we will publish the accepted submissions via a CEUR workshop proceedings, which was done from the first workshop iteration [12]. We hope the workshop encourages participants to adopt learnersourcing tools, use the discussed datasets, or start their own learnersourcing projects. Participants will be invited to join a dedicated Slack channel and mailing list, which has also been maintained since the first workshop iteration, providing continuous access to learnersourcing advancements and datasets. These resources are expected to support participants' future research and potentially inspire future Learning @ Scale workshops or competitions. We want to continue to cultivate a community of active collaboration and engagement in learnersourcing.

6 CALL FOR PARTICIPATION

Join our full-day workshop on learnersourcing! Featuring invited speakers, interactive activities, paper presentations, and discussions, we will delve into the field's opportunities and challenges. Attendees will engage in hands-on development of learnersourcing activities suited to their own courses or systems and gain access to various learnersourcing systems and datasets for exploration.

This workshop aims to foster discussions on new types of learnersourcing activities, strategies for evaluating the quality of student-generated content, the integration of LLMs with the field, and approaches to scaling learnersourcing to produce valuable instructional and assessment materials. We welcome participants from all backgrounds, as learnersourcing benefits from the diverse perspectives of students, instructors, researchers, and practitioners.

While no submission is required to participate in this workshop, we encourage attendees to share their insights and experiences with learnersourcing through 4-6 page submissions, including research papers, work-in-progress reports, and position papers. Submissions might cover topics such as:

- Strategies for engaging and motivating student participation in learnersourcing activities

- Exploration of innovative learnersourcing content formats
- Methods for evaluating the quality of student-generated content
- Ways to encourage high-quality student contributions
- Approaches to enable collaboration and content sharing across institutions
- How LLMs can assist in the cold start problem for content creation
- Leveraging LLMs to assist in the different stages (creation, evaluation, etc.) of the learnersourcing process

This workshop is an opportunity to contribute to the evolving field of learnersourcing, share experiences, and collaborate with peers dedicated to enhancing education through learnersourcing.

7 ORGANIZERS

Steven Moore is a PhD candidate 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.

Anjali Singh is a PhD candidate at the University of Michigan and is advised by Dr. Christopher Brooks and Dr. Xu Wang. Her research focuses on the use of human and machine intelligence to improve Data Science education, in both formal and informal learning settings like MOOCs.

Xinyi Lu is a PhD student at the University of Michigan advised by Dr. Xu Wang. Her research focuses on learning about how instructors perceive and interact with AI tools and building effective Human-AI tools to elaborate AI and CS techniques into teaching practice.

Hyoungwook Jin is an MS candidate in the School of Computing at KAIST, advised by Dr. Juho Kim. His research supports personalized learning environments at scale, leveraging AI agents as tutors and tutees to build personalized and adaptive interactive learning systems.

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.

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.

Christopher Brooks is an Associate 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 Coursera, where he investigated student choice in the generation of multiple-choice questions.

Xu Wang is an Assistant Professor in Computer Science and Engineering and the School of Information at the University of Michigan. She develops human-AI collaborative techniques to support education, such as empowering educators to create effective learning experiences.

Juho Kim is an Associate Professor in the School of Computing at KAIST. His research focuses on building interactive and intelligent systems that support interaction at scale, aiming to improve the ways people learn, collaborate, discuss, make decisions, and take action online.

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

REFERENCES

[1] Solmaz Abdi, Hassan Khosravi, Shazia Sadiq, and Gianluca Demartini. 2021. Evaluating the quality of learning resources: A learnersourcing approach. IEEE Transactions on Learning Technologies 14, 1: 81–92.

[2] Paul Denny, John Hamer, Andrew Luxton-Reilly, and Helen Purchase. 2008. PeerWise: students sharing their multiple choice questions. In Proceedings of the Fourth international Workshop on Computing Education Research (ICER '08), 51–58. https://doi.org/10.1145/1404520.1404526

[3] Paul Denny, Sami Sarsa, Arto Hellas, and Juho Leinonen. 2022. Robosourcing Educational Resources -- Leveraging Large Language Models for Learnersourcing. Retrieved March 1, 2024 from http://arxiv.org/abs/2211.04715

[4] Hyoungwook Jin and Juho Kim. 2023. CodeTree: A System for Learnersourcing Subgoal Hierarchies in Code Examples. Retrieved March 1, 2024 from https://jhw123.github.io/files/CSCW2024%20CodeTree.pdf

[5] Hassan Khosravi, Gianluca Demartini, Shazia Sadiq, and Dragan Gasevic. 2021. Charting the design and analytics agenda of learnersourcing systems. In LAK21: 11th International Learning Analytics and Knowledge Conference

[6] Hassan Khosravi, Paul Denny, Steven Moore, and John Stamper. 2023. Learnersourcing in the age of AI: Student, educator and machine partnerships for content creation. Computers and Education: Artificial Intelligence: 100151.

[7] Juho Kim. 2015. Learnersourcing: improving learning with collective learner activity. Massachusetts Institute of Technology.

[8] Juho Kim, Robert C. Miller, and Krzysztof Z. Gajos. 2013. Learnersourcing subgoal labeling to support learning from how-to videos. In CHI'13 Extended Abstracts on Human Factors in Computing Systems, 685–690.

[9] Xinyi Lu, Simin Fan, Jessica Houghton, Lu Wang, and Xu Wang. 2023. ReadingQuizMaker: A Human-NLP Collaborative System that Supports Instructors to Design High-Quality Reading Quiz Questions. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1–18. https://doi.org/10.1145/3544548.3580957

[10] Steven Moore, Huy Anh Nguyen, and John Stamper. 2021. 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, 209–220.

[11] Steven Moore, John Stamper, Christopher Brooks, Paul Denny, and Hassan Khosravi. 2022. Learnersourcing: Student-generated Content@ Scale. In Proceedings of the Ninth ACM Conference on Learning@ Scale, 259–262.

[12] Steven Moore, John Stamper, Paul Denny, Hassan Khosravi, and Christopher Brooks. editors. Proceedings of the Workshop on Learnersourcing: Student-Generated Content @ Scale 2022. CEUR Workshop Proceedings, 2022, Vol. 3410. https://ceur-ws.org/Vol-3410/.

[13] Nea Pirttinen, Arto Hellas, and Juho Leinonen. 2023. Experiences from Learnersourcing SQL Exercises: Do They Cover Course Topics and Do Students Use Them? In Proceedings of the 25th Australasian Computing Education Conference, 123–131. https://doi.org/10.1145/3576123.3576137

[14] Anjali Singh, Christopher Brooks, and Shayan Doroudi. 2022. Learnersourcing in Theory and Practice: Synthesizing the Literature and Charting the Future. In Proceedings of the Ninth ACM Conference on Learning@ Scale, 234–245.

[15] Anjali Singh, Christopher Brooks, Yiwen Lin, and Warren Li. 2021. 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, 221–233.

[16] Xu Wang, Srinivasa Teja Talluri, Carolyn Rose, and Kenneth Koedinger. 2019. UpGrade: Sourcing student open-ended solutions to create scalable learning opportunities. In Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale, 1–10.