# Connecting Collaborative & Crowd Work with Online Education

**Joseph Jay Williams**
Harvard University, HarvardX
125 Mt. Auburn St Rm. 422
Cambridge, MA 02139 USA
joseph_jay_williams@harvard.edu

**Markus Krause**
Liebniz University
Appelstrasse 9A
30167 Hannover Germany
markus@hci.uni-hannover.de

**Praveen Paritosh**
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043 USA
pkparitosh@gmail.com

**Jacob Whitehill**
Harvard University, HarvardX
125 Mt. Auburn St Rm. 422
Cambridge, MA 02139 USA
jacob_whitehill@harvard.edu

**Justin Reich**
Harvard University, HarvardX
21 Hawthorne Ave
Arlington, MA 02476 USA
justin_reich@harvard.edu

**Juho Kim**
MIT
32 Vassar Street, 32-G707
Cambridge, MA 02139  USA
juhokim@mit.edu

**Piotr Mitros**
edX
141 Portland St., 9th Floor
Cambridge, MA 02139 USA
piotr@mitros.org

**Neil Heffernan**
Worcester Polytechnic Institute
Computer Science Department
100 Institute Road,
Worcester MA. 01609 USA
nth@wpi.edu

**Brian C. Keegan**
Harvard Business School
Soldiers Field Road
Boston, MA 02163 USA
bkeegan@hbs.edu

## Abstract

Human behavior increasingly involves digital online software, where the activities and resources that support (1) *learning*, (2) *work*, and (3) *collaboration* overlap and are placed in far greater proximity than the physical world – often just a browser-tab or window away. What scientific and practical gains in 21st century *learning*, *work*, and *collaboration* can be achieved by integrating and contrasting these three areas' relevant technologies, scientific communities, and industry practitioners?

For example: How can software for collaborative work incorporate learning? Which methods are effective for coordinating diverse experts to iteratively improve online educational resources? How can online learning improve the skill set and labor force for crowd work? What kinds of computational frameworks exist to jointly optimize the *learning* of skills and the *use* of these skills to achieve practical goals?

This workshop tackles such questions by bringing together participants from industry (e.g., platforms similar to oDesk, Amazon Mechanical Turk); education, psychology, and MOOCs (e.g., attendees of AERA, EDM, AIED, Learning at Scale); crowdsourcing and collaborative work (e.g., attendees of CHI, CSCW, NIPS, AAAI's HCOMP).

**Motivation**
We describe the trends and
opportunities that make this
the right time for such a workshop.

## Crowdsourcing and Online education are tackling similar problems at scale

The online education and crowdsourcing communities are addressing similar problems in educating, motivating, and evaluating students and workers. The online learning community succeeds in increasing the supply side of the cognitively skilled labor market, and the crowdsourcing at scale community creates a larger marketplace for cognitively skilled work [11].

Linking online platforms for crowd work with platforms for Massive Open Online Courses (MOOCs) has the potential to provide knowledge and training at a massive scale to contributors; collect data that identify expert skills; engage contributors in simultaneously working and learning in a social environment; and organize large communities around online courses on specific topics. These all provide new opportunities to support and deploy sophisticated algorithms for crowd learning and work.

## Increasing complexity of crowdsourcing tasks could be met by innovations for learning at scale

As tasks continue to grow in scale and complexity, as simple tasks become increasingly automated. For example, deep convolutional neural networks trained on massive datasets collected by the reCAPTCHA project have achieved 99.8% accuracy [1]. As machine learning algorithms continue to prove, the emphasis on crowdsourcing may increasingly shift toward more cognitively demanding tasks. Three major challenges were identified at last year's Crowdsourcing at Scale workshop: (1) training contributors for complex tasks, (2) defining and detecting experts, and (3) improving the contributor experience.

Such challenges could in part be addressed by recent theoretical and empirical progress in learning at scale, one form of which has been Massive Open Online Courses (MOOCs). Presenters at this workshop will report on recent work, as reflected in the proceedings of the first ACM Learning at Scale 2014 conference, the MOOCshop workshop at the International Conference of the Learning Sciences, and the Learning Innovation at Scale workshop at CHI [13].

In order to address more complex tasks and expand the active market for cognitively skilled labor with higher value and pay, we need systems for developing materials and technologies to educate and assess contributors [8]. Learning at scale can help address this bottleneck and provide systematic rather than one-off or ad hoc solutions. Hundreds of MOOCs already exist, and many more are in production, on topics from programming to machine learning. These provide broadly accessible educational materials, can include assessments of learning, and bring together an online community or crowd around a topic. The many MOOC participants who already have knowledge of a topic are themselves an invaluable crowd of trained minds.

## Crowdsourcing can enable the next-generation of online education

In the other direction, running quality courses with thousands of students demands an efficient combination of instructor and student effort with computer algorithms to accomplish personalized teaching, learning, assessment, feedback, and support. Producing high quality content for a greater diversity of learners (across knowledge level, educational goals, country, culture, or occupation) in a range of subjects is an arduous challenge for individuals [14]. Crowdsourcing

techniques could help aggregate the effort and intelligence of tens of thousands of teachers and instructors who teach very similar courses and workshops every year in grades K-12, universities, and the workplace [9].

Crowdsourcing can also enable more personalized education, provided by students for students, in forms such as peer grading [8] and answering questions via discussion forums and online chat. Besides the direct use of crowdsourcing in MOOCs, we believe that some of the most useful contributions from the crowdsourcing domain are tried-and-tested methodologies to direct groups, simplify tasks, provide motivation, ensure trust, and guarantee quality.

## Topics

*Extending crowdsourcing methods for online education*
Q: How can the crowd of students in a Massive Open Online Courses (MOOC) as well as low-cost external crowds (e.g. Amazon Mechanical Turk) be used to support education at scale with thousands of learners?

- Utilizing a MOOC's crowd of students
- Quality Control: Cheating and Fraud Detection [5]
- Internal Peer Grading and Grading by External Crowds
- Optimally Combining Human & Machine Grading [6]
- Supporting Asynchronous and Real-time Peer Tutoring and Feedback
- Using student behavior to identify misconceptions and correct them [7]

*Extending Crowdsourcing to Collaborative Production of Online Educational Content for MOOCs*
Q: How can online resources be produced and improved through "high-skill" crowdsourcing and collaborative development, and efficient division of human-machine computation?

- Leveraging the accessibility of the Internet to bring together diverse crowds (teachers, tutors, students, interdisciplinary researchers) to tailor, customize, and personalize online education. [14]
  - Curriculum design
  - Customization of examples, explanations, etc. by students
  - Curation of content
- Leverage learners' collective input (either from natural interactions with content, or from carefully designed learner prompts) to create new content or improve the learning platform itself
- Learners' click-level interactions with MOOC videos have been analyzed to detect interaction peaks, which are points in a video with a significantly higher amount of learner activity (e.g., play button clicks or the number of viewers) than nearby points. Researchers observed that these peaks correspond to important or confusing points in a video [2].
- LectureScape is a MOOC video player that uses collective interaction peaks to improve video navigation, in-video search, and summarization [3].
- For how-to videos, Crowdy introduces a multi-stage learnersourcing workflow that prompts

learners who are actively watching the video to contribute summaries of steps. The collective input provides labels for groups of steps and their scope in a how-to video [12].

Q: Which software and technologies are well suited to supporting such crowdsourcing?

Q: What kinds of crowdsourcing patterns and division of labor processes & systems work well for producing online resources?

*Crowdsourcing Activities that simultaneously help learners and use their behavior to improve learning resources*
Q: How can we structure learners' interactions with or online resources to allow them to learn from, while iteratively improving, educational materials?

- Soliciting learner explanations and problem-solving to improve learning
- Gamification: real-time, rich feedback[4,5], large tasks with microbreaks

*How MOOCs can support Crowdsourcing*

- Training contributors new skills for jobs
- Vocational/market-driven education
- Using forums, communities and other mechanisms to scale learning on the job

## Organizers

Joseph Jay Williams, Harvard University, USA: Joseph is a Research Fellow at HarvardX. He does experimental and computational cognitive science research on online learning and "MOOClets" across platforms from Khan Academy to Moodle to EdX to ASSISTments, and has organized a **workshop on online learning at CHI 2014**, and an invited panel on **Collaborative Development of Online Educational Resources at the American Education Research Association**.

Markus Krause, Leibniz University, Germany: Markus is a postdoctoral research fellow at the Leibniz University. He investigates personalized online education at scale and the intersection between online work and online learning. He is also a professional game designer and worked as an art director for different game studios. Markus has organized conferences and workshops on human computation, online education, crowdsourcing, and digital games.

Praveen Paritosh, Google, USA: Praveen is a research scientist at Google working on human and machine intelligence. With Matt Lease and Tatiana Josephy, he organized a successful workshop at HCOMP 2013, Crowdsourcing at Scale [www.crowdscale.org]; and is on the organizing and program committees of AAAI.

Jacob Whitehill is a research fellow at HarvardX. His research interests are in machine learning, computer vision, automatic facial expression recognition, and their intersection with education and crowdsourcing. He completed his PhD at UCSD in 2012, MSc at the University of the Western Cape (South Africa) in 2007, and BS at Stanford in 2001, all in Computer Science. Prior to joining Harvard he was a research scientist and co-founder at Emotient.

Justin Reich is an educational researcher interested in the future of learning in a networked world, and the Richard L. Menschel HarvardX Research Fellow, a Fellow at the Berkman Center for Internet and Society, a Lec-

turer at the Harvard Graduate School of Education, and a lecturer in the Scheller Teacher Education Program at MIT. Justin is the co-founder of EdTechTeacher, a professional learning consultancy devoted to helping teachers leverage technology to create student-centered, inquiry-based learning environments. He has led the *Distributed Collaborative Learning Communities* project, a Hewlett Foundation funded initiative to examine how social media are used in K-12 classrooms.

Juho Kim is a Ph.D. candidate in EECS at MIT, focusing on Human-Computer Interaction (HCI) research. He works in the User Interface Design Group in CSAIL (Computer Science and Artificial Intelligence Laboratory), advised by Prof. Rob Miller and Prof. Krzysztof Gajos. The Samsung Fellowship is supporting his graduate studies. He designs interactive technologies powered by large-scale learning interaction data. His research introduces learnersourcing, where learners' natural learning interactions collectively generate novel educational contents not possible before. His approach combines a) crowdsourcing to collect and process a large amount of learner input, b) social computing to incentivize learners through collaboration, and c) content-based video analysis techniques such as computer vision and natural language processing to complement learner input.

As Chief Scientist at edX and its technical co-founder, Piotr Mitros is charged with developing and applying technology to optimize the learning process. Mitros led the creation of the original MITx platform and helped lead the creation of its pedagogy. He brings a broad interdisciplinary background that combines teaching, engineering, computer science, and math, and has been interested in teaching and education since he was a child.

Neil Heffernan is an Associate Professor at Worcester Polytechnic Institute in the Computer Science Department. He co-directs the new PhD + Masters Programs in Learning Sciences and Technologies and is well known for the web-based system he created called ASSISTments. ASSISTments has been used, in partnership with a dozen others universities, as a shared scientific instrument to conduct randomized controlled trials comparing different ideas on how to improve student learning. He also has expertise in Artificial Intelligence, Machine Learning, Intelligent Tutoring Systems, Educational Data Mining, Experiential Design, Psychometrics and the Cognitive Science of Mathematical Cognition.

Brian C. Keegan is a computational social scientist and research associate at the Harvard Business School's HBX massively open online course (MOOC) platform. He was previously a post-doc in David Lazer's lab at Northeastern University where his work focused on online political behavior and web experiments. He received his Communication PhD in 2012 from Northwestern University and his dissertation examined Wikipedia's coverage of breaking news events. His research has been published in CSCW, ICWSM, WWW, American Behavioral Scientist, and PLoS ONE.

## References
[1]  Goodfellow, I. J., Bulatov, Y., Ibarz, J., Arnoud, S., & Shet, V. (2013). Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks. arXiv preprint arXiv:1312.6082.

[2]  Kim, J., Guo, P. J., Cai, C. J., Li, S., Gajos, K. Z., & Miller, R. C. (2014). Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th annual ACM symposium on User interface software and technology (UIST '14).* ACM, New York, NY, USA, 563-572.

[3]  Kim, J., Guo, P. J., Seaton, D. T., Mitros, P., Gajos, K. Z., & Miller, R. C. (2014). Understanding in-video dropouts and interaction peaks inonline lecture videos. *In Proceedings of the first ACM conference on Learning @ scale conference (L@S '14).* ACM, New York, NY, USA, 31-40.

[4]  Krause, M., Mogalle, M., Pohl, H., & Williams, J. J. (2015). A Playful Game Changer : Fostering Student Retention in Online Education with Social Gamification. In *Proceedings of the second ACM conference on Learning@ scale conference (L@S'15)*. Vancouver, Canada: ACM Press.

[5]  Krause, M. (2014). *Homo Ludens in the Loop: Playful Human Computation Systems* (p. 232). Hamburg, Germany: tredition GmbH.

[6]  Mitros, P., Paruchuri, V., Rogosic, J., & Huang, D. (2013). An Integrated Framework for the Grading of Freeform Responses. In *Proceedings of the Sixth Conference of MIT's Learning International Networks Consortium.*

[7]  Mitros, P., & Sun, F. (2014, July). Creating Educational Resources at Scale. In *2014 IEEE 14th International Conference on Advanced Learning Technologies (ICALT),* (pp. 16-18). IEEE.

[8]  Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). Tuning Peer Grading. In *Proceedings of the 6th International Conference on Educational Data Mining. Memphis, Tennessee*.

[9]  Porcello, D., & Hsi, S. (2013). Crowdsourcing and Curating Online Education Resources. *Science, 341*(6143), 240-241.

[10] Razzaq, L. & Heffernan, N. (2010). Open content Authoring Tools. In Nkambou, Bourdeau & Mizoguchi (Eds.) *Advances in Intelligent Tutoring Systems.*pp 425-439. Berlin: Springer Verlag.

[11] Retelny, D., Robaszkiewicz, S., To, A., Lasecki, W. S., Patel, J., Rahmati, N., & Bernstein, M. S. (2014, October). Expert crowdsourcing with flash teams. In Proceedings of the 27th annual ACM symposium on User interface software and technology (pp. 75-85). ACM.

[12] Weir, S., Kim, J., Gajos, K. Z., & Miller, R. C. (2015). Learnersourcing Subgoal Labels for How-to Videos. In *Proceedings of the 18th ACM conference on Computer supported cooperative work & social computing (CSCW '15)*, to appear.

[13] Williams, J.J., Kizilcec, R.F., Russell, D. M., & Klemmer, S. R. (2014a). Learning Innovation at Scale. *Workshop at ACM CHI Conference on Human Factors in Computing Systems.*

[14] Williams, J. J., Linn, M., Edwards, A., Trumbore, A., Chae, H. S., Natriello, G., Saxberg, B., & Mitros, P. (2014b). How online resources can facilitate interdisciplinary collaboration. Panel at American Educational Research Association.
(tiny.cc/onlinecollaboration)