

The Future of Adaptive Learning: Does the Crowd Hold the Key?

Neil T. Heffernan¹ · Korinn S. Ostrow¹ · Kim Kelly¹ ·
Douglas Selent¹ · Eric G. Van Inwegen¹ · Xiaolu Xiong¹ ·
Joseph Jay Williams²

Published online: 2 February 2016

© International Artificial Intelligence in Education Society 2016

Abstract Due to substantial scientific and practical progress, learning technologies can effectively adapt to the characteristics and needs of students. This article considers how learning technologies can adapt over time by crowdsourcing contributions from teachers and students – explanations, feedback, and other pedagogical interactions. Considering the context of ASSISTments, an online learning platform, we explain how interactive mathematics exercises can provide the workflow necessary for eliciting feedback contributions and evaluating those contributions, by simply tapping into the everyday system usage of teachers and students. We discuss a series of randomized controlled experiments that are currently running within ASSISTments, with the goal

✉ Korinn S. Ostrow
ksostrow@wpi.edu

Neil T. Heffernan
nth@wpi.edu

Kim Kelly
kkelly@wpi.edu

Douglas Selent
dselent@wpi.edu

Eric G. Van Inwegen
egvaninwegen@wpi.edu

Xiaolu Xiong
xxiong@wpi.edu

Joseph Jay Williams
joseph_jay_williams@harvard.edu

¹ Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, USA

² HarvardX, Harvard University, 125 Mt. Auburn Street, Cambridge, MA 02138, USA

of establishing proof of concept that students and teachers can serve as valuable resources for the perpetual improvement of adaptive learning technologies. We also consider how teachers and students can be motivated to provide such contributions, and discuss the plans surrounding PeerASSIST, an infrastructure that will help ASSISTments to harness the power of the crowd. Algorithms from machine learning (i.e., multi-armed bandits) will ideally provide a mechanism for managerial control, allowing for the automatic evaluation of contributions and the personalized provision of the highest quality content. In many ways, the next 25 years of adaptive learning technologies will be driven by the crowd, and this article serves as the road map that ASSISTments has chosen to follow.

Keywords Crowdsourcing · Learnersourcing · Feedback · Learning gains · Online learning platform · Adaptive learning technologies · ASSISTments

Evolving Adaptive Learning Technologies Through Crowdsourced Contributions

For this Special Issue, we were asked to predict elements that would drive the next 25 years of AIED research. Clairvoyance is difficult, if not impossible, and if we were to provide readers with definitive strategies to guide the next quarter century of research, we would likely suggest far more “misses” than “hits.” Instead, we use this work to examine the modest steps that ASSISTments, a popular online learning platform, will be taking to accommodate issues of growing importance. Over the next 25 years, it is our hope that adaptive learning technologies will extend support for best practices in K-12 learning through rigorous experimentation to identify and implement personalized educational interventions in authentic learning environments. We anticipate that while big data will be used to improve these platforms (i.e., through educational data mining and learning analytics), innovations in this area will be restricted by pedagogy and by fine-grained, personalized support for all learners. Still, growth rooted in best practices will be necessary to keep the field from growing stagnant.

Specifically, this article considers how improvements for a perpetually evolving educational ecosystem can be solicited dynamically and at scale through *crowdsourcing* (Kittur et al. 2013; Howe 2006). We provide a brief background on crowdsourcing, noting the issues inherent to the concept, and the novelty of its use in educational domains. We follow this discussion with a detailed description of the ASSISTments platform and its feedback capabilities in their current form. We then highlight a number of randomized controlled experiments that have run or are currently running within ASSISTments that outline the steps that the ASSISTments team is taking to implement crowdsourcing within the platform. We believe that others in the AIED community should consider similar work in the coming years. We then outline the process by which ASSISTments plans to implement crowdsourcing, through an infrastructure we refer to as PeerASSIST. We explain how crowdsourced feedback contributions will be collected and how the platform will use sequential design as a managerial control to isolate and deliver high quality contributions to other learners. We conclude our discussion by linking our work within ASSISTments to general implications for the AIED community and the coming 25 years of research.

Recent research has suggested that large improvements to adaptive learning technologies can be produced through a multitude of small-scale, organic contributions from distributed populations of teachers and students. Users of these platforms receive content via online and blended education systems and, in return, provide data on learning and interactions (i.e., log files). It would be relatively simple to incorporate more elaborate user contributions, in the form of solution explanations or ‘work shown,’ as pedagogical innovations that underlie systemic change (Howe 2006; Von Ahn 2009). Thus, when considering the next 25 years of AIED, especially at scale, safety will be in numbers, and ASSISTments will be following the crowd.

Crowdsourcing

Extensive discussion surrounds the challenge of clearly defining crowdsourcing (Estellés-Arolas and González-Ladrón-de-Guevara 2012). In this article we use the term crowdsourcing to contrast obtaining curriculum or pieces of content designed by a single expert (or a small team of experts) (Porcello and Hsi 2013) with obtaining contributions from many people, who tend not to be restrictively vetted or selected, and whose efforts are voluntary. There is tremendous evidence for the power of crowdsourcing in human-computer interaction research (Doan et al. 2011; Howe 2006; Kittur et al. 2013), with recent work covered by many publications at venues like HCOMP (Conference on Human Computation and Crowdsourcing), CSCW (Computer Supported Cooperative Work and Social Computing), CHI (Computer-Human Interaction), and Collective Intelligence (see also Malone and Bernstein 2015). Despite the trending popularity of crowdsourcing, adaptive learning settings have not taken advantage of the approach as a viable framework for success.

Not to be confused with the “wisdom of crowds,” or the assumption that the whole is greater than the sum of its parts (Surowiecki 2004), crowdsourcing does not necessarily require a “wise” crowd, or one with cognitive diversity, independence, decentralization, and aggregation, as described by the framework set forth by Surowiecki (2004). Many instances of crowdsourcing have proven successful without crowd member independence or the aggregation of opinions (Saxton et al. 2013). It therefore follows that any crowd, even those comprised of novices rather than domain experts, may serve as a helpful resource. By sourcing content contributions from users within adaptive learning platforms, it is possible to expand the breadth and diversity of available material beyond that born of just a few designers, supporting the personalization of online educational content (Organisciak et al. 2014; Weld et al. 2012). Rather than designing a theoretical framework for the sound implementation of crowdsourcing across platforms or in differing scenarios (see Saxton et al. (2013) for a meta-analysis exemplifying structured crowdsourcing models), we focus primarily on the logistics of, and issues surrounding, outsourcing content creation to active users of an adaptive learning platform.

In addition to the academic literature in human-computer interaction, many well-known websites and Web 2.0 services (i.e., Facebook, Flickr) involve crowdsourcing activities. Perhaps the most prominent success story built on the crowdsourcing approach is Wikipedia, the free online encyclopedia that relies on crowdsourcing to author and edit content. Wikipedia has surpassed the capabilities of previous electronic

encyclopedias (i.e., Encyclopedia Britannica) by taking an approach that was initially criticized and met with skepticism: a wide range of users, all free to create, edit, flag, and delete content. This approach, now common to all “wikis,” constitutes a *knowledge base building model* (Saxton et al. 2013), requiring high levels of crowd collaboration with little-to-no compensation in return.

In contrast, Stack Overflow and Yahoo Answers are two examples of crowdsourcing sites designed to allow users to interact and to provide others with assistance, thereby building a knowledge base. Such sites have shown compelling benefits (Anderson et al. 2012). For instance, Stack Overflow is among the top 50 most visited websites on the Internet and is used by 26 million programmers each month (<http://stackexchange.com/about>). Within this implementation of crowdsourcing, any user is able to ask questions related to programming, and others in the community are able to provide answers. Additionally, users can “upvote” or “downvote” questions and answers to promote the most accurate and helpful content. Further, questions can be linked, marked as duplicates, flagged as inappropriate, or commented upon with general responses. Stack Overflow then uses an algorithm to rank users according to the “value” of his or her answers, thereby helping to efficiently highlight quality content from domain experts in the crowd.

In many other large technological platforms, processes for crowdsourcing have provided valuable solutions. Popular examples can be categorized by various model types within the theoretical framework set forth by Saxton et al. (2013), including those focused on *collaborative software development* like game design (Von Ahn and Dabbish 2008) and the programming of mobile apps by sourcing the efforts of experts with specialized skills (Retelny et al. 2014), those focused on *citizen media production* (i.e., YouTube, Reddit), and those focused on *collaborative science projects* like the digitization and translation of books and addresses, and image identification at scale (Griswold 2014). Such crowdsourcing activities also interface with discussions around “Big Data” and “Data Science” (Manyika et al. 2011; Boyd and Crawford 2012) as novel kinds of data and analyses emerge as social network interactions (Tan et al. 2013) and crowdsourcing behaviors (Franklin et al. 2011) grow in popularity.

While the previous examples offer powerful and compelling uses of crowdsourcing, the concept still faces challenges in the realm of education. Issues that arise within educational domains, including managerial control, or how to evaluate and enforce high quality user contributions, continue to plague crowdsourcing systems in other domains (Saxton et al. 2013). For instance, Stack Overflow cannot outwardly measure which answers have more of an effect on learning outcomes. Although users might assume that highly “upvoted” content is the most reliable, there is no qualitative way to survey users after having read each answer to determine differences in learning gains. Similarly, open authorship on “wikis” allows users to supply inaccurate content or to destroy accurate content through malicious edits. Without a principled approach for evaluating the quality of contributions beyond user opinion, Wikipedia faces skepticism from those in education about the reliability and the veracity of its content.

Improving Education Through “Teachersourcing”

Despite the lack of its use within educational domains, crowdsourcing holds great promise for the future of adaptive education, with few substantial obstacles (Williams

et al. 2015a). Teachers and experts can curate and collect high quality educational resources online (Porcello and Hsi 2013), with research showing success in authoring expert knowledge for intelligent tutors and educational resources by using crowds of teachers (Floryan and Woolf 2013; Aleahmad et al. 2008). However, the majority of adaptive learning technologies that offer personalized instruction lack the infrastructure required to obtain sufficient contributions from the crowd and to then return customized instruction to match students' needs. For example, to solve a problem requiring students to add fractions with unlike denominators, adaptive learning systems typically provide scaffolded instruction that walks the student through finding a common denominator, creating equivalent fractions, and then adding the fractions. However, the Common Core State Standards (NGACBP and CCSSO 2010) emphasize multiple approaches to problem solving, often with varying complexity. For example, one student may use a manipulative, such as fractions of a circle, to find equivalent fractions and then carry out the addition. Another student may take a more sophisticated approach by listing all equivalent fractions for each fraction in order to find a common denominator. A third student may instead use an algorithm to find the least common multiple and carry through with the addition using this as the denominator. An adaptive learning system that is assisting a student with this problem should know all potential approaches, know which approach is most appropriate given the student's actions, and provide the assistance that will optimize benefit for each student. This is where the idea of implementing crowdsourced content or feedback within an educational context can grow exceedingly necessary. A single teacher may not be the most apt at explaining all topics to all students. If multiple approaches exist to solve a problem, and the teacher consistently teaches only a single approach or method, students may fail to grasp what they would perhaps otherwise understand when taught using a different approach (Ma 1999).

Crowdsourcing feedback material from teachers would allow for an expanse in the probability that students will learn from an effective teacher, or possibly from an effective combination of teachers (Weld et al. 2012). Some platforms in the AIED community are already beginning to consider crowdsourcing, and a number of researchers in the community have shown interest in the topic. An academic collaboration has paired Professor Kong at Worcester Polytechnic Institute with Yahoo Answers to make progress in better predicting the quality of questions, the helpfulness of answers, and the expertise of users (Zhang et al. 2014). However, few adaptive learning technologies have considered this approach, and perhaps even fewer have considered crowdsourcing content from learners.

An Alternative Approach: “Learnersourcing”

Crowdsourcing feedback does not necessarily have to stop at teachers, or those considered domain experts. We believe that students can provide high quality worked examples of their solution path for a problem, or essentially “show their work.” Not only might the process of explaining their actions help to solidify their understanding of the content, but the feedback they provide can in turn be connected to the problem for the benefit of future students (Kulkarni et al. 2015). Student users spanning classrooms around the world offer a wealth of information; they can provide versatile explanations that would allow the system to incorporate all potential approaches for solving a

particular problem. Currently in most adaptive learning systems, when a student requests feedback in the form of a hint or scaffold, only a single approach is provided. Crowdsourcing student explanations has the potential to expand the capability of these systems to provide multiple, vetted approaches to the right students at the right times.

Engaging in “learnersourcing” (a term coined by Juho Kim and the CSAIL team at MIT, see Kim (2015)) may also be beneficial to students, if pedagogically useful activities like prompts for self-explanation are used to elicit student contributions (Williams and Lombrozo 2010). One line of work has had learners organically generate outlines for videos, by prompting them to answer questions like “What was the section you just watched about?”, having those answers vetted by other learners, and using the resulting information to dynamically build an interactive outline that can be delivered alongside the video (Weir et al. 2015). Weir et al. (2015) showed that this type of learnersourcing workflow could produce outlines for videos that lay out subgoals for learning in a way that is indistinguishable from outlines painstakingly produced by experts. We use the term “learnersourcing” in the present work simply to signify the crowd as comprised of student users of an online learning platform.

In theory, crowdsourcing could play an integral role in the future of adaptive learning. However, the issues surrounding the actual practice of crowdsourcing feedback within adaptive learning technologies are complex. What type of a system must exist for crowdsourcing to be easy and natural for users? After collecting a variety of feedback approaches for a particular problem, how should the system go about dispensing the proper feedback to the proper students at the proper times? We consider these issues, as well as others, as we discuss the intended future of harnessing the crowd within ASSISTments.

Issues Inherent to Learnersourcing

Although crowdsourcing has offered solutions for tasks that range from menial to complex, and while the technique will surely continue to prove effective moving into the future, one may argue that the use of crowdsourcing (especially learnersourcing) within adaptive learning technologies may carry a number of risks. Through a meta-analysis of 103 websites that implement some form of crowdsourcing, Saxton et al. (2013) established a comprehensive taxonomy for guiding the framework of crowdsourced designs. The approach to crowdsourcing that will guide the future of research within ASSISTments falls into their *knowledge base building model* framework. Considering the author’s summary of this type of model, “information- or knowledge-generation processes are outsourced to community users, and diverse types of incentive measures and quality control mechanisms are utilized to elicit quality knowledge and information that may be latent in the virtual crowd’s ‘brain’” (Saxton et al. 2013). Essentially, feedback creation could be outsourced to student users and can be incentivized through a grading rationale, with content quality managed by algorithms that promote the subsequent presentation of explanations that produce the greatest learning.

Saxton et al. (2013) suggest three primary issues with regard to crowdsourcing: 1) the “what” being outsourced, 2) the collaboration required from the crowd, and 3) managerial control over the quality of crowd based contributions. When outsourcing content creation to teachers, we are retrieving data from (more or less) domain experts.

However, when branching to learnersourcing, we are accepting contributions from “experts-in-training.” We argue that learnersourcing is acceptable within adaptive learning technologies if approached properly; we have lowered the complexity of the task at hand by relabeling feedback creation as “self-explanation.” The complexity of providing self-explanations for solutions to problem content may vary drastically in relation to the content in question. For instance, in mathematics, asking a student to explain how he or she solved a perimeter problem may be far less complex than asking a student to explain how he or she solved a logarithmic function. As these solutions are being sourced from the learner, or an “expert-in-training,” it grows more difficult to collect accurate and high quality content. Crowd collaboration can serve to alleviate some of this risk while establishing a framework for managerial control. Devising a voting methodology for the strongest content would allow for crowd collaboration, but may lead to social strife in classrooms if feedback is not collected anonymously (i.e., the potential for “downvotes” to high quality content as a form of bullying or social dominance). As Saxton et al. (2013) note, not all crowdsourcing systems require collaboration as part of a successful model, and as such, we consider it possible to establish learnersourcing that does not necessarily rely on the crowd’s collaborative impact. Content control within learnersourcing is easily the most difficult issue to tackle, and managerial control has been considered one of the greater challenges of crowdsourcing in general (Howe 2006; Howe 2008; McAfee 2006).

In the present work, we discuss a single method for implementing crowdsourcing within an online learning platform. We do not suggest that ASSISTments is the only platform capable of learnersourcing, nor do we suggest that we have found the ideal framework for implementation in other adaptive learning technologies. The framework that we set forth may or may not be generalizable to other platforms. However, we outline the steps that our team will be taking in the coming years (note that a large portion of this work is not yet substantiated) in hopes that the AIED community will consider crowdsourcing and the related issues as driving forces for research on the effects of feedback within adaptive learning technologies over the next quarter century.

Implementing Crowdsourcing Within Assistments

In the remainder of this article, we discuss how we hope to extend the ASSISTments platform to enable large-scale improvements through crowdsourcing from teachers and students. ASSISTments is an online learning platform offered as a free service of Worcester Polytechnic Institute. The platform serves as a powerful tool providing students with *assistance* while offering teachers *assessment*. Doubling its user population each year for almost a decade, ASSISTments is currently used by hundreds of teachers and over 50,000 students around the world with over 10 million problems solved last year. At its core, the premise of ASSISTments is simple: allow computers to do what computers do best while freeing up teachers to do what teachers do best. In ASSISTments, teachers can author questions to assign to their students, or select content from open libraries of pre-built material. While the majority of these libraries provide certified mathematics content, the system is constantly growing with regard to other domains (i.e., chemistry, electronics), and teachers and researchers are able to author content in any domain.

Specifically, the ASSISTments platform is driving the future of adaptive learning in some unique ways. The first is the platform's ability to conduct sound educational research at scale efficiently, ethically, and at a low cost. ASSISTments specializes in helping researchers run practical, minimally invasive randomized controlled experiments using student level randomization. As such, the platform has allowed for the publication of over 18 peer-reviewed articles on learning since its inception in 2002 (Heffernan and Heffernan 2014). While other systems provide many of the same classroom benefits as ASSISTments, few merit an infrastructure that also allows educational researchers to design and implement content-based experiments without an extensive knowledge of computer programming or other specialized skills with an equally steep learning curve. Recent NSF funding has allowed for researchers around the country to design and implement studies within the system, moving the platform towards acceptance as a shared scientific instrument for educational research.

By articulating the specific challenges for improving K-12 mathematics education to a broad and multidisciplinary community of psychology, education, and computer science researchers, leaders spanning these fields can collaboratively and competitively propose and conduct experiments within ASSISTments. This work can occur at an unprecedentedly precise level and large scale, allowing for the design and evaluation of different teaching strategies and rich measurement of student learning outcomes in real time, at a fraction of the cost, time, and effort previously required within K-12 research. While leading to advancements in the field through peer-reviewed publication, this collaborative work simultaneously augments content and infrastructure, thereby enhancing the system for teachers and students.

Pathways for Student Support Provide Potential for Crowdsourced Contributions

Previous work has aptly described feedback as “information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding” (Hattie and Timperley 2007). Students may receive one of many types of feedback within an ASSISTments assignment, depending on settings selected by the content designer (i.e., a teacher or researcher). The most basic form of support is *correctness feedback*; students are informed if they are correct or incorrect when they answer each question (this feature can be shut off by placing questions in ‘test’ mode when necessary). When more elaborate feedback is desired, questions may include *mistake messages* created by the content author, or sourced from teachers and classes that have isolated “common wrong answers.” These messages are automatically delivered to a student in response to particular mistakes, as shown in Fig. 1. Additionally, elaborate feedback can come in the form of on-demand *hints* that must be requested by the student and are presented sequentially (i.e., students may see the option “Show hint 1 of 3”). Hints are typically presented with increasing specificity before presenting the student with the correct answer (the “Bottom Out Hint”), allowing the student to move on to the next problem in the assignment rather than getting stuck indefinitely. Alternatively, ASSISTments offers a form of elaborate feedback that is typically used to present worked examples, or to break a problem down into smaller, more solvable sub-steps. This type of feedback is called *scaffolding*, and is presented when the student makes an incorrect response or requests that the problem be broken down into steps. A comparison of hint feedback and scaffolding is presented in Fig. 2.

A meta-analysis of 40 studies on item-based feedback within computer-based learning environments recently suggested that elaborated feedback, or that providing a student with information beyond the accuracy of his or her response, is considerably helpful for student learning, reporting overall effect sizes of 0.49 (Van der Kleij et al. 2015). To root this theory in ASSISTments terminology, elaborated feedback would include *mistake messages*, *hints*, and *scaffolds*, but not *correctness feedback*. It is also likely that the three types of elaborated feedback available within ASSISTments provide students with differential learning benefits, as they function differently with regard to timing and content specificity. Van der Kleij et al.'s (2015) examination of three previous meta-analyses revealed a gap in feedback literature: although feedback has been shown to positively impact learning, not all feedback provides the same impact. As such, it is possible that providing the worked solution for a problem is more beneficial to students than providing less specific hints. When considering learnersourcing, the type of feedback collected from a student, as well as its quality, should be taken into consideration as moderating the subsequent learning of other students that receive that content.

Assignment: Problem #PSA259R

Problem ID: PRA259R [Comment on this problem](#)

Solve for a

$$9a + 10 = 28$$

Check your sign
 Positive / Positive = Positive
 Negative / Negative = Positive
 Positive / Negative = Negative
 Negative / Positive = Negative

Step 1: Correct
 $9a + 10 = 28$
 $-10 \quad -10$

Step 2: Correct
 $\frac{9a}{9} = \frac{18}{9}$

Sign: Incorrect
 $a = 18/9$ IS NOT -2

Type your answer below (mathematical expression):

✘ Sorry, try again: "-2" is not correct

Fig. 1 An example of a *mistake message*. This type of feedback responds with tailored information that pinpoints exactly where the student made a mistake. If the student is unable to arrive at the correct answer with this guidance, standard hints are also available (note the “Show hint 1 of 3” button). The steps shown trace through the student’s probable solution path to isolate and correct the misunderstanding. In this example, the student was on the right track with steps 1 and 2, but incorrectly attributed the negative sign to his or her solution

Assignment: Problem #PSA4S9W

Problem ID: PRA4S9W [Comment on this problem](#)

Find the mean (average) of the numbers below.

11, 4, 13, 8

To find the mean, add all the numbers and divide the sum by how many numbers there are.

First, find the sum of the numbers.

Sum = $11 + 4 + 13 + 8 = 36$

[Comment on this hint](#)

Next, we need to know how many numbers we have. For this problem, there are 4 numbers.

[Comment on this hint](#)

Finally, divide the sum of all the numbers, 36, by how many numbers there are, 4.

Mean = $\frac{36}{4} = 9$

Type in 9.

[Comment on this hint](#)

Type your answer below (mathematical expression):

Assignment: Problem #PSA4AUZ

Problem ID: PRA4AUZ [Comment on this problem](#)

Find the mean (average) of the numbers below.

11, 4, 13, 8

Type your answer below (mathematical expression):

12

✖ Sorry, try again: '12' is not correct

Problem ID: PRA4AUZ - 1104185 [Comment on this problem](#)

Looks like you could use some help. Let's look at a similar Example Problem.

Example Problem:

Find the mean (average) of the numbers below.

20, 0, 9, 15, 10, 6

Step 1 of 3

To find the mean, add all the numbers and divide by how many numbers there are.

The first step is to sum up the numbers that were given.

$20 + 0 + 9 + 15 + 10 + 6 = 60$

Now it's your turn. Find the sum of all the numbers in your problem: 11, 4, 13, 8

Enter the sum in the box below.

Type your answer below (mathematical expression):

Fig. 2 A comparison of Hints and a Scaffold within an identical problem. Note that three hints are shown on the left, as requested by the student. On the right, the student provided an incorrect response and was automatically given a scaffold with a worked example on how to solve a similar problem. If the student is unable to answer this sub-step they can choose to have the answer revealed and move on to the next portion of the main problem, presented as a second scaffold

Figure 3 depicts a well-established model of learning from feedback, as proposed by Bangert-Drowns et al. (1991). Within this model, students begin at an initial knowledge state (the dashed circle), and when presented with a question, practice the information retrieval required to form a response. The student then receives feedback regarding their response that he or she can use to evaluate their response and adjust their knowledge accordingly. This process is iterative with each question, beginning again at the student's adjusted knowledge state (the dashed circle). This model is worthy of attention when designing adaptive learning technologies because the type of feedback supplied after each item will affect the learning process considerably. We present this model because it further highlights the risk of sourcing feedback from learners. If a learnersourced contribution is incorrect, or of extremely low quality, the contribution should not be presented to other students as feedback. It is crucial that learnersourced contributions only be displayed as credible feedback when it is clear that they produce measureable gains in students' knowledge state. Later in this article, we propose that it is possible to determine the effectiveness of learnersourced contributions through randomized controlled experimentation and the use of sequential design.

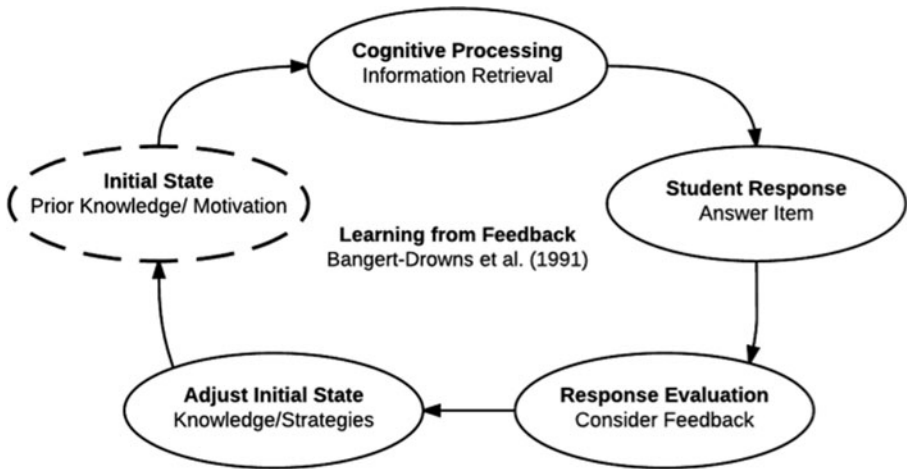


Fig. 3 Model of learning from feedback as proposed by Bangert-Drowns et al. (1991). Students begin at an initial knowledge state (dashed circle). When presented an item, information retrieval occurs and the student forms a response. Through feedback, the student evaluates his or her response and adjusts their knowledge accordingly. The process begins again at the student’s adjusted knowledge state

While students using ASSISTments benefit from the aforementioned elaborated feedback, teachers benefit from a variety of actionable reports on students’ progress. An example of an item report, the most commonly used report within ASSISTments, is shown in Fig. 4. This report has a column for each problem (i.e., “item”) and a row for each student, along with quantitative data tracking student and class performance. The first response logged by each student is provided for each problem, and teachers are able to monitor feedback usage and assignment times. Teachers often use the item

Student/Problem [Unanonymize]	Average Data driven	PRAHE5Y Data driven	PRAHE5Z Data driven	PRAHE52 Data driven
Problem Average	60%	27%	61%	84%
Common Wrong Answers		1/9^10, 56% +feedback	1/5^13, 58% +feedback	
Correct Answer(s)		1/3^10	1/5^3	1/16^2
XXXXX *	50%	✗ 1/9^10	✗ 1/5^13	✗ 1/162
XXXXX *	45%	✗ 1/9^10	✓ 1/5^3	✓ 1/16^2
XXXXX *	55%	✓ 1/3^10	✗ 1/5^13	✓ 1/16^2

Fig. 4 An item report that shows the first three problems and the first three students from a larger class and assignment. Each column represents a problem and each row represents a student. Each item has a percent correct, and if applicable, a common wrong answer. In the student row, the student average and first attempt for each problem is reported. For example, the second student answered the first problem incorrectly (he or she said 1/9^10) and the second and third problems correctly on the first attempt. Finally the “+feedback” link affords the teacher the opportunity to write a *mistake message* for the common wrong answer displayed

report in the classroom as a learning support because it provides actionable data. The report can be anonymized, as shown in Fig. 4, which randomizes student order and hides student names for judgment free in-class use. This report allows instructors to pinpoint which students are struggling and which problems need the most attention during valuable class time. The common wrong answers featured in this report are especially important in helping instructors diagnose students' misconceptions. They are shown in the third row of the table in Fig. 4.

From this type of report, teachers and students can see the percentage of students who answered the problem with a particular wrong answer (common wrong answers are those that at least three students made if representative of more than 10 % of the students in the class). In Fig. 4, only 27 % of the students answered the first problem correctly, leaving 73 % answering incorrectly. About half of the students who had an incorrect answer shared a common misconception and answered $1/9^{10}$. This problem seems worthy of class discussion. There is also a "+feedback" link available for teachers to write a *mistake message* for students who attempt this problem in the future, tailoring feedback based on the misconception displayed. Many teachers work through this process with their students, helping them to learn why the misconception is incorrect and how to explain the error to another student. This practice is what makes us believe that it is possible to learnersource feedback within systems like ASSISTments. The benefits of this type of learnersourcing would be both immediate (i.e., students learn to explain their work and pinpoint misconceptions) and long lasting (i.e., students that attempt this problem in the future can access elaborate feedback that targets their misconceptions).

The Potential Role of Video in Crowdsourced Contributions

Within ASSISTments, and in many similar adaptive learning platforms, content and feedback are facing a digital evolution. The recent widespread availability of video has spearheaded a variety of intriguing innovations in instruction. Projects like MOOCs (Massive Online Open Courses) and MIT's OpenCourseWare™ have exposed students to didactic educational videos on a massive scale. Video lectures can be created by the best lecturers around the world and provided to anyone, allowing professors that were once a powerful resource to a limited audience to now impact any willing learner. These lectures can reach very remote parts of the world and can be accessed by those that would otherwise never have the opportunity to attend a world-class university. The universal power of the video lecture suggests that there is a "time for telling" (Schwartz and Bransford 1998), and that eager learners can use technology to access the knowledge of experts and understand the bulk of the story.

However, many learners require more than just the storyline; students often need reinforcement and support while practicing what they have learned. We advocate for the use of video beyond lectures and into the realm of short tutorial strategies as lecturing is only a small portion of an instructor's job that can be captured on video. By only focusing on lectures, thousands of students lose out on unique explanations and extra help that can be provided through individually tailored tutoring. The greatest teachers spend a large portion of their time tailoring instruction to a struggling student's individual needs. Adaptive learning technologies need to consider the problem of

capturing and delivering these just-in-time supports for students working in class and at home, and we argue that videos offer a starting point.

When ASSISTments first began, all tutorial strategies were presented using rich text. However, with content authors and student users gaining more prevalent access to video, both in the classroom and at home, ASSISTments has recently experienced an increase in volume of video explanations. Recent technological advances have made it easy for almost anyone to create and access video as support for learning. The platform has responded by making it easier for users to create videos while working within particular problems. The ASSISTments iPad application has recently been upgraded to include a built-in feature that allows users to record Khan Academy style “pencasts” (a visual walkthrough of the problem with a voice over explanation) while working within a problem. In the near future, the app will allow for these recordings to be uploaded to YouTube and stored within our database as a specific tutorial strategy for that problem. Although this linking system is still under development, the use of video within ASSISTments is already expanding through more traditional approaches to video collection and dissemination. Teachers have started to record their explanations, either in the form of a pencast or by recording themselves working through a problem on a white board, uploading the content to a video server, and linking the content to problems or feedback that they have authored. In the past year, ASSISTments has witnessed the use of videos as elaborate explanations (i.e., *hints*, *scaffolds*), as *mistake messages* to common wrong answers, and even for instruction as part of the problem body.

But why would the production of video by crowds of teachers (or even students) be helpful? Consider the following use case:

A tutor is holding an after school session for five students who need extra help as they prepare for their math test. The tutor circulates around the small classroom, working with each student while referencing an ASSISTments item report on her iPad. She notices that one of the students answered a problem incorrectly and that his solution strategy includes a misconception about the problem. While tutoring him through the mistake, the tutor uses the interface within the ASSISTments app to record the help session, explaining where the student went wrong and how to reach the correct solution (essentially a conversational *mistake message*). The recording includes both an auditory explanation and a visual walkthrough of the problem as the tutor works through the misconception. The explanation takes about 20 seconds to provide, but because it has been captured, it must only be provided once. Following this instance of helping the student, the tutor quickly uploads her video to YouTube and links the material to the current problem. Within five minutes, another student at the extra help session reaches the same problem and tries to solve it using the same misconception. The newly uploaded feedback video is provided as a *mistake message* and the student is able to correct her own error by watching the video and attempting the problem again. Meanwhile, the tutor is able to help a third student on a different problem, rather than having to provide that first help message repetitively.

This use case is the perfect embodiment of the vision that ASSISTments holds for the future of adaptive tutoring. The process does not exclude the human tutor from the

feedback process, but rather harnesses the power of explanations given *once* to help students across *multiple* instances. We have purposely used the noun “tutor” here rather than “teacher” to signify that students may also be able to provide video feedback to help their peers through tough problems. By using this approach iteratively across many problems, or to collect numerous contributions for a particular problem, we argue that adaptive learning technologies can expand their breadth of tutoring simply by accessing the metacognitive processes already occurring within the crowd.

How can we convince teachers (and students) that the process of collecting feedback and building a library of explanations is useful? Suppose that the goal is to collect feedback from various users to expand the library of *mistake messages* to cover every common wrong answer for every problem used within remedial Algebra 1 mathematics courses. If we consider problems from only the top 30 basic Algebra 1 math textbooks in America, estimating 3000 questions per book, it leaves a total of 90,000 questions requiring feedback. High quality teachers across the country have already generated explanations to many of these problems, but they have been lost on individual students rather than recorded and banked for later use by all students. If every math teacher in the country were to explain five math questions per day, roughly 30 million explanations would be generated per year. Even if just one out of every 300 instructors captured an explanation, feedback would be collected for all 90,000 questions within a single year. Students working through these problems could also be tasked with contributing by asking them to “show their work” on their nightly homework (a process that many teachers already require) or capturing in-class discussions surrounding common misconceptions. By implementing crowdsourcing, perhaps as described here through the collection of video feedback, adaptive learning technologies can potentially access rich user content that would otherwise be lost.

Guiding the Crowd

We anticipate that in the coming years, adaptive learning technologies will incorporate mechanisms for interactivity in eliciting contributions at scale, or *directed* crowdsourcing (Howe 2006). In our platform, we are hoping to achieve this by extending ASSISTments’ existing commenting infrastructure, which already provides teachers and researchers with the ability to interact with learners. By leveraging this system, we anticipate allowing learners to “show their work,” or provide elaborate feedback to peers that can be delivered as *hints*, *scaffolds*, or *mistake messages*. This process takes a complex task (content creation) and dilutes it into elements common to traditional mathematics homework. Crowdsourcing simple tasks requires a much different framework than that required for solving complex problems (Saxton et al. 2013). By scaling down the task requested of each learner, the process of learnersourcing becomes much more viable. We suggest that other adaptive learning technologies seeking to implement crowdsourcing consider task complexity and how to best access the ‘mind of the crowd.’

Currently within ASSISTments, each time a student works on a problem or is provided a *hint*, they are also provided a link from which they can write a comment. Students’ comments are collected and delivered both to the student’s teacher and to the problem’s author, as shown in Fig. 5. Teachers are able to act on comments by helping students individually, while content authors (if not the teacher) are able to use the

Comments From Other Users on Problems I've Built				See Comments From My Students on Problems I've Assigned	
Time	Who	Type	Comment	Actions	Problem ID
May 29, 2014 04:58 AM	Student	I learned something from this hint	I was doing 9.4-6.6 (Flag this comment)	Test Drive - Edit - Delete	PRAJBX2
May 29, 2014 04:54 AM	Student	Drastic changes are needed for this hint	Completely unhelpful. (Flag this comment)	Test Drive - Edit - Delete	PRAJGYJ
May 29, 2014 03:39 AM	Student	I don't understand this hint	I did that equation Mr. J. I got b for the x value as 0.07349 and divided it by the 5Eb which is 0.10101 I got .727 and it is not the right answer. (Flag this comment)	Test Drive - Edit - Delete	PRAJ9JN

Fig. 5 Comments from users on specific problems. Some of the comments are routine while others give the content author genuinely helpful information. This commenting infrastructure could be built out to allow students to “show their work” or to upload useful videos that would subsequently become *hints*, *scaffolds*, or *mistake messages* for their peers

comments (which have been anonymized) to enhance the quality of their questions. Students working within ASSISTments have already written 80,000 comments while solving roughly 20 million problems within the last five years. The commenting infrastructure includes a pull down menu as a sentence starter (see Fig. 6) as well as a text field where students can provide their comment. Students use this feature to request assistance, to communicate their confusion, or to provide input on the question or answer (as shown in Fig. 5). It is possible that based on the depth of their understanding, asking students to “show their work” through a commenting structure may not be helpful to the student providing the contribution (Askey 1999), but that the worked solution would ultimately prove insightful for other students when presented as elaborated feedback.

The goal for the future of ASSISTments is to use a similar infrastructure to learnersource feedback. We hope to harness the power of YouTube, or similar video servers, alongside ASSISTments problem content while using sequential design and developing multi-armed bandit algorithms to aid in subsequent feedback delivery (discussed in a later section). This approach will build off of functionalities that already exist within the ASSISTments platform, but it will allow users to efficiently create and add feedback to the system. Figure 7 depicts a mockup of an alternate crowdsourcing interface, distinct from the commenting infrastructure, that could allow content authors to easily create and link elaborate feedback to problems in the form of *mistake messages*. This environment offers a more in-depth approach that seeks to source content from domain experts (i.e., teachersourcing) rather than learners. In Fig. 7, the

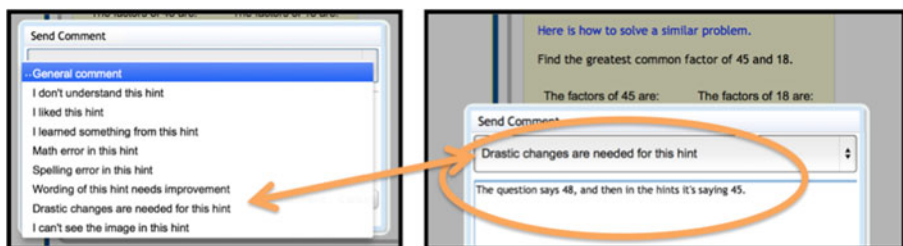


Fig. 6 The current structure for commenting within ASSISTments. Students that want to leave a comment select from the pull down menu first, then type in a more specific comment

Problem Statement:

Solve for c

$C - 2 = -18$

Common Mistake
Of the 547 students who got the problem wrong,

30% said: **-20**
 Cannot determine

22% said: **20**
 Cannot determine

11% said: **-6**
 Cannot determine

Mistake Message – Create a video, draw a picture, type words.

Font Sizes - B / U S

https://www.youtube.com/watch?v=OLdISVjHoaM

Font Sizes - B / U S

https://www.youtube.com/watch?v=1dZT5Zx4k-w

Font Sizes - B / U S


Many people enter -6 but it is not right.

Write an explanation on how to solve this problem. If you want to do it in steps just add a new box.

Add a new text box.

Font Sizes - B / U S

Watch my video. Pay attention to how I keep track of my signs. 😊




Save for later
Submit

Fig. 7 A mockup presenting an infrastructure for crowdsourcing tutorial strategies from domain experts. The problem content is stated at the top of the image. The user is able to create video or text feedback tailored to the three common wrong answers and is also provided the option to create a more generic explanation

content author is informed of three common wrong answers and given the space to respond accordingly to each scenario. The Figure shows that 30 % of students responded “-20,” although the correct answer would be “-16.” In response, the content author recorded and linked a YouTube video with tutoring specific to the error. The content author then uploaded a separate video link for the 22 % of students that responded with “20” as a common wrong answer. At the bottom of Fig. 7, the “explanation” section allows the content author to add a more general comment or video that offers elaborate feedback as an on-demand *hint*. When students tackle this problem in the future, those that make common wrong answers will receive tailored feedback, while those that request a hint will receive the general explanation. This approach to crowdsourcing is more adept to teachers and content authors, as they provide the domain expertise required to tease out the misconceptions behind common wrong answers. As such, we suggest that the AIED community consider teachersourcing (or crowdsourcing from domain experts) independently from learnersourcing, rather than considering all users as part of the same crowd.

While our schematics provide insight into how the actual process of crowdsourcing could work within an online learning platform, we are left with questions about how to learn which contributions are the most useful, for which learners, and under what contexts? We do not propose that the approaches presented here are the only methods for collecting student and teacher contributions, nor are we claiming that ASSISTments will be the only platform capable of these types of crowdsourcing. In the present work,

 Springer

we simply discuss the paths taken by the ASSISTments team to build interfaces to collect user explanations and leverage those contributions as feedback content. In the next section, we discuss a variety of randomized controlled trials that have been conducted within ASSISTments in an attempt to theorize on some of these important issues. We follow this discussion with an outline of our approach to delivering personalized content and feedback using sequential design and multi-armed bandit algorithms.

Evaluating Crowdsourced Content via Randomized Controlled Experiments

What rigorous options are available to evaluate the contributions made by the crowd? ASSISTments is unique in the technological affordances it provides for randomized experiments that compare the effects of alternative learning methodologies on quantifiable measures of learning (Williams et al. 2015b). Experimental comparisons can therefore be used within the platform to evaluate the relative value of crowdsourced alternatives, just as they are used to adaptively improve and personalize other components of educational technology (Williams et al. 2014). The promise of this approach is reinforced by numerous studies within ASSISTments that have already identified large positive effects on student learning, by varying factors like the type of feedback provided on homework (Mendicino et al. 2009; Kelly et al. 2013; Kehrer et al. 2013). A series of similar experiments currently serve as a proof of concept for various iterations of teachersourcing and learnersourcing elaborate feedback.

Comparing Video Feedback to Business as Usual

Ostrow and Heffernan (2014) inspired the use of video feedback by designing a randomized controlled experiment to examine the effectiveness of various feedback mediums. This study sought to examine the effects on learning outcomes if identical feedback messages were presented using short video snippets. Student performance and response time were analyzed across six problems pertaining to the Pythagorean theorem. As shown in Fig. 8, feedback was matched across delivery mediums. All students had the opportunity to receive both text and video feedback during the course of the assignment, but only saw feedback if they requested assistance or if they answered a problem incorrectly. Learning gains were examined on the second question across students who received feedback on the first question ($n = 89$). Video feedback improved student performance on the next question, although results were not statistically significant (two-tailed, $p = 0.143$). Still, results suggested a moderate effect size (0.32), and it is likely that the finding would grow statistically significant with an increase in power. Following the problem set, students were asked a series of survey questions to judge how they viewed the addition of video to their assignment. Based on self-report measures, 86 % of students found the videos at least somewhat helpful and 83 % of students wanted video in future assignments. Multiple problem sets in differing math domains have since been modified to include video feedback in an attempt to replicate these findings; analyses are not yet available. The results of this study, coupled with the ability to easily record

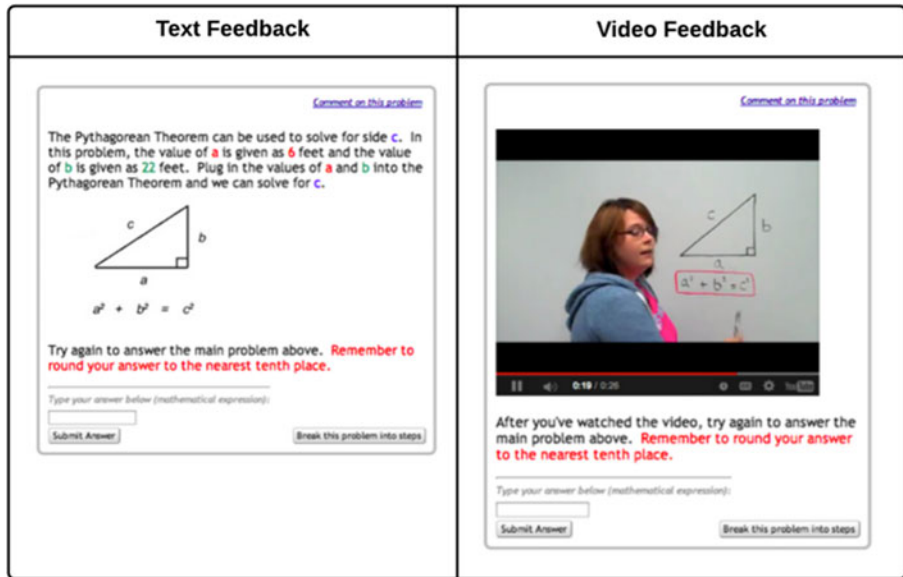


Fig. 8 Text and Video Feedback conditions as experienced by students (Ostrow and Heffernan 2014). Isomorphic problems featured matched content feedback across mediums, and struggling students showed greater benefits from receiving video feedback

and upload video content within ASSISTments, leave our team confident that video feedback can be implemented as viable crowdsourced content.

Comparing Contributions from Different Teachers: Proof of Concept

Selent and Heffernan (2015) took video feedback a step further to try to understand the potential benefits of crowdsourced *mistake messages* made by domain experts. These messages were made by a teacher that now works as part of the ASSISTments team, following a structure similar to that depicted in Fig. 7 as a proof of concept. The goal of this work was to determine if video tutoring used as *mistake messages* for common wrong answers, paired with access to the correct answer through a “Bottom Out Hint,” would prove more effective than just providing students with the correct answer. As shown in Fig. 9, each video was 20–30 seconds in length, offering a single, tailored message to misconceptions students might have when solving one-step equation problems. Students in the control group received a problem set featuring feedback that was restricted to the correct answer, to keep them from getting stuck on a problem. Those in the experimental condition received the same problem set with feedback altered to include tailored video mistake messages. These videos explained the process the student had used to arrive at their incorrect answer and how to return to the correct solution path. In a sample of 649 students (n control = 328, n experimental = 321), no significant differences were observed in completion rates for the assignment, the number of problems required for completion, or the accuracy and attempt count on the next question following a student’s initial incorrect response (i.e., following their experience of either a video message or the answer). Thus, while the addition of teacher videos to realign common misconceptions was not *harmful* to student

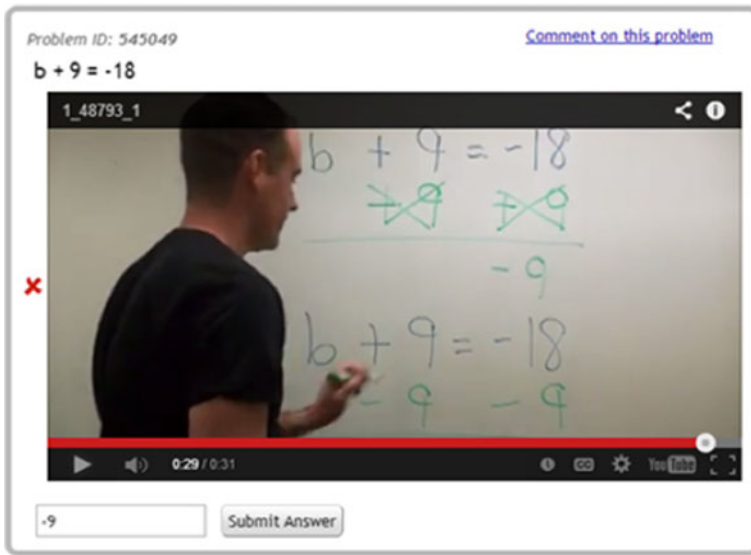


Fig. 9 Teacher created video used as a mistake message tailored to the common wrong answer of “-9.” In the 31 second clip, the teacher notes that the student added 9 to both sides when he or she should have subtracted 9 from both sides. Multiple videos, or similar feedback delivered using alternative mediums, can be teachersourced for use as elaborate feedback

learning, this study did not prove that video was helpful when used as mistake messages. These null findings may be due in part to the limited sample of students that received tailored messaging based on common wrong answers. Still, this study suggested that an infrastructure like that depicted in Fig. 7 would be viable to teachersource feedback content.

Comparing Contributions from Students: Proof of Concept Designs

One of the more unique studies currently running within ASSISTments considers learnersourced contributions in a controlled environment to determine which contribution leads to greater learning. This randomized controlled experiment examines two versions of feedback for the same problem on elapsed time, sourced from two different students. These students were not directed in their explanations, they were simply asked to “show their work” in solving the problem to benefit their peers that may still be struggling. As shown in Fig. 10, the resulting explanations were rather different when students were asked to determine the amount of elapsed time between 7:30 am and 4:15 pm. Student A (left) chose to break down the elapsed time across a number line, while Student B (right) used a clock face to show the progression of time. While results are not yet available for this work, we present this research design as proof of concept for a basic method of testing the effectiveness of learnersourced contributions. This design can easily be scaled to consider a much larger crowd, comparing numerous instances of elaborate feedback from student users. Of course, teachers would need to abide by the necessary permission requirements at the school level in order for students to be recorded in the classroom environment, but we see this as a minor obstacle within learnersourcing feedback.

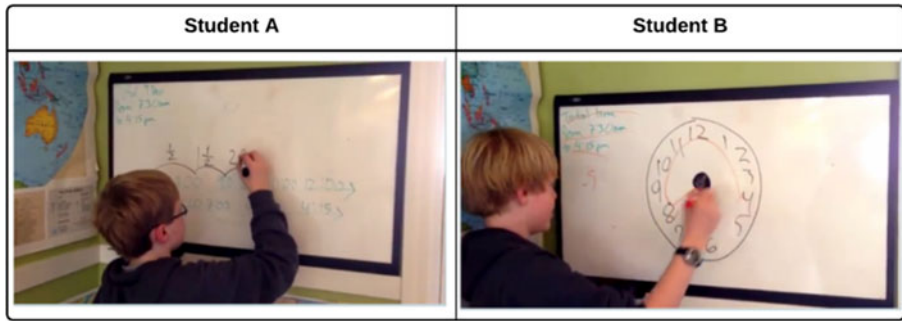














Fig. 10 Student A solves an elapsed time problem using a method based on measuring steps between chunks of time that have passed. Student B solves the same problem using a method based on the way hands move around a clock face. Through a randomized controlled design, we hope to determine which feedback instance leads to greater student learning

A second, more elaborate design, has also been implemented to examine the quality and effectiveness of learnersourced feedback provided as *hints*. Students in an AP Chemistry class were randomly assigned problem sets on two unrelated topics following an AB crossover design. For the first topic the student experienced, they were required to show and explain their work. For the second topic, they were simply required to provide an answer. Thus, half of the sample created explanations for Topic A and provided answers for Topic B, while the other half created explanations for Topic B and provided answers for Topic A. Before the crossover, the strongest student explanations were selected by the teacher and made available to students as they attempted to provide answers for the alternate topic. As a control, a portion of students continued to receive the text hints traditionally provided by ASSISTments. A posttest was to be conducted to determine if writing explanations lead to better learning than providing answers alone and to determine whether learnersourced contributions lead to better learning than traditional text hints. For this iteration of the study, the posttest was not ultimately assigned to the sample population and therefore results were not substantiated. However, this study served as a basis for a design that can be reused to assess the quality and usefulness of learnersourced feedback.

Collective vs. Individual Teachers’ Contributions: “Patchwork Quilts” of Feedback

In 2014, seven teachers were funded by a grant initiative to increase the amount of video feedback within ASSISTments with the intention of establishing a bank of crowdsourced explanations to use in various randomized controlled research designs. Numerous studies are currently running in ASSISTments to investigate the effects of being taught by multiple teachers. Versions of what we term the “Patchwork Quilt” design have been built to present feedback videos from mixes of two or more teachers across a set of problems. For instance, in the design shown in Table 1, students are randomly assigned to a “Teacher” condition or to the control (displayed across rows). Three teachers (Teacher A, B, & C) were asked to create video feedback for three isomorphic problems (displayed across columns). Videos from Teacher A and Teacher B are shown in Fig. 11 for comparison. Although both teachers approached video

Table 1 The “Patchwork Quilt” design. Across three isomorphic problems, students are randomly assigned to receive video feedback from Teacher A, Teacher B, Teacher C, or a mix of all three teachers. The control condition features the text feedback traditionally provided within ASSISTments for comparison. Through a posttest following problem 3, differences in learning outcomes can be measured to determine the effectiveness of teachersourced content

	Problem 1	Problem 2	Problem 3
Teacher A			
Teacher B			
Teacher C			
All Teachers			
Control (Text)	<div style="border: 1px solid black; padding: 10px; background-color: #ffffcc;"> <p>The least common denominator is 40 because it is the smallest number into which both 8 and 10 divide evenly.</p> $\frac{5}{8} + \frac{7}{10}$ <p>Find equivalent fractions using the denominator 40.</p> <p style="text-align: right;">Comment on this hint</p> </div>		

creation using pencasts, the formats are noticeably different. Just as the design in the previous section explored the effectiveness of learnersourced contributions, this randomized controlled design explores the effectiveness of teachersourced feedback at a small scale ($n = 3$). Comparison of learning outcomes will be possible in these designs by implementing posttests following instances of crowdsourced feedback.

Motivating Participation in Learnersourcing

As we crowdsource explanations from students to enrich the content in ASSISTments, it is necessary to ask why a student would want to provide an explanation as the time and

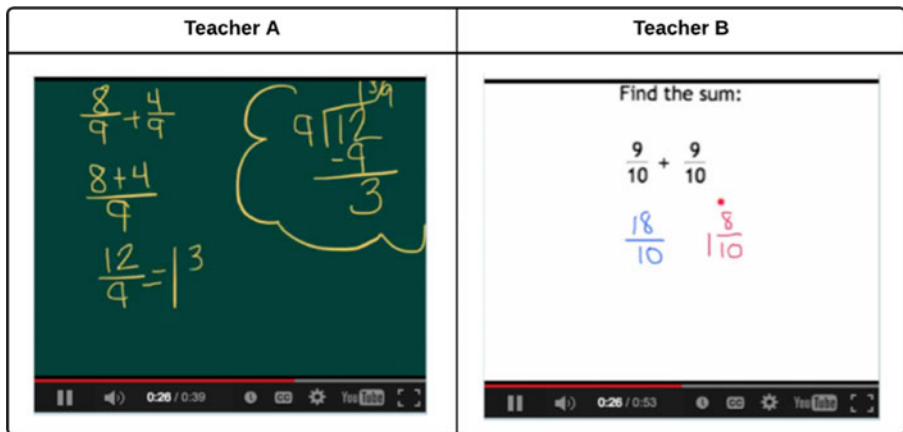


Fig. 11 Videos created by Teacher A and Teacher B on isomorphic problems. Both questions feature fraction addition with common denominators. The teachers use different teaching approaches and slightly different video styles to present elaborate feedback

effort required is nontrivial. Within ASSISTments' implementation of learnersourcing, we have devised several methods to incentivize student participation. Our goal is to provide incentives that encourage students to supply a high volume of high quality explanations.

The simplest 'incentive' is to do nothing other than provide students the ability to create explanations and notify them that their contributions will be shared with their peers. This approach would be voluntary and would not require a reward structure. We believe that this approach would show limited success simply based on altruism.

A stronger incentive would require the design of a rating system that students could use to rate the contributions of their peers. Students that write high quality explanations would be highly rated by their peers, while those that write low quality explanations would receive lower ratings. This approach to incentivization is also voluntary in nature and implements only a social reward structure. This approach also allows room for error as it calls on crowd collaboration to designate contribution quality (Saxton et al. 2013) and it may present social risks if contributions are directly linked to students.

Another potential incentivization is to get students to explain their mistakes by providing an extra opportunity to earn credit within an assignment. While this approach would source a higher volume of feedback messages, it could lower the quality of contributions. How do we create an environment where students both want to provide feedback and are likely to provide useful feedback? One of the basic types of problem sets within ASSISTments is the Skill Builder. Skill Builders are assignments that have an exit requirement of n problems right in a row, with 3 problems set as the default. A common complaint from students who complete Skill Builder assignments is that they will answer two problems correctly, and then make a mistake on the third, thereby resetting their progress. Data mining has suggested that a student that gets two consecutive correct answers has an 84 % chance of correctly answering the third question (Van Inwegen et al. 2015). A slight difference exists between the student that accurately answers the first two questions in the assignment (88.5 % chance of accurately answering the third problem) and the student that achieves two consecutive

correct answers at a later point in their assignment (82.6 % chance of accurately answering the next problem). With these probabilities in mind, problem sets can be manipulated to allow students a second chance to answer a third consecutive question, at the cost of providing a *mistake message* to assist their peers. The goal behind adaptive learning technologies providing second chance problems in the context of learnersourcing also benefits evaluating the strength of contributions: if the student is able to answer the “redo,” there is a high probability that the feedback they provide will be useful to other students. The student was able to self-correct and explain their misconception. On the other hand, students that answer the “redo” incorrectly are not likely to provide useful feedback. Performance on a second chance problem can therefore serve as an initial curator for weeding out feedback content that has low efficacy or accuracy. Providing students an opportunity to learn from their mistakes has been shown to improve learning (Attali and Powers 2010), and the process serves as a viable way to elicit feedback from students in the context of assignments within adaptive learning technologies.

A more mandatory incentive is to *force* students to write explanations for the problems that they solve as part of a predefined grading rationale. Although this may seem demanding, it is traditional practice in most classrooms; teachers almost always require students to show their problem solving process in order to receive full credit for their assignments. Without proof that a student has worked through the problem on their own, it is impossible to know how he or she arrived at an answer and whether or not they simply copied a peer. This incentivization integrates a student’s normal workflow with the creation of explanations. Expanding on this idea, teachers could have the ability to edit and improve upon student explanations by grading the work, tapping into traditional teacher grading workflow.

Regardless of incentivization, in order to implement learnersourcing on a larger scale in our world (i.e., from all students, across all content within ASSISTments), it is still necessary for our team to design a proper crowdsourcing infrastructure for use by teachers and students. This goal sparked the birth of PeerASSIST, a feature currently being developed to allow students to provide assistance to their peers through explanations and mistakes messages.

IMPLEMENTING LEARNERSOURCED FEEDBACK WITHIN ASSISTMENTS: PEERASSIST

Since ASSISTments began, the standard method of instruction has included hints or scaffolding to help students solve problems or to break problems down into smaller steps. This approach will be overhauled by the implementation of PeerASSIST. The envisioned workflow of the new feature, as shown in Fig. 12, begins as an option on the tutor interface; a button that allows the student to “Explain How to Solve This Problem.” When the student clicks on this button, an input window opens prompting the student for feedback. The content generated by the student might be a worked example of the problem, an explanation regarding the solution or a common wrong answer, hints regarding the proper approach, or even a motivational message to encourage their peer. When the student submits his or her feedback, it is linked to the current problem and sent to the ASSISTments database. When another student in

the same class begins the problem, an additional option will be added to the second student's tutor interface that will "Show My Classmate's Explanation." If the student clicks on this button, PeerASSIST will randomly provide a piece of student generated feedback for that problem (it is possible that problems would accumulate multiple explanations, some better than others, that could be tested for efficacy and accuracy through random provision). Current design protocol does not allow a student to ask for peer assistance more than once per problem. However, the student can default to traditional ASSISTments assistance (hints or scaffolding) that exist for the problem.

Within PeerASSIST, students will also be able to voice whether or not the hints provided by their peers are helpful. Each instance of peer feedback will include "Like" and "Dislike" buttons, allowing the user to judge the efficacy and accuracy of the feedback. There will also be a "Report" button, allowing students to flag inappropriate content within peer-generated feedback to isolate that piece of content for potential removal from the system. If an instance of feedback is reported by more than one student, it will automatically be removed from the pool of explanations linked to that problem. Teachers will also be able to review and veto PeerASSIST feedback generated by their students on a page specifically designed for feedback management.

The remaining issue that exists within PeerASSIST is determining which explanation to display if a problem has multiple instances of student generated feedback. An obvious approach would be to randomly select an explanation to use each time a student requests peer assistance (much like the randomized controlled experimentation

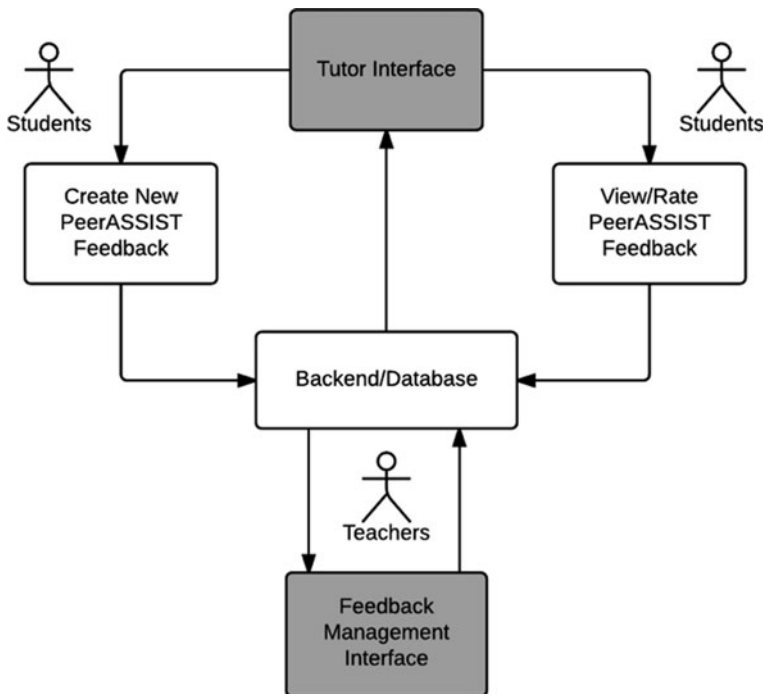


Fig. 12 The PeerASSIST data flow. Students generate feedback for other students. Feedback is linked to a particular problem and provided randomly to students that subsequently struggle with the same problem. Students are able to judge the feedback provided by their peers, and teachers are able to manage feedback created by their students using a management interface

already presented). This approach would be easy to implement and explain. However, if a PeerASSIST explanation has been “Disliked” many times, there is little reason to continue to display that contribution. Further, the information linked to each PeerASSIST explanation has the potential go beyond “Likes” and “Dislikes.” Certain researchers may be more interested in learning specific outcomes for specific instances of feedback. Thus, the system must rely on an approach that will explore the learning outcomes brought about by student-generated feedback while supplying students the best assistance available.

Algorithms for Evaluating Crowdsourced Contributions

Once feedback content has been sourced, how do we deem explanations as effective? The solution is not to examine how much the explanation helps the student through the question that he or she is struggling with, but rather to consider increases in the probability that the student answers their next problem accurately, on their first attempt, without any help. This problem of managerial control is not specific to our domain and has existed for a long time in the design of experiments. In a general context the question becomes, “How many samples should we draw and which populations should the samples be drawn from?” This question was originally proposed by Herbert Robbins in his landmark paper on sequential design (Robbins 1952). Sequential design of experiments occurs when the sample size is not predetermined but is a function of the samples themselves, as opposed to being fixed before an experiment is conducted.

There are several advantages to using sequential design. Sequential design allows for an experiment to use a fewer number of samples and allows for the experiment to end earlier. Resources such as time, money, and the number of samples required are saved. Another advantage to this approach is that if a particular condition in an experiment is detrimental, it can be avoided more efficiently. This often occurs in medical trials where a treatment is ultimately found to be harmful (Wegscheider 1998). There is no reason to continue providing a harmful treatment and it is essentially unethical. Using sequential design of experiments minimizes and prevents the undue provision of harmful treatment. However, a disadvantage of sequential design is that constant significance testing throughout the course of the experiment can result in high Type-I error rates (although this can be prevented through various forms of error correction).

The sequential design problem is more commonly known as the multi-armed bandit story. Multi-armed bandits are presented when a person enters a casino to play a slot machine and each potential machine has a different payout rate, as depicted in Fig. 13. The player needs to determine which machine’s lever (or “arm”) to pull that will provide the greatest payout rate in order to maximize his or her profits. In this scenario, slot machines have earned the term “bandits” because regardless of payout rate, they essentially steal money from the player (Lai and Robbins 1985). This problem is also known as the exploration/exploitation trade-off in the area of reinforcement learning. In this context, the gambler needs to explore various slot machines to determine which machine has the best payout, but must also exploit the machine with the best-known payout rate. Considering sequential design, the number of populations is equivalent to

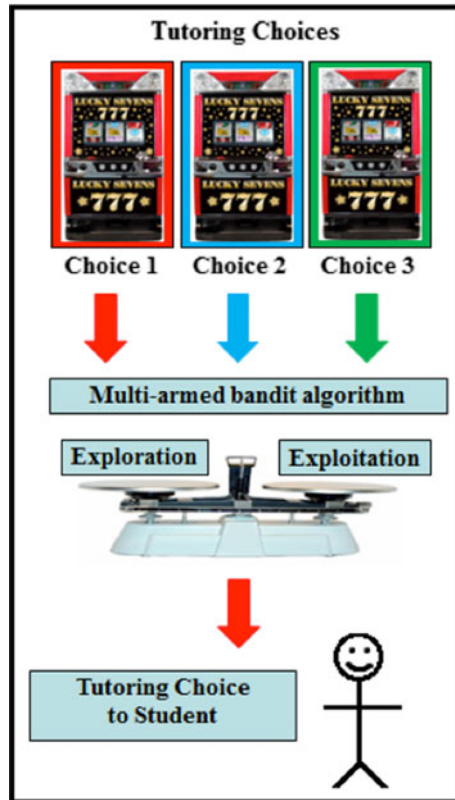


Fig. 13 An example of how a multi-armed bandit algorithm can be used when crowdsourcing student explanations. In this example there are three slot machines representing three different student-generated tutoring strategies. A multi-armed bandit algorithm is run balancing exploration and exploitation to determine which of the three tutoring strategies is given to the next student

the number of slot machine arms that can be pulled. A sample from a chosen population is analogous to a pull on a chosen arm of a slot machine. In the context of learnersourcing within ASSISTments, the pool of content available to assign to students represents these populations (arms to pull) and a sample from the population is equivalent to assigning a piece of content to a student.

It is important that we use sequential design when assigning content to students for several reasons. The first and most important reason is to quickly filter out “bad feedback” content while exposing as few students as possible. Aside from malicious or purely erroneous content, “bad feedback” would be considered any content that results in unnecessary confusion or misinformation, which can be detected by measures of how well students perform on the next problem following feedback. It would be unethical to use design types in which we would continue to expose children to content known to be “bad.” The use of sequential design will also allow us to conduct experiments in which we do not know the amount of content or the number of students a priori. This versatility is essential in order to conduct experiments in a crowdsourcing environment, where new content and new students are continually entering the system.

Lessons Learned and a Call to the Community

Beyond the experiments presented here, additional proof of concept work has been conducted to better understand the complexities and consequences of teachersourcing and learnersourcing feedback within ASSISTments. The studies presented herein, many of which are still in progress, were not included for the consideration of particularly significant or null findings, but rather, to present the direction in which our team is moving. Many of the remaining concerns about the implementation of crowdsourcing can be explored as research questions that we pose to the AIED community for exploration across adaptive learning technologies in the coming years:

1. How do we ensure the accuracy of learnersourced feedback?
2. What is the efficacy of learnersourced feedback?
3. Are students willing to spend their time generating feedback for other students?
4. Are students willing to use feedback that has been generated by a peer?
5. How do we ensure the accuracy of teachersourced feedback?
6. What is the efficacy of teachersourced feedback?
7. Can crowdsourcing be implemented as an effective use of teachers and students time?

This is by no means an exhaustive list for the community's consideration, and it is likely fair to say that a range of possible outcomes will exist for each of these concerns spanning content domains, age ranges, and types of learners. It is possible that crowdsourcing will be useful and/or successful in certain scenarios but not in others. It is also possible that crowdsourcing will prove a more viable strategy for particular adaptive learning technologies. As we have presented here, we suggest that the community considers an approach to crowdsourcing (specifically learnersourcing) that simplifies the complex task of content creation into the simple task of having students "show their work." The Common Core Standards for Mathematics (NGACBP and CCSSO 2010) require that students are able to explain their reasoning in addition to answering questions. Thus, more and more, students are providing written explanations of their work as part of normal instruction. As exemplified by our proof of concept study designs, ASSISTments has the potential to gather teachersourced and learnersourced contributions and rigorously test their effectiveness. While our platform is somewhat novel in this regard, and much of our work is still underway, other adaptive learning technologies will also serve as excellent resources for studying crowdsourced content in the coming 25 years of AIED research.

Closing Thoughts

While predicting the future is an impossible task, considering the trends in amongst domains it is safe to say that the future of adaptive learning will be strongly driven by the crowd. Current technologies that rely on the crowd for expert knowledge and system expansion are prevailing, and the trend will soon spill over into educational domains. As such, we have presented our plan for bringing ASSISTments into the next

quarter century while highlighting the complexities of crowdsourcing for consideration by the AIED community.

Especially in the realm of mathematics, students around the world have historically been required to ‘show their work’ when completing homework or answering test problems. In the age of adaptive learning technologies, these worked examples can be captured and used as powerful feedback for other, subsequently struggling students. This practice would benefit all parties: explaining a solution allows the student to solidify his or her understanding of the problem, receiving peer explanation increases motivation and employs proper solution strategies in struggling students, and the adaptive learning platform experiences perpetual evolution and expanse. Perhaps most intriguing, all of this promise stems from only minor adjustments to the workflow that is already taking place in classrooms around the world, as teachers and students use online learning platforms like ASSISTments to conduct day-to-day learning activities. Simple steps can be taken to bring adaptive learning technologies to the next level: simplifying the collection of video feedback, running randomized controlled experiments to understand what works, building out an infrastructure like PeerASSIST to capture the explanations that students are already preparing, and employing sequential design to deliver the right feedback to the right students at the right times. The crowd can be a limitless force and it is better to have teachers and students on our side and ultimately working with us rather than against or alongside us.

Harnessing the knowledge of the crowd will enhance adaptive learning platforms moving forward. The next 25 years within the AIED community should be marked by research that brings underlying fields together to understand best practices, establish collaborative scientific tools for the community, and integrate users through content creation and delivery. The current application of stringent research methodologies to improve learning outcomes is severely lagging what the educational research community requires. The inclusion of sound experimental design and crowdsourced content within adaptive learning systems has the potential to simultaneously produce large-scale systemic change for education reform, while advancing the collaborative knowledge of those researching AIED.

Acknowledgments We would like to thank all the 150+ students at WPI that helped us create ASSISTments, and then do research with it. We would also like to thank NSF (0231773, 1109483, 0448319, 0742503, 1031398, 1440753, 1252297, 1316736, and 1535428), GAANN, The Spencer Foundation, The US Department of Education (R305K03140, R305A07440, R305C100024 and R305A120125), The Office of Naval Research, the Bill and Melinda Gates Foundation, and EDUCAUSE for funding the research described here. A complete list of funders is here: <http://www.aboutus.assistments.org/partners.php>. Thanks to S.O. & L.P.B.O. We would also like to thank the editor of this Special Issue and the three anonymous reviewers that helped to strengthen our work prior to publication.

References

- Aleahmad, T., Aleven, V., and Kraut, R. (2008). Open community authoring of targeted worked example problems. In Woolf, Aimeur, Nkambou, & Lajoie (eds) *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*. Springer-Verlag. pp. 216–227.
- Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2012, August). Discovering value from community activity on focused question answering sites: a case study of stack overflow. In

- Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 850–858). ACM.
- Askey, R. (1999). Knowing and teaching elementary mathematics. *American Educator*, 23, 6–13.
- Attali, Y., & Powers, D. (2010). Immediate feedback and opportunity to revise answers to open-ended questions. *Educational and Psychological Measurement*, 70(1), 22–35.
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213–238.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679.
- Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4), 86–96.
- Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189–200.
- Floryan, M., & Woolf, B. P. (2013, January). Authoring Expert Knowledge Bases for Intelligent Tutors through Crowdsourcing. In *Artificial Intelligence in Education* (pp. 640–643). Springer Berlin Heidelberg.
- Franklin, M. J., Kossmann, D., Kraska, T., Ramesh, S., & Xin, R. (2011, June). CrowdDB: answering queries with crowdsourcing. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data* (pp. 61–72). ACM.
- Griswold, A. (2014). How Luis Von Ahn Turned Countless Hours of Mindless Activity Into Something Valuable. *Business Insider*, *Strategy*. Retrieved on November 14, 2015, from <http://www.businessinsider.com/luis-von-ahn-creator-of-duolingo-recaptcha-2014-3>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Heffernan, N., & Heffernan, C. (2014). The ASSISTments ecosystem: building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24(4), 470–497.
- Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, 14(6), 1–4. Retrieved November 14, 2015 from <http://www.wired.com/2006/06/crowds/>
- Howe, J. (2008). *Crowdsourcing: Why the power of crowd is driving the future of business*. New York,: Crown Business.
- Kim, J. (2015). Learnersourcing: Improving Learning with Collective Learner Activity. MIT PhD Thesis. Retrieved from <http://juhokim.com/files/JuhoKim-Thesis.pdf>.
- Kehrer, P., Kelly, K. & Heffernan, N. (2013). Does immediate feedback while doing homework improve learning. In Boonthum-Denecke, Youngblood (Eds), Proceedings of the twenty-sixth international Florida artificial intelligence research society conference. AAAI Press 2013. pp 542–545.
- Kelly, K., Heffernan, N., Heffernan, C., Goldman, S., Pellegrino, J., & Goldstein, D. S. (2013). Estimating the effect of web-based homework. In Lane, Yacef, Motow & Pavlik (Eds) *The Artificial Intelligence in Education Conference*. Springer-Verlag. pp. 824–827.
- Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J.,... & Horton, J. (2013, February). The future of crowd work. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. pp. 1301–1318. ACM.
- Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., et al. (2015). Peer and self assessment in massive online classes. In H. Plattner, C. Meinel & L. Leifer (Eds.), *Design thinking research* (pp. 131–168). Springer-Verlag Berlin Heidelberg: New York.
- Lai, T. L., & Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1), 4–22.
- Ma, L. (1999). *Knowing and teaching elementary mathematics*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Malone, T. W., & Bernstein, M. S. (2015). *Handbook of collective intelligence*. Massachusetts Institute of Technology. The MIT Press: Cambridge, MA.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: the next frontier for innovation, competition, and productivity*. McKinsey Global Institute. Retrieved from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
- McAfee, A. P. (2006). Enterprise 2.0: the dawn of emergent collaboration. *MIT Sloan Management Review*, 47(3), 21–28.
- Mendicino, M., Razzaq, L. & Heffernan, N. T. (2009). Improving Learning from Homework Using Intelligent Tutoring Systems. *Journal of Research on Technology in Education (JRTE)*. 41(3), 331–346.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). Common core state standards for mathematics. National governors association center for best practices, Council of Chief State School Officers, Washington D.C.

- Organisciak, P., Teevan, J., Dumais, S., Miller, R. C., & Kalai, A. T. (2014, May). A Crowd of Your Own: Crowdsourcing for On-Demand Personalization. In the *Second AAAI Conference on Human Computation and Crowdsourcing*.
- Ostrow, K. S. & Heffernan, N. T. (2014). Testing the Multimedia Principle in the Real World: A Comparison of Video vs. Text Feedback in Authentic Middle School Math Assignments. In Stamper, J., Pardos, Z., Mavrikis, M., McLaren, B.M. (eds.) *Proceedings of the 7th International Conference on Educational Data Mining*. pp. 296–299.
- Porcello, D., & Hsi, S. (2013). Crowdsourcing and curating online education resources. *Science*, 341(6143), 240–241.
- Retelny, D., Robaszekiewicz, 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.
- Robbins, H. (1952). Some aspects of the sequential design of experiments. In *Herbert Robbins Selected Papers* (pp. 169–177). New York: Springer.
- Saxton, G. D., Oh, O., & Kishore, R. (2013). Rules of crowdsourcing: models, issues, and systems of control. *Information Systems Management.*, 30(1), 2–20.
- Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, 16(4), 475–5223.
- Selent, D. & Heffernan, N. T. (2015) When More Intelligent Tutoring in the Form of Buggy Messages Does Not Help. In Conati, Heffernan, Mitrovic & Verdejo (eds) *The 17th Proceedings of the Conference on Artificial Intelligence in Education (AIED 2015)*. Springer. pp. 768–771.
- Surowiecki, J. (2004). *The wisdom of crowds*. New York: Doubleday.
- Tan, W., Blake, M. B., Saleh, I., & Dustdar, S. (2013). Social-network-sourced big data analytics. *IEEE Internet Computing*, 5, 62–69.
- Van der Kleij, F.M., Feskens, R.C.W., & Eggen, T.J.H.M. (2015). Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-Analysis. *Review of Educational Research*, AERA. 85 (4). pp. 475–511.
- Van Inwegen, E., Wang, Y., Adjei, S. & Heffernan, N.T. (2015) The Effect of the Distribution of Predictions of User Models. In Santos, Boticario, Romero, Pechenizkiy, Merceron, Mitros, Luna, Mihaescu, Moreno, Hershkovitz, Ventura, & Desmarais (eds.) *Proceedings of the 8th International Conference on Educational Data Mining (EDM 2015)*.
- Von Ahn, L., & Dabbish, L. (2008). Designing games with a purpose. *Communications of the ACM*, 51(8), 58–67.
- Von Ahn, L. (2009, July). Human computation. In *Design Automation Conference, 2009. DAC'09. 46th ACM/IEEE* pp. 418–419. IEEE.
- Wegscheider, K. (1998). Advantages and disadvantages of sequential designs. *Herzschrittmachertherapie und Elektrophysiologie*, 9(4), 248–254.
- Weir, S., Kim, J., Gajos, K., & Miller, R. (2015). Learnersourcing subgoal labels for how-to videos, *proceedings of the 18th ACM conference on computer supported cooperative work & social computing*.
- Weld, D., Adar, E., Chilton, L., Hoffmann, R., Horvitz, E., Koch, M., Landay, J., Lin, C. H., & Mausam. (2012). Personalized online education- a crowdsourcing challenge. AAAI Workshops, North America. Retrieved from <<https://www.aaai.org/ocs/index.php/WS/AAAIW12/paper/view/5306/5620>>
- Williams, J., J., Li, N., Kim, J., Whitehill, J., Maldonado, S., Pechenizkiy, M., Chu, L., & Heffernan, N. (2014). MOOClets: a framework for improving online education through experimental comparison and personalization of modules (working paper no. 2523265). Retrieved from the Social Science Research Network: <http://ssrn.com/abstract=2523265>
- Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: evidence from category learning. *Cognitive Science*, 34, 776–806.
- Williams, J. J., Krause, M., Paritosh, P., Whitehill, J., Reich, J., Kim, J., et al. (2015a). Connecting collaborative & crowd work with online education. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing* (pp. 313–318). ACM.
- Williams, J. J., Ostrow, K., Xiong, X., Glassman, E., Kim, J., Maldonado, S. G., et al. (2015b). Using and designing platforms for in vivo educational experiments. In D. M. Russell, B. Woolf & G. Kiczales (Eds), *Proceedings of the 2nd ACM Conference on Learning at Scale*, pp. 409–412.
- Zhang, J., Kong, X., Luo, R.J., Chang, Y., & Yu, P.S. (2014). NCR: A Scalable Network-Based Approach to Co-Ranking in Question-and-Answer Sites. *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management*. pp 709–718.