International Journal of Artificial Intelligence in Education Volume# (YYYY) Number
IOS Press

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

Neil Heffernan
Korinn Ostrow
Kim Kelly
Douglas Selent
Eric VanInwegen
Xiaolu Xiong
Worcester Polytechnic Institute, *100 Institute Road*, *Worcester, MA 01609*

Joseph Jay Williams
HarvardX, Harvard University, 125 Mt. Auburn Street, Cambridge, MA 02138

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. Using the context of ASSISTments, an adaptive tutoring system, we explain how interactive mathematics exercises can provide the workflow necessary for eliciting feedback contributions and evaluations of those contributions, while teachers and students use the platform in everyday education. We discuss randomized controlled experiments that are currently running within the ASSISTments platform with the goal of offering a 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, a planned infrastructure that will help ASSISTments to harness the power of the crowd. Algorithms from machine learning such as multi-armed bandits will provide a mechanism for automatic evaluation and personalization of alternative micro-designs. We feel that the future of adaptive learning technologies will be driven by the crowd, and this article offers an attempt at a road map.

Keywords. Crowdsourcing, Feedback, Adaptive Tutoring System, ASSISTments

EVOLVING ADAPTIVE LEARNING TECHNOLOGIES THROUGH CROWDSOURCED CONTRIBUTIONS

Over the next 25 years, it is our hope that adaptive learning technologies like intelligent tutoring systems will expand support for best practices in K-12 learning through rigorous experimentation to identify and implement personalized educational interventions in real world classrooms. However, we anticipate that while data will be continually used to improve these platforms, innovations in this arena will be limited by the pedagogy and micro-designs in educational content and interactions that can provide fine-grained and tailored support for learners. Still, growth rooted in best practices will be necessary to keep the field from growing stagnant.

This article considers how improvements for a perpetually evolving educational ecosystem can be solicited dynamically and at scale through *crowdsourcing* (Kittur, et al., 2013). Recent research suggests that large improvements can be produced through many small-scale and organic contributions from distributed populations of teachers and students. Users of adaptive learning technologies not only receive content via online and blended education systems and send back data about learning and interactions, but can also contribute many small changes and pedagogical innovations that underlie systemic change (Howe, 2006; Von Ahn, 2009).

Crowdsourcing

Crowdsourcing is a powerful alternative to a curriculum or pieces of content designed by single or small teams of experts (Porcello & Hsi, 2013), yet is not often taken advantage of in adaptive learning settings. The approach suggests that the whole is greater than the sum of its parts; we know more together than any single person knows alone. By sourcing contributions from users within an adaptive learning platform, it is possible to expand the breadth and diversity of available material beyond that born of just a few designers, supporting functions such as the personalization of online educational content (Organisciak, 2014; Weld, Adar & Chilton, 2012).

A prominent success story that uses this approach is Wikipedia, a free online encyclopedia that relies on crowdsourcing to author and edit content. Wikipedia has gone far beyond what was capable in previous electronic encyclopedias like Britannica by using an approach that as initially criticized and met with skepticism: a wide range of users, all free to create, edit, and flag content. In many large technological platforms, processes for crowdsourcing have provided valuable solutions (VonAhn & Dabbish, 2008), and can even accomplish tasks like programming mobile apps by crowdsourcing experts with specialized skills (Retelny et al, 2015).

A slightly different, yet still successful, crowdsourcing model is used by services like Stack Overflow and Yahoo Answers, two websites that are designed to allow users to interact and provide assistance and have shown a variety of compelling benefits (Anderson, et al., 2012). For instance, Stack Overflow is one of the top 50 most visited sites on the Internet and is used by 26 million programmers each month (http://stackexchange.com/about). Within this crowdsourcing format, any user is able to ask questions related to programming, and others in the community can provide answers. Additionally, users can 'upvote' or 'downvote' questions and answers to promote accurate and helpful content. Further, questions can be linked, marked as duplicates, flagged as inappropriate, or commented on with general responses. Stack Overflow then uses an algorithm to rank users according to the 'value' of the answers they provide, thereby helping to highlight the best answers from the 'best' community members with more efficiency.

While these are powerful and compelling uses of crowdsourcing, the concept still face challenges in the domain of education. Stack Overflow cannot outwardly measure which answers more effectively cause learning. Users might assume that the most 'upvoted' content is the most reliable, but there is no qualitative way to test users after each they read each answer to determine the learning outcome. Similarly, open authorship on sites like Wikipedia makes it very easy for users to author inaccurate content or to destroy accurate content with malicious edits. Without principled ways of evaluating the quality of contributions beyond opinion, Wikipedia faces skepticism from those in education about the reliability and the veracity of 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 a few substantial obstacles (Williams, et al., 2015a). Teachers and experts can curate and collect high quality educational resources online (Porcello & Hsi, 2013), with research showing successful attempts to author expert knowledge for intelligent tutors by using crowds of teachers (Floryan & Woolf, 2013). However, the majority of adaptive learning systems 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, 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 & CCSSO, 2010) emphasize multiple approaches to problem solving, often with varying complexity. For example, one student may use a manipulative, such as fraction pieces of a circle, to find equivalent fractions and then carry out the addition portion. Another student may take a more sophisticated approach by listing all of the 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. Any adaptive learning system that is assisting a student with this problem should know all potential approaches, know which approach is most appropriate for the student, 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 exceeding complex. 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 on one approach or method, some students may fail to grasp what they would perhaps otherwise understand when taught using a different approach.

Crowdsourcing question content and feedback material from teachers could 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., 2013). 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 to better predict the quality of questions, the helpfulness of answers, and the expertise of users (Zhang, et al., 2014).

An Alternative Approach: “Learnersourcing”

Crowdsourcing feedback does not have to stop at teachers. We believe that students can provide quality worked examples of how they solved 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 as feedback for the benefit of future students (Kulkarni, et al., 2014). 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 may also be beneficial to students, if pedagogically useful activities like prompts for self-explanation are used to elicit student contributions (Williams & 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, Kim, Gajos & Miller, 2015). Weir, et al. (2015) show that this type of learnersourcing workflow can produce outlines for videos that lay out subgoals for learning in a way that is indistinguishable from outlines produced painstakingly by experts.

In theory, crowdsourcing could play an integral role in the future of adaptive learning. However, the questions surrounding the actual practice of crowdsourcing feedback within and adaptive tutor are complex. What type of a system must exist for crowdsourcing to be easy and natural to 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 questions as well as others as we discuss the intended future of harnessing the crowd within ASSISTments.

IMPLEMENTING CROWDSOURCING WITHIN ASSISTMENTS

The present article considers how we hope to extend the ASSISTments platform to enable large-scale improvement through crowdsourcing from teachers and students. ASSISTments is an online tutoring system 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 instructors to do what instructors do best. In ASSISTments, instructors can author question content to assign to students, or select from open libraries of pre-built material.

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 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 & 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 extensive knowledge of computer programming. Recent NSF funding has allowed for researchers from around the country to design and implement studies within the system, moving the platform towards acceptance as a shared scientific instrument for educational research. There is an immense community of researchers who can take advantage of such systems; last year, more than 14,000 researchers gathered at the American Educational Research Association conference alone.

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 the platform. 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

Students receive a variety of support within the ASSISTments platform. 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). Next, questions may include *mistake messages* created by the author of the problem, or sourced from teachers and classes that have discovered 'common wrong answers.' These messages are automatically delivered to the student in response to a particular mistake, as shown in Figure 1.

Additionally, explanatory feedback can come in the form of *hints* that must be requested by the student and are presented sequentially. Hints are typically presented with increasing specificity before presenting the student with the correct answer (via the Bottom Out Hint), allowing the student to move on to the next problem in the assignment rather than becoming indefinitely stuck. Alternatively, ASSISTments offers a form of explanatory 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 the problem be broken down into steps. A comparison of hint feedback and scaffolding is presented in Figure 2.

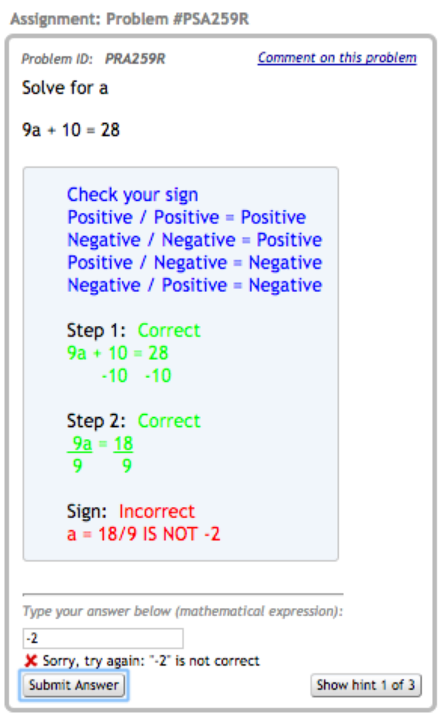


Figure 1. An example of a mistake message. This type of feedback responds with tailored information that pinpoints exactly where the student made a mistake. If the student is unable to arrive at the correct answer with this guidance, standard hints are also available.

While students benefit from enriched 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 Figure 3. 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 report in the classroom as a learning support because it provides actionable data. The report can be anonymized, as shown in Figure 3, 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 Figure 3.

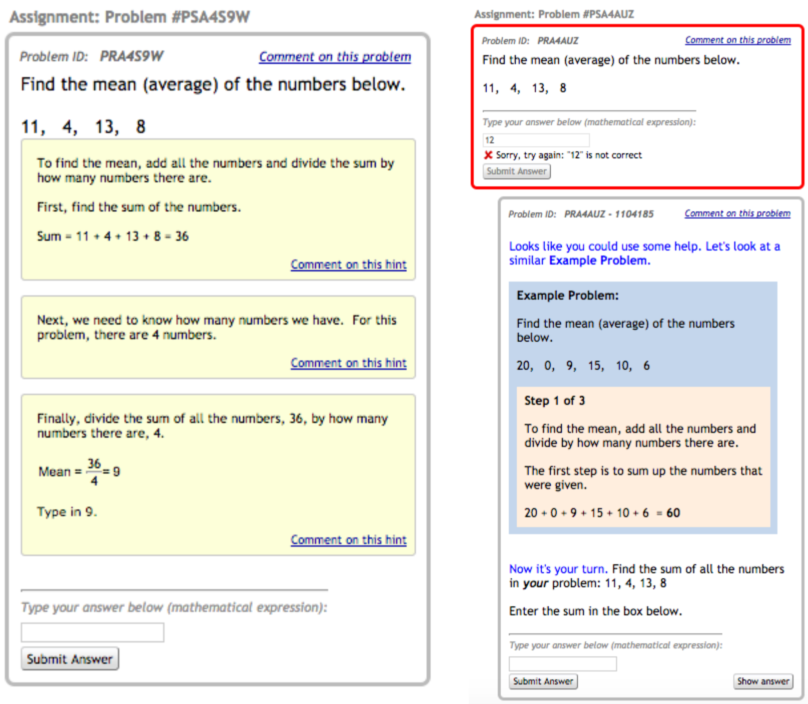


Figure 2. A comparison of Hints and a Scaffold within an identical problem. Note that three hints are shown on the left, as requested by the student. On the right, the student provided an incorrect response and was automatically given a scaffold with a worked example on how to solve a similar problem. 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.

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

Figure 3. 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 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.

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 Figure 3, 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. The data suggests that this problem is worthy of class discussion. There is also a "+feedback" link available for instructors to write a mistake message to 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 suggests that it is possible to crowdsource feedback from teachers and students within systems like ASSISTments. The benefits can be both immediate (i.e., the students learn to explain their work and pinpoint misconceptions) and long lasting (i.e., students who attempt this problem in the future now have enriched feedback that targets their misconceptions).

The Potential Role of Video in Crowdsourced Contributions

Within ASSISTments and 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 teachers and professors who were once a powerful resource to a limited audience to now impact any student who is willing to learn. These lectures are reaching some very remote parts of the world and are being watched by those who would otherwise never have the opportunity to attend a world-class university. The universal power of the video lecture suggests that there is clearly a 'time for telling' (Schwartz & Bransford, 1998), and that eager learners can use this technology to access the knowledge of experts and understand the bulk of the story.

However, many learners require more than just the story; they need reinforcement and support while practicing what they have learned. We advocate the use of video beyond the lecture and into the realm of short tutorial strategies. After all, lecturing is only a small portion of an instructor's job that can be captured on video. By only focusing on the lecture, thousands lose out on unique explanations and extra help that can be provided through individual tutoring. The greatest teachers spend a large portion of their time tailoring instruction to a struggling student's individual needs. The future of adaptive learning technology needs to consider the problem of capturing and delivering these just-in-time supports, for students working in class and at home, and video may 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 the 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 using our system, and keeping those videos directly connected to particular problems. The ASSISTments iPad app has a built-in feature that allows users to record Khan Academy style "pen casts" (a visual walkthrough of the problem with a voice over explanation) while working on the problem. Ideally, the app will then allow for the recording to be uploaded to YouTube and stored as a specific tutorial strategy for that problem. This linking system is still under development. However, 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 pen cast or by recording themselves working through a problem on a white board, uploading the content to a video server, and linking to the content in problems or feedback that they have authored. In the past year, ASSISTments has witnessed the use of videos as explanations, 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 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 his iPad. He notices that one of the students answered a problem incorrectly and that her solution strategy includes a misconception about the problem. While tutoring her through the mistake, he 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. The recording includes both an auditory explanation and the visual walkthrough of the problem as he works through the student's work on the iPad. The explanation takes about 20 seconds to provide, but because it has been captured, it must only be provided once. Following this instance of tutoring alongside the student, the instructor quickly uploads his video to YouTube and links the material to the solved problem. Within two minutes, another student at the extra help session reaches the problem and tries to solve the problem using the same misconception. The newly uploaded feedback video is provided as a mistake message and the student is able to correct his own error by watching the video and working through the problem again. Meanwhile, the tutor is able to help a third student on a different problem, rather than having to provide that first type of help repetitively.

This use case is the vision that ASSISTments holds for the future of adaptive tutoring. The process does not exclude the instructor from the feedback process, but rather harnesses the power of explanations given *once* to help students across *multiple* instances. How can we convince teachers that the process of collecting feedback and building a library of explanations is useful? Suppose that the goal is to collect feedback uploads from various teachers (and possibly even students) to expand the library of explanations 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 3,000 questions per book, it leaves a total of 90,000 questions requiring feedback. Explanations to many of these problems have already been generated by high quality instructors across the nation, they are just being lost on individual students rather than recorded and banked for later use by all students. If each math instructor in the nation 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.

GUIDING THE CROWD

We anticipate that the future of adaptive learning should examine mechanisms for interactivity in eliciting contributions at scale, or *directed* crowdsourcing (Howe, 2006). This can be achieved by extending ASSISTments' existing commenting infrastructure, which interacts with student users. We hope to leverage a similar system of interactivity to garner feedback from student users, or to allow users to 'show their work' on problems.

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 instructor and to the problem's author, as shown in Figure 4. Teachers are able to act on comments by helping students individually, while content authors are able to use the comments (which have been anonymized) to enhance the quality of their questions. Already within the system, students have written 80,000 comments on roughly 20 million problems solved over the last five years. The commenting infrastructure includes a pull down menu as a sentence starter (see Figure 5) as well as a text field where students can write their comment.

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 SEb which is 0.10101 i got .727 and it is not the right answer. (Flag this comment)	Test Drive - Edit - Delete	PRAJ9JN

Figure 4. Comments from users on specific problems. Some of the comments are routine while others give the author genuinely helpful information.

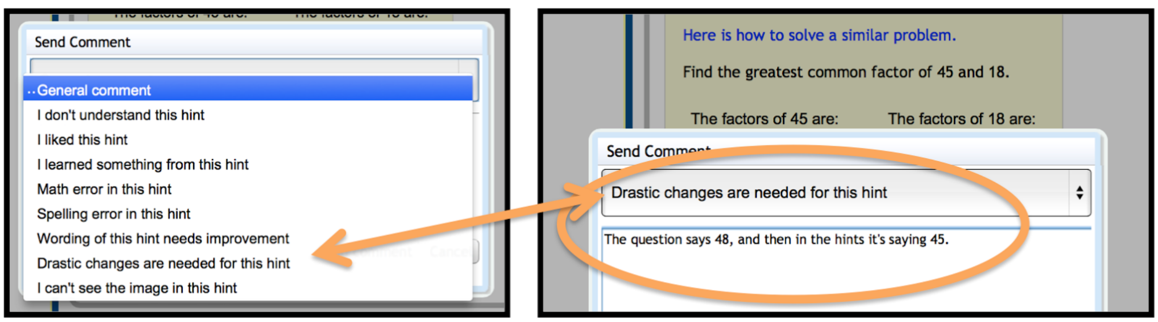


Figure 5. This figure shows the current comment feature. Students who want to leave a comment select from the pull down menu first, then type in a comment.

The goal for the future of ASSISTments is to use a similar infrastructure to crowd source tutoring feedback from teachers and students. The vision is a new configuration tools coming together for a crowdsourcing platform. We hope to harness the power of YouTube, or similar video servers, alongside ASSISTments problem content while developing a K-armed bandit algorithm to aid in feedback delivery. This new approach will build largely off of functionalities that already exist within the ASSISTments platform, but it will allow professors, instructors, undergraduate teaching-assistants, and possibly even students to efficiently create and add tutoring videos to the system.

Figure 6 depicts a mockup of what we envision the crowdsourcing interface to look like for instructors being asked to create video tutorial strategies. Teachers will be informed of the common wrong answers and the percentage of students who shared the misconception. In Figure 6, 30% of the students responded "-20" although the correct answer solving for C would be "-16." Therefore, the instructor can record a YouTube video with tutoring specific to the error, "It looks like you subtracted the 2 from both sides when you should have added the 2 to both sides." Figure 6 also shows that the instructor uploaded a different video link for the 22% of students who responded with "20" as their answer. At the bottom, an "explanation" section allows the instructor to add video that features a worked example of the problem, alongside an encouraging message. Students who make common wrong answers will receive tailored feedback, while students that simply ask for a hint will receive this '"explanation" video.

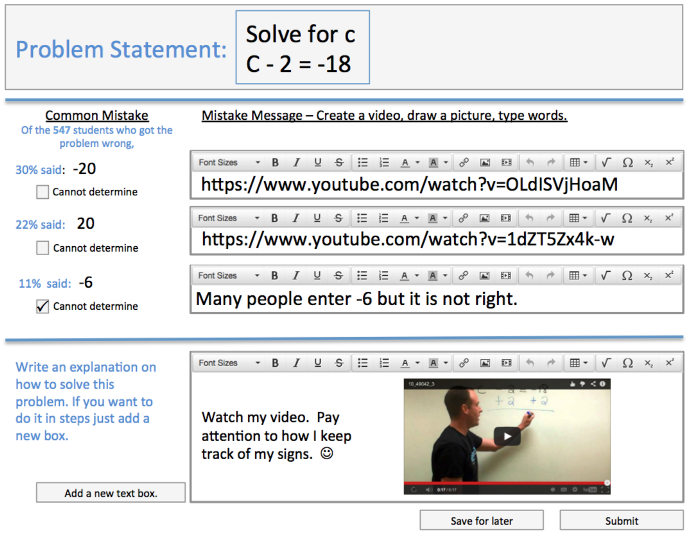


Figure 6. A mockup of one infrastructure for crowdsourcing tutorial strategies from instructors. The problem is stated at the top. The teacher is able to create video or text feedback tailored to the three most common wrong answers and is also provided the option to create a more generic explanation.

While this schematic provides insight into how the actual process of crowdsourcing could work within an adaptive tutor, we are left with questions surrounding how to learn which videos are most useful, for which students, and under what contexts? The next section discusses a variety of randomized controlled trials that have been conducted within ASSISTments in an attempt to theorize on some of these important issues.

EVALUATING CROWDSOURCED CONTENT VIA RANDOMIZED CONTROLLED EXPERIMENTS

What rigorous options are available to evaluating the contributions from a 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 could therefore be used 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 on ASSISTments that have already identified large positive effects on student learning, by varying factors like feedback on homework from correctness feedback to full tutorials provided through scaffolding (Mendicino, Razzaq & Heffernan, 2009; Kelly, et al., 2013; Kehrer, Kelly & Heffernan, 2013). A series of similar experiments are currently serving as proof of concept for crowdsourcing text and video feedback from teachers and students. Last summer, seven teachers were funded by a grant initiative to increase the amount of video feedback within the system, with the intention of running the randomized controlled experiments discussed here while establishing an initial bank of crowdsourced explanations.

Comparing Video Feedback to Business as Usual

Ostrow & 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 Figure 7, feedback was matched across medium, with video comprised of the lead researcher delivering each step of tutoring feedback as a tutor. All students received the same set of questions in a variety of orders, allowing for the opportunity for all participants to receive both text and video feedback during the course of the assignment. Students 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. Results from an analysis of 89 students who completed the assignment and were able to access video content revealed that video feedback led to near significant increases in student's accuracy on the next problem. Students spent significantly longer consuming video feedback but answered their next questions more efficiently. 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 designed and implemented in an attempt to replicate these findings; analyses are not yet available.

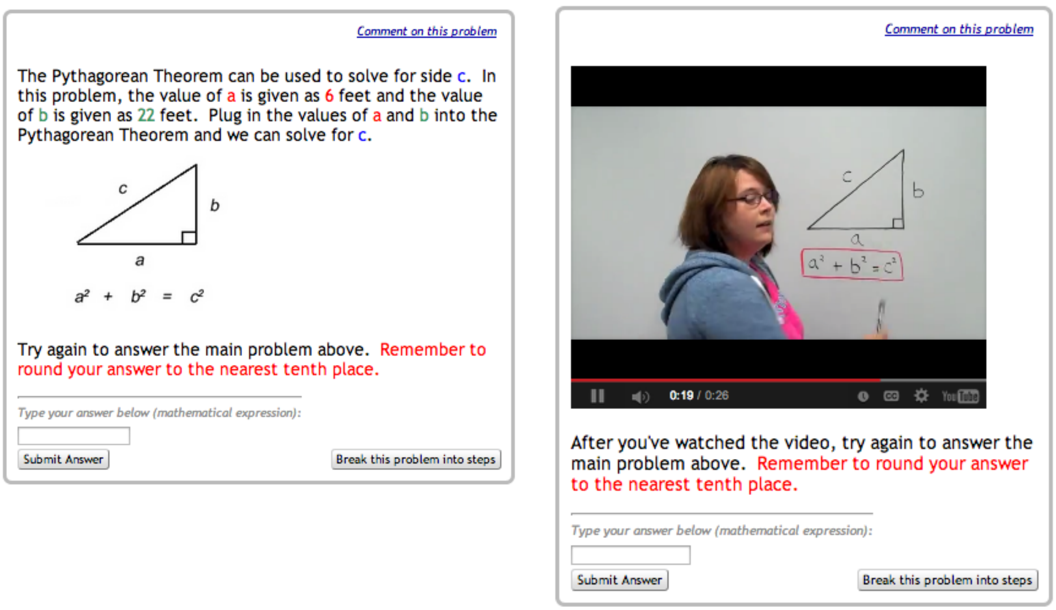


Figure 7. Text and Video Feedback conditions as experienced by students, *Testing the Multimedia Principle in the Real World: A Comparison of Video vs. Text Feedback in Authentic Middle School Math* (Ostrow & Heffernan, 2014). Isomorphic problems featured matched content feedback across mediums.

Comparing Contributions from Different Teachers: Proof of Concept

Selent & Heffernan (2015) took video feedback a step further to try to understand the potential benefits of crowd sourced mistake messages made by teachers. These messages were made by a teacher who now works as part of the ASSISTments team, following a structure similar to that depicted earlier in Figure 6 in an attempt to determine the effectiveness of the approach. The goal of this research 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 that just providing students with the correct answer (a feedback approach currently used within ASSISTments). As shown in Figure 8, 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. Each of the first twelve problems in the problem set include 2 or 3 mistake message videos covering the common wrong answers (resulting in 23 videos created in total). These videos explained the process the student had used to arrive at their incorrect answer and how to start on 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 first 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 learning, we do not yet have proof that video is helpful in this context. However, this process has shown that a system like that depicted in Figure 6 would be viable for teachers to supplement problems with tailored mistake messages.

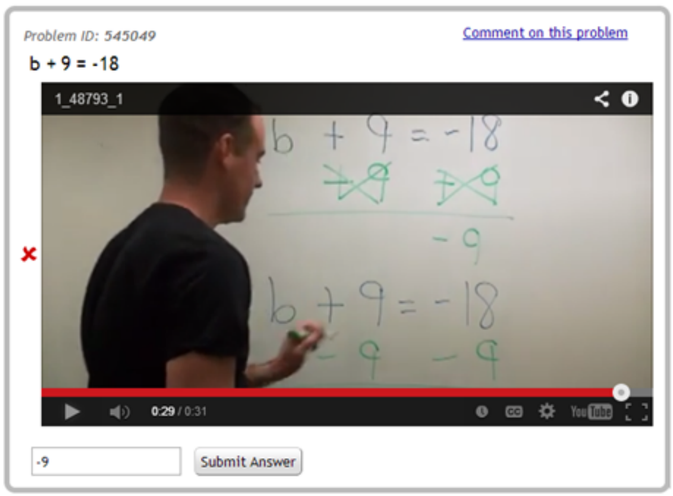


Figure 8. Teacher created video used as a mistake message tailored to the common wrong answer of "-9." In the short clip, the teacher notes that the student added 9 to both sides when they should have subtracted 9 from both sides.

Comparing Contributions from Students: Proof of Concept

One of the more unique studies currently running uses what we have termed the "If-Then-Else structure" and is paving the way for crowdsourced tutoring feedback from *students* within the ASSISTments platform. This study 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 explain their problem solving steps to their peers. As shown in Figure 9, the resulting explanations were rather different. The goal of this study is to observe differences in learning gains across multiple solutions for the same problem, thereby learning how to select the best crowdsourced content. Results are not yet available for this work. Eventually, this idea can be scaled to a larger number of users, sourcing video from students around the country to help peers within their classroom as well as within other schools.

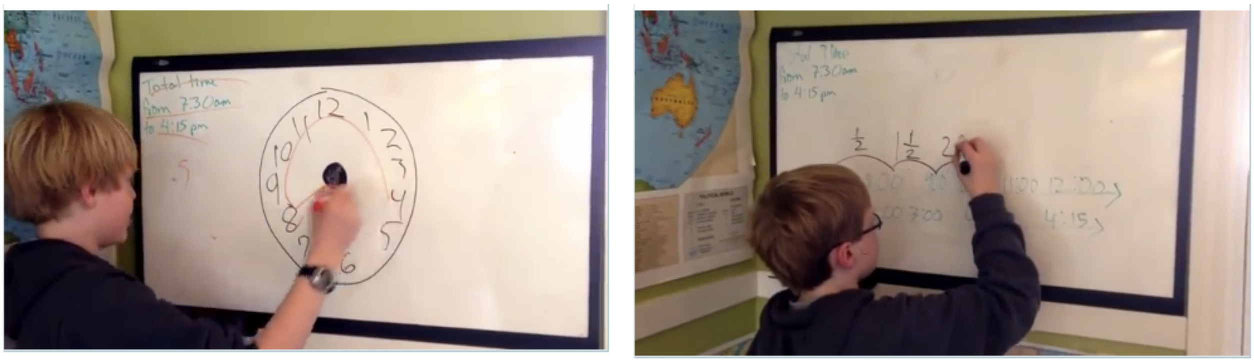


Figure 9. Student A solves an elapsed time problem using a method based on the way hands move around a clock. Student B solves the same problem using a method based on measuring steps between chunks of time that have passed.

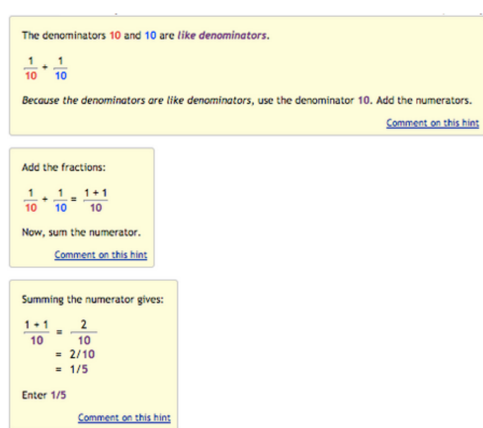
Collective vs. Individual Teachers' Contributions: "Patchwork Quilts" of Feedback

A variety of studies are currently running in the ASSISTments platform to investigate the effects of being taught by multiple teachers or sources. Versions of this "Patchwork Quilt" design have been built to house feedback videos from two or three teachers across a set of problems. For instance, in the design shown in Table 1 below, students are randomly assigned to a "Teacher" condition. Three teachers (A, B, & C) were asked to create video feedback for three isomorphic problems. Videos from Teacher A and Teacher B are shown in Figure 10 for comparison. Although both teachers approached video creation using pencasts, the formats are noticeably different. Students are randomly assigned to receive feedback from Teacher A, Teacher B, Teacher C, or from a mix of all three teachers. This randomized control trial design explores the effectiveness of crowdsourcing tutoring explanations at a small scale (N=3). Comparison of learning gains is accomplished using pre- and posttests implemented before and after the experience of video feedback.

Teacher A | Teacher B

Figure 10. 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.

Table 1. Across three isomorphic problems, students have the potential to receive video feedback from Teacher A, Teacher B, Teacher C, or a mix of all three teachers. The control condition features text feedback traditionally provided in the ASSISTments platform.

	Problem 1	Problem 2	Problem 3
Videos by Teacher A			
Videos by Teacher B			
Videos by Teacher C			
Videos by a Mix of Teachers			
Control (Text)	 The denominators 10 and 10 are *like denominators*. $\frac{1}{10}+\frac{1}{10}$ *Because the denominators are like denominators*, use the denominator 10. Add the numerators. Comment on this hint Add the fractions: $\frac{1}{10}+\frac{1}{10}=\frac{1+1}{10}$ Now, sum the numerator. Comment on this hint Summing the numerator gives: $\frac{1+1}{10}=\frac{2}{10}$ = 2/10 = 1/5 Enter 1/5 Comment on this hint		

Lessons Learned From Crowdsourcing Student Contributions

Beyond the experiments presented here, proof of concept work has been conducted to better understand the complexities and consequences of crowdsourcing feedback from students. Many of the potential concerns about allowing students to assist their peers can be explored as research questions:

1. How do we ensure the accuracy of student-generated material?
2. What is the efficacy of student generated material?
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. Can crowdsourcing be implemented as an effective use of teachers' time?
6. Is writing explanations of mistakes an effective use of student time?

It is probably fair to say that a range of possible outcomes exists for each of these concerns (i.e., students may be willing to generate feedback in some situations but not in others, therefore resulting in variability in the quality of feedback). But in order to answer these research questions properly, it is necessary to consider the efficacy and value of student sourced feedback. It is also necessary to fine-tune the methods employed to generate feedback from students while they work within ASSISTments.

One method that has proven successful has been requesting that students explain their work. The Common Core Standards for Mathematics (NGACBP & CCSSO, 2010) tell us that students should be able to explain their reasoning in addition to answering a question. Thus, more and more, students are providing written explanations of their work as part of normal instruction. The ASSISTments platform has the ability to gather these explanations and put them to good use. Explanations sourced from students can often translate directly into hint feedback or mistake messages for other students who are struggling with the same problem.

However, sourcing explanations from students carries very real concerns about the accuracy and efficacy of sourced content, as noted by research questions 1 & 2. In an initial attempt to examine the quality and effectiveness of student explanations provided as hints, a randomized controlled trial was conducted. 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 students created explanations for Topic A and provided answers for Topic B, while the other half of students created explanations for Topic B and provided answers for Topic A. Before the crossover, strong student explanations were selected by the teacher and made available to students as they attempted to provide answers for the respective topics. Feedback was presented as on demand hints for students who felt they were struggling. 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 if student created examples lead to better learning than traditional text hints. Due to instructor error, this posttest was not ultimately assigned and therefore results are not presented here. However, this rudimentary study served as a basis for a design that can be reused to assess the quality and usefulness of crowdsourced, student generated feedback.

Motivating Participation in Feedback Generation

As we crowdsource explanations from students to enrich the content in ASSISTments, it is necessary to ask why a student would want to help. Of course there are likely to be a few altruistic students who wish to go above and beyond, but the goal is to solicit feedback from all students. A potential approach to this goal is to entice students to explain their mistakes by providing an extra opportunity to earn credit within an assignment. While this approach would source more explanations and feedback messages, it could lower the quality of the feedback that is generated. 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* right in a row, with 3 being the default requirement. 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 who gets two consecutive correct answers has an 84.0% chance of correctly answering the third question (Van Inwegen, et al., 2015). A slight difference exists between the student that gets the first two questions in the assignment right (88.5% of getting the third) and the student who achieves the two consecutive correct answers at a later point during the assignment (82.6% of getting the third). With these probabilities in mind, a problem set was designed to allow students a second chance to answer a third 'consecutive' question, at the cost of providing a mistake message of feedback explaining their error.

ASSISTments does not yet have a permanent infrastructure to allow students a 'redo' by submitting an explanation or feedback. However, by using the new If-Then-Else navigator (Donnelly, 2015), it is possible to create assignments that follow this design. Provided the student has correctly answered two consecutive questions, the navigator allows for students to be routed into an ungraded open response question for the collection of mistake messages, before being returned to the most recent problem for a second chance. If the student answers their second chance correctly, it is also possible to provide a challenge question to ensure their knowledge before they are able to successfully complete the assignment. The goal behind a system providing a second chance on the same problem is simple: if the student is able to answer their '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 where they went wrong. On the other hand, students who answer the 'redo' incorrectly are not likely to provide useful feedback. The student was not able to pinpoint their mistake, and therefore the explanation they have generated is not likely to help other students. Performance on a second chance problem can therefore serve as an initial curator for weeding out feedback content that has low efficacy or accuracy. This process serves several purposes. Students may be more motivated to complete their assignment and to provide quality feedback when they can avoid having to restart their correct-in-a-row sequences. Further, students who are able to fix their mistakes may have nearly the same level of mastery as students who did not make a mistake in the first place. Plus, providing students an opportunity to learn from their mistakes has been shown to improve learning (Attali & Powers, 2010), and the process serves as a viable way to elicit feedback from students in the context of a typical assignment within ASSISTments.

Still, in order to implement this crowdsourcing strategy on a larger scale (i.e., from all students, across all content within ASSISTments), it is necessary 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 CROWDSOURCED STUDENT FEEDBACK: PEERASSIST

Since ASSISTments began, the standard method of instruction has included hints or scaffolding to help students solve a problem or to break the problem down into smaller steps. This approach will be overhauled by the implementation of PeerASSIST. The envisioned workflow of the new feature, as shown in Figure 11, 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 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 her feedback, it is linked to the problem she is working on and sent to the ASSISTments database. When another student in the same class begins the problem, an additional option will be added to 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 improper content within the peer-generated feedback. 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.

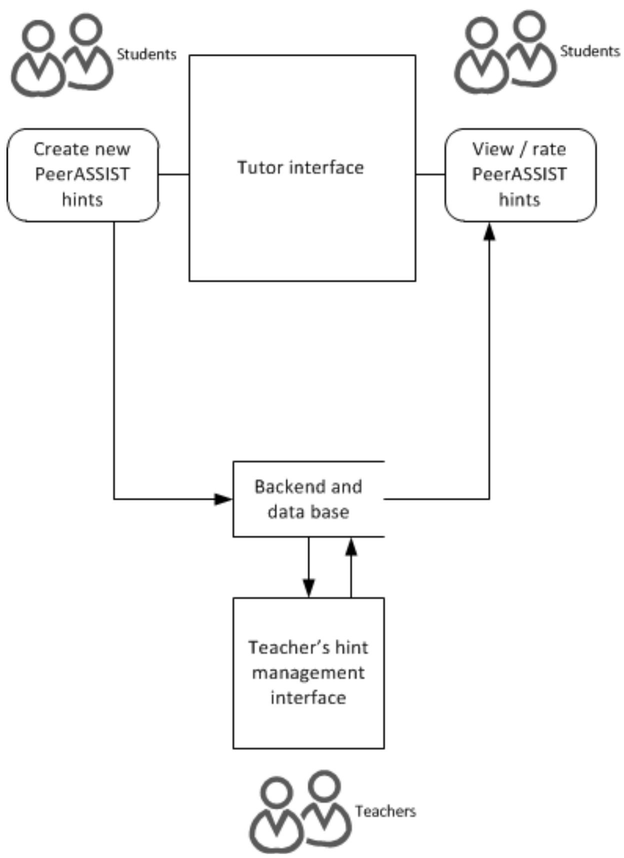


Figure 11. The PeerASSIST data flow. Students generate feedback for other students. Feedback is linked to a particular problem and provided randomly to students who struggle with the same problem. Students are able to judge the feedback provided by their peers, and teachers are able to manage feedback created by their students using a management interface.

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. This approach would be easy to implement and explain. However, if a PeerASSIST explanation has been "Disliked" many times, there is little reason to display it again. Further, the information linked to each PeerASSIST explanation has the potential go beyond "Likes" and "Dislikes." Within the Intelligent Tutoring System and Educational Data Mining communities, researchers would be more interested in learning specific outcomes for each instance 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 by the increase in the probability that the student gets the next problem correct, on their first attempt, without any help. Additionally, if crowdsourcing is implemented, how do we choose which content to assign to students?

This problem 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 (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 (often people) are saved. Another advantage to this approach is that if a particular condition in an experiment is detrimental, it can be avoided. This can often occur 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 a large type-I error of falsely rejecting the null hypothesis, although this can be prevented with 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 Figure 12. The player needs to determine which level ("arm") to pull that will provide the greatest payout rate in order to maximize his profits. The slot machines have earned the term 'bandits' because regardless of payout rate, they essentially steal money from the player (Lai & 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 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 crowdsourcing with 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.

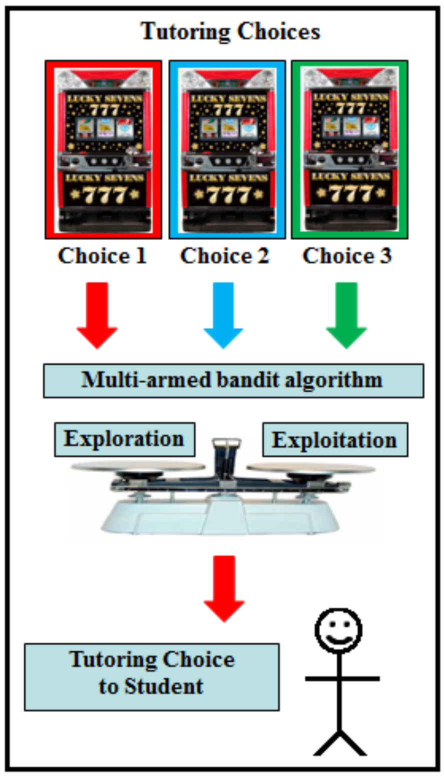


Figure 12. An example of how a multi-armed bandit algorithm can be used when crowdsourcing student explanations. In this example there are three slot machines representing three different student-generated tutoring strategies. A multi-armed bandit algorithm is run balancing exploration and exploitation to determine which of the three tutoring strategies is given to the next student.

CLOSING THOUGHTS

We feel 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. 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, struggling students. This practice would benefit all parties: explaining a solution allows the student to solidify 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 adaptive 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.

We feel that harnessing the knowledge of the crowd will enhance adaptive learning platforms moving forward. The next 25 years within educational technology should be marked by risks that seek to bring 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 within related fields.

ACKNOWLEDGEMENTS

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, and 1440753), 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.

REFERENCES

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.

Attali, Y. & Powers, D. (2010). Immediate feedback and opportunity to revise answers to open-ended questions. *Educational and Psychological Measurement*, *70*(1), 22-35.

Donnelly, C. (2015). Enhancing Personalization Within ASSISTments. Master's Thesis, Computer Science Department. Retrieved on May 11, 2015 from http://www.wpi.edu/Pubs/ETD/Available/etd-042315-135723/

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.

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. Int J Art Intel in Ed.

Howe, J. (2006). The rise of crowdsourcing. *Wired magazine*, *14*(6), 1-4.

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 Artificial Intelligence in Education (pp. 824-827). Springer Berlin Heidelberg.

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., ... & Klemmer, S. R. (2015). Peer and self assessment in massive online classes. In*Design Thinking Research* (pp. 131-168). Springer International Publishing.

Lai, T. L., & Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, *6*(1), 4-22.

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. Washington, DC: Authors.

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

Robbins, H. (1952). Some aspects of the sequential design of experiments. In *Herbert Robbins Selected Papers*. pp. 169-177. Springer New York.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and instruction*, *16*(4), 475-5223.

Selent, D., Heffernan, N. (2015).When More Intelligent Tutoring in the Form of Buggy Messages Does Not Help. *Proceedings of the 17th International Conference on Artificial Intelligence in Education*. 2015.

Van Inwegen, E., Adjei, S., Wang, Y., & Heffernan, N. (2015). Using partial credit and response history to model user knowledge. *Proceedings of the Eighth International Conference on Educational Data Mining*. 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, Vancouver, BC, Canada.

Weld, D., Adar, E. & Chilton, L. (2012). Personalized online education—a crowdsourcing challenge. *HComp'12 Proceedings of the AAAI Workshop on Human Computation*, pp.159–163.

Williams, J. J., Krause, M., Paritosh, P., Whitehill, J., Reich, J., Kim, J., Mitros, P., Heffernan, N., & Keegan, B. C. (2015a). Connecting Collaborative & Crowd Work with Online Education. *Workshop conducted at the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing.* [Extended Abstract] [Website: tiny.cc/crowdworklearning]

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., Ostrow, K., Xiong, X., Glassman, E., Kim, J., Maldonado, S. G., Reich, J., & Heffernan, N. (2015b). Using and Designing Platforms for In Vivo Educational Experiments. *Proceedings of the Second ACM Conference on Learning@Scale*.

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.