The objective of this project was to address design gaps in previous programs that address scientific argumentation in middle schools. We adopted gamification mechanics (e.g., stimulus and response and progression) to design a computer-assisted program to better support students' development of scientific argumentation. In this paper, we describe the development of a gamified tool to support the representation of collaborative scientific argumentation for middle school students and describe how a faculty member in the Instructional Technology program and a faculty member and students in the Department of Computer Science collaborated and delivered project work through the Agile methodology. We developed two iterations of the prototype and conducted usability tests of the two iterations of the design with adults and middle school students in the United States. This study presents how the design evolved from the first to the second iteration based on usability testing results. The adult participants (e.g., science pre-service teachers and an in-service teacher) and student participants provided suggestions for further improvement of the second iteration of the prototype.

PI-SUI HSU is an Associate Professor in the Department of Educational Technology, Research and Assessment (ETRA) at Northern Illinois University. Her research interests focus on technology integration in science education, particularly in scientific argumentation.

REVA FREEDMAN is an Associate Professor in the Department of Computer Science at Northern Illinois University. She specializes in the development of an intelligent tutoring system.

DARIN BROCKMANN, ZACHARY HUENEKE, DEAN LABARBERA, BEN KLOGA, RUI ZHANG, and IAN SULLIVAN are undergraduate students in the Department of Computer Science. They specialize in the development of intelligent tutoring systems.

MARGOT VAN DYKE is a middle school science teacher. Her research interests focus on instructional design and integration of technology in science teaching and learning.

INTRODUCTION

This paper presents a design case of a gamified scientific argumentation tool for middle school students. We are an interdisciplinary team that involves faculty and students from the Instructional Technology program and the Computer Science program. We share how we delivered the project through Agile methodology and present challenges in the design process. We also describe the design process of the gamified tool in both the first and second iterations of the design process. We share the users’ experience in both iterations of the prototype and indicate how we made revisions based on the test results. The design detailed in this case is one that was conducted as part of a research study. The current paper describes the specific design in this design case.

CONTEXT

The objective of this project was to develop a gamified scientific argumentation tool for middle school students in the United States. The Next Generation Science Standards (National Research Council, 2012) identified scientific argumentation as one of the eight essential science practices for students. Scientific argumentation is a form of logical discourse that involves arriving at an agreed-upon position among members of a group (Kuhn, 1993) and is practiced when scientists build on and refute one another’s theories and empirical evidence to arrive at scientific conclusions. As adolescence is a critical age in which argumentation skills develop (Belland et al., 2011), researchers (Scheuer et al.,
Suthers et al., 2008) have developed computer-assisted tools (e.g., Digalo, Belvedere, Araucaria) to support such endeavors. Computer-assisted tools have shown great potential as cognitive tools for helping learners demonstrate their thinking visually, share the cognitive load (Jonassen & Carr, 2020) and support the development of middle school students’ scientific argumentation skills.

In previous studies, although each of the computer-assisted tools (e.g., Digalo, Belvedere, Araucaria) has a different way of constructing argumentation maps, there are a number of common features across these tools. For example, contributions are displayed as boxes or nodes that represent argument components, such as claims. The arrows represent the relation between the argument components (e.g., supports or refutes). The different components of arguments and relations can be easily distinguished via their visual appearance.

Extending this line of research, since 2016 the first author has led the “Argue like a Scientist with Technology” (ALAST) team to conduct studies to examine ways to enhance students’ development of scientific argumentation skills. This project included adapting a proprietary concept mapping tool as the basis for a graph-oriented tool (Figure 1) in a computer-assisted project-based argumentation curriculum (Hsu et al., 2016; Hsu et al., 2018). Hsu et al. (2016) investigated how middle school students were engaged in collaborative argumentation with the support of an online graph-oriented program and how this intervention led to their development of science argumentation skills and science knowledge. Although the impact of an online graph-oriented program on the development of scientific argumentation was positive, we observed that matching a shape to a corresponding scientific argumentation component and dragging and dropping a shape into the proprietary tool imposed an extra cognitive load on the students. In the proprietary tool, the students had to follow the instructions in a handout about what shape they would use to represent a

---

**FIGURE 1.** Screenshot of the proprietary tool.

---

<table>
<thead>
<tr>
<th>Shapes and Arrows</th>
<th>Argumentation Skill</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position (light bulb)</td>
<td>An opinion or conclusion on the main question</td>
<td></td>
</tr>
<tr>
<td>Reason (rectangle and arrow)</td>
<td>A claim that supports the position</td>
<td></td>
</tr>
<tr>
<td>Evidence (cloud and arrow)</td>
<td>A separate idea or example that supports a reason (or counterclaim or rebuttal)</td>
<td></td>
</tr>
<tr>
<td>Counterargument (signified by star and “x”)</td>
<td>A claim that refutes another position or gives an opposing reason</td>
<td></td>
</tr>
<tr>
<td>Rebuttal (signified by oval and “xx”)</td>
<td>A claim that refutes a counterargument by demonstrating that it is invalid, lacks as much force or correctness as the original argument or rests on a false assumption.</td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 2.** Argument elements by corresponding shapes and arrows in a proprietary tool.
certain scientific argumentation component (Figure 2). Figure 3 indicates a sample of scientific argumentation in the proprietary tool.

In terms of design gaps, as shown in previous computer-assisted tools (e.g., Digalo, Belvedere, Araucaria, proprietary concept mapping tool), students need to drag and drop different objects or copy lines from the object palette to the drawing canvas to represent their argumentation. While addressing the issue of cognitive load and design gaps, we adopted gamification to design a computer-assisted program to better support students' development of scientific argumentation. Gamification is the application of game design elements and game principles in non-game contexts (Robson et al., 2015). In this study, we aimed to contribute to design knowledge on how to employ game design elements such as stimulus and response, progression, and collaboration to scaffold students' development of scientific argumentation (Arnab et al., 2015). This study also contributed insights into the challenge of design work during the COVID-19 pandemic (Mangalaraj et al., 2014). The challenges included turnover of undergraduate students, identifying a solution to keep design knowledge, and explaining new concepts in online meetings, which will be elaborated in the next section.

**AGILE METHODOLOGY AND CHALLENGES IN DESIGN TEAMWORK**

We formed the team in January 2020. The original design was conceptualized by the first author who was a faculty member in the Instructional Technology program at a state university located in Northern Illinois in the United States. The first author collaborated with the second author in the Computer Science program. They recruited three undergraduate students in the Computer Science program who showed interest in this project. Later in Spring 2020, one more undergraduate student in the Computer Science program joined the team. In Fall 2020, two more students in the Computer Science program joined the design team. We chose an interdisciplinary team because we needed strengths in both educational technology and computer science. The latter was needed because we needed to build the gamification features from scratch. However, this process required a longer development timeline than modifying an existing tool.

We delivered the project through Agile methodology (Crowder & Friess, 2014; Cullen, 2021). Agile methodology is a project management process, mainly for software development, in which demands and solutions evolve through the collaborative effort of cross-functional teams and their customers. For example, when the first author discussed the initial conceptualization of the design and the goal of the program, the student programmers were encouraged to propose solutions. Gamification was one of the solutions they proposed. We discussed a number of game elements such as stimulus and response, progression, and collaboration to scaffold students' development of scientific argumentation. We broke the task down into small and manageable milestones in the process. One student developed the back-end database. Two students developed the front-end interface. Three students worked on integration of the database and the front-end interface. The entire team met two hours per week to ensure that the student team delivered a working program based on the content expert's (first author) expectations and to reflect collectively on the development process. During the emergency shut-down due to the COVID-19 pandemic, we made the quick transition from inperson meetings to online meetings and have met every
week since. As such, we were able to test the first iteration of the program in Summer 2020 and receive early feedback to revise the program and test the second iteration in Summer 2021. We conducted alpha testing and beta testing for each iteration of the prototype.

Usability testing includes a range of tests and evaluation methods, and one of the main purposes of usability testing is to identify issues that keep users from meeting the goals of a software program. We used a number of tests and evaluation methods to identify issues in the prototype. There were two phases of usability testing: alpha testing and beta testing. The usability testing in both phases focused on how the prototype improved the process of representation of scientific argumentation and how the prototype could be further improved. The testing process and results are described in the following sections.

There are a number of challenges in design teamwork. The turnover of undergraduate students is common because they eventually graduate and new students join the team. Although some of the students stayed on the team, others moved on with their careers. Thus, identifying a solution to keep design knowledge was a challenge. We decided to use GitHub as a repository and asked each team member to provide detailed documentation of their code before they left. With that data, another student could pick up the existing design knowledge and continue to develop the program. Another challenge was that the first author discussed the project using ideas and terms from education that the Computer Science students were not familiar with. The second author, who had previously worked in an educational institute, helped to translate these terms and concepts. This process created a synergy among team members; thus, we always allocated some of the meeting time to explaining terms and concepts in both fields. The third challenge was that online meetings limited how we shared ideas. In inperson meetings, we frequently sketched on a whiteboard to express how the program should be structured and demonstrate the ongoing design. In online meetings, it was difficult to sketch; instead, we used more text to explain our design and lay out the plan. Additionally, undergraduate students needed time to pick up the knowledge and skills required, and they needed to fit the project in with their coursework and other obligations. An essential part of the Agile methodology is allowing us to set our own deadlines. Thus, although the program has developed as planned to support the development of scientific argumentation, it has taken almost two years to develop a high-quality product that met users’ needs.

**DESIGN PROCESS IN THE FIRST ITERATION**

With the goal of developing a gamified program to support middle school students’ development of scientific argumentation skills, we selected the Unity game engine as the platform to develop it. In addition to Unity, we used several applications such as JavaScript, PHP, and the C# programming language to develop the front end and back end of the program. We also developed an API (Application Programming Interface) to allow integration of different components. Meanwhile, we also developed the landing page, which allows access to the login screen, the teacher registration page, the student registration page, and an experiment definition page.

A number of researchers (McNeill et al., 2006; Sampson et al., 2012) have defined essential elements of argumentation: claim, evidence, reasoning, counterargument (alternate claim), and rebuttal (challenge to the alternate claim). A **claim** is an answer to a question; **evidence** includes relevant measurements or observations that support a claim; and **reasoning**, which often incorporates scientific principles and ideas, functions as the linkage between the claim and evidence. **Counterargument** refers to an assertion that counters another claim or gives an opposing claim. **Rebuttal** is an assertion that refutes a counterargument by demonstrating that the counterargument is not valid, lacks as much force or correctness as the original argument, or is based on a false assumption (Hsu et al., in press).

The first development tasks included review of scientific argumentation element objects (called containers), icons, and relationship signifiers (arrows). Based on the agreed-upon definition for the argumentation elements (McNeill et al., 2006) and previous studies (Hsu et al., 2016), we developed gender and culturally neutral graphic icons to represent the scientific argumentation elements. The claim is represented by a flag. The evidence is represented by a magnifying glass. Reasoning is represented by a light bulb. A counterargument is represented by crossed swords, and rebuttal is represented as a shield.

This paragraph presents what the actual game-like aspect of the design looks like as it is running. We evaluated the workspace layout and adopted a hierarchical design for the scientific argumentation process. The hierarchical structure focuses on a parent-child relationship and presentation for the various scientific argumentation components. Figure 4 shows a sample scientific argumentation process from the first iteration of the prototype. When the prototype was tested in the classroom, two sets of subjects were assigned to work on the same page using icons of different colors. Figure 4 shows that a blue team and an orange team were engaged in collaborative scientific argumentation. To scaffold students’ development of scientific argumentation, two types of gamification mechanics, stimulus, and response as well as progression, are built into the program. Stimulus and response were manifested in the boxes the represent a claim, evidence, reasoning, a counterargument, and a rebuttal. For example, the Claim box for each team appeared on the screen when students logged into the program. They
typed in their claim in the white space. The students on the orange team had to click on the small Evidence icon in the upper-left corner of the Claim box. Then the Evidence box appeared, and they progressed to adding evidence. The same procedure applied to the progression to reasoning. With progression, they had to start with a claim, followed by a piece of evidence and then by reasoning. They cannot skip evidence to reach reasoning. The blue team could click on the small Counterargument icon in the upper-right corner of the Claim box to make a Counterargument box appear. The blue team could enter their counterargument and provide their reasoning and evidence. Then the orange team could click on the small Rebuttal icon in the upper-right corner of the Counterargument box. The Rebuttal box would appear. With progression, the blue team was able to provide a counterargument based on the claim that the orange team developed. The orange team was able to provide a rebuttal based on the counterargument that the blue team provided. The first iteration of the program was completed at the end of Spring 2020. The following section presents users' experience with the first iteration of the prototype in Summer 2020.

**USERS’ EXPERIENCE OF THE FIRST ITERATION OF THE PROTOTYPE**

It was challenging to recruit participants during the pandemic. The first author reached out to several doctoral students and preservice science teachers whom she knew and reached out to several science teachers with whom she had collaborated. Two graduate students, two undergraduate preservice teachers and one science teacher agreed to participate in the alpha testing. The science teacher taught a summer camp that involved ten students. She was able to recruit five students to participate in the beta testing. In the usability testing, Nielsen (2012) suggested that testing with five people lets us find almost as many usability problems as we would find using many more participants.

**Alpha Testing**

The purpose of alpha testing was to examine whether the five adult participants were able to navigate the interface and identify the essential features needed to support the representation of scientific argumentation. Due to the stay-at-home order, we set up a Microsoft Teams Meeting with each participant and conducted online observations and interviews for up to two hours. Each research meeting was recorded and saved for later data analysis.

We conducted performance measurement with each adult participant. A performance measure is a technique used to obtain quantitative data about test participants’ performance of tasks during usability tests (Soken et al., 1993). Soken et al. suggest that this method provides quantitative data useful for doing comparative testing as well as testing against predefined benchmarks. The technique can be used in combination with post-test interviewing and questionnaires so that both quantitative and qualitative data are obtained.

We followed the guideline of including five adult participants and followed up with an interview with each adult participant. We began the performance measurement by providing orientation to the concept of scientific argumentation and then providing a list of nine tasks for each adult to test with the prototype (see Figure 4). We observed their performance and tracked whether they did each task correctly. We assigned one point to each task performed correctly. The maximum score the participants could receive was nine and the minimum score was zero. The tasks included “Type in a claim in the Claim box.” and “Click on the Evidence icon of the Claim box. Add your evidence.” We provided each adult participant a description of each task and asked them to perform each task one by one. Additionally, we also asked them to think aloud and provide comments.

The data indicated that all adult participants completed all tasks correctly except the first task and received eight points for the performance measure. All participants had trouble re-centering the screen on the claim box. We provided a hint to use the arrow keys, so they were able to complete this task. The participants suggested providing multiple ways to navigate the program due to a variety of devices (e.g., iPad, Chromebook) used in schools as well as accessibility issues. For example, the program should allow using arrow keys on
the keyboard or a touchpad to pin in or stretch out when a mouse is not available in the classroom.

After the participants completed the performance measure, we followed up with an interview, including these sample questions:

- Did you enjoy the program?
- What was your least favorite part about the program?
- Did this program help you understand scientific argumentation?

Overall, all five adult participants enjoyed using the program and pointed out that it was easy to use. Carrie (pseudonym) commented, “The program was interactive and allows for integration in formal and informal science learning contexts.” Mary (pseudonym) said, “It is very kid-friendly, particularly the shapes and the colors. It is easy to read, organize, and intuitive for students.”

Becky (pseudonym) commented on the potential use of this program for explaining scientific argumentation to others:

The program provides a good representation of scientific argumentation and does a good job of connecting different icons with different components of scientific argumentation. The program can allow teachers to model the scientific argumentation process and allow teachers to show the process in a visual way.

Carrie also commented:

This program allows students to do it in real-time or outside the classroom. It also allows for peer interaction. It would also allow shy students to engage in the scientific argumentation process.

Mary added that she liked the visual component of the program and the ability to follow specific patterns. She said, “Branching out from one component to another component really helps explanation.”

However, they also commented on their least favorite parts of the program. Becky, Cassie, and Mary had trouble zooming out the screen to see everything and had to use arrow keys. Peter (pseudonym) and Bob (pseudonym) pointed out that they had to scroll left and right constantly to see the whole scientific argumentation process.

Additionally, we asked each participant to select the design of a container from three designs based on aesthetics (e.g., color scheme) and placement of icon for each scientific argumentation component (Figures 5, 6 and 7).

Becky chose design three as her first choice because of its modern and asymmetrical design. Carrie, Peter, Bob, and Mary chose design one as their first choice and design three as their second choice. Carrie explained:

Although design one is a bit old school, the box has an icon on the top, coupled with the term. Having icons on two separate corners ensure that students won’t mis-click, particularly for middle school students.
Mary added that design three has a clean and modern look. Bob also pointed out that design three and its color scheme are student oriented. It seems that all participants agreed that design one and design three were acceptable. Design two was the least favorite for all participants. Bob commented, “The color is plain.” Carrie added, “The layout is just like writing a paper.” Based on their feedback, we also agreed that the colors on the containers should have good contrast and intensity and should consider people with color-blindness.

We also asked all participants about how multiple containers for the same scientific argumentation components (e.g., evidence) should be represented. We had brainstormed several ideas. One was to display all containers on the same screen and to require users to pan around the screen to see all of them. Another was to arrange all containers as stacked-up cards and hide the inactive ones. The discussion was inconclusive because it was challenging to visualize the display. We decided to seek additional input from student participants.

**Beta Testing**

The purpose of beta testing was to examine whether student participants were able to navigate the interface and identify the essential features in order to support representation of scientific argumentation. Due to the stay-at-home order in Illinois, we conducted the study in an online setting. We collaborated with a middle school science teacher who sent a Google Meet link to five volunteer students and us. We conducted online observations and interviews with a focus group of five student participants for 90 minutes. Each meeting was recorded and saved for later data analysis. One student remembered the time incorrectly and was not able to participate in the study.

We conducted the performance measure with a focus group given that middle school students tend to share their thoughts when their peers are present (Markopoulos & Bekker, 2002). Therefore, we did not choose to conduct the study with individual students.

We followed the same procedure as the adult participants and began conducting performance measures by providing orientation to the concept of scientific argumentation. Then we provided a list of nine tasks for each student to test with the prototype (see Figure 4). We observed the students’ performance, tracked whether they did each task correctly, and assigned one point to each task performed correctly. The maximum score students could receive was nine and the minimum score was zero. The tasks included “Type in a claim in the Claim box.” and “Click on the Evidence icon of the Claim box. Add your evidence.” We provided each adult participant a description of each task and asked them to perform each task one by one. Additionally, we also asked each of them to think aloud and provide comments.

The data indicated that all student participants completed all tasks correctly and received nine points on the performance measure, which outperformed the adult participants. All participants were able to recenter the screen on the claim box without any hints. Evan, Elena, and Sophie (pseudonyms) used arrow keys and WASD keys a lot when they played games such as Minecraft. The WASD system allows participants to use those letters of the alphabet to move the cursor. Thus using arrow keys to re-center, the screen was not new to them. Sonia did not play games but used arrow keys to navigate other programs. We found that they used four common ways to navigate various programs, including keyboard, touchpad, touchscreen, and mouse. They preferred the keyboard and mouse and suggested having at least two methods available for navigating the program.

After completing the performance measure, we followed up with an interview, including these sample questions:

- Did you enjoy the program?
- What was your least favorite part about the program?
- Did this program help you understand scientific argumentation?

Overall, all student participants enjoyed the program and pointed out that the program was easy to use. Elena (pseudonym) commented, “The program was very easy to use.” Evan (pseudonym) said, “It is organized and helps with visual learners.” All agreed that they felt confident using this program to explain scientific argumentation to their friends because it allowed them to progressively click on different icons to show their friends how to build scientific argumentation. Additionally, Sonia pointed out that her least favorite aspect of the program was the loading time. It took a while for her to load the program and scroll back and forth from right to left and left to right. When we probed for plausible reasons, this was attributed to the device (e.g., iPad) or the bandwidth at home.

We asked all students to select the design of a container from three designs (see Figure 5, 6 and 7) based on aesthetics and placement icons for each scientific argumentation component. All students ranked design three as their favorite design, then design one and lastly design two. Evan commented, “Design three looks more open. Colors are appealing.” Elena said, “Design three has softer color. It is kid-friendly and has a bubbly look.” Sophia added, “The colors seem calmer. The design is stylish.” Evan commented on design one: “I don’t like the font in design one.” Sonia pointed out that “Design two has a serious look.” Although they preferred design three over the other two designs, they pointed out that the green and blue in design three did not have good contrast. They also suggested giving students choices for selecting their own colors for their container.
DESIGN PROCESS IN THE SECOND ITERATION

Based on the results of alpha testing and beta testing in Summer 2020, we identified a number of areas that required revision of the design. As shared by both the adult and the student participants, it was challenging to scroll back and forth from right to left and left to right because the space on the screen was limited (Figure 4), particularly when there were multiple containers for components of scientific argumentation (e.g., counterarguments). With that, we made revisions in terms of the spatial arrangement. Figure 8 shows that the blue team is at the top and the green team is at the bottom in the revised version.

To ensure that the interface was easy to navigate, we examined three aspects: how tree structures grow, how the different containers stack up, and a reasonable maximum number of each type of container. First, in terms of tree structures, we revised the left-right design (Figure 4) to a top-down (Figure 8) design. For example, if the blue team on the top clicked on the small Evidence icon in the lower-left corner of the Claim icon, the Evidence box would appear, allowing the blue team to enter their evidence. If they then clicked on the small Reasoning icon in the lower-left corner of the Evidence box, the Reasoning box would appear, which allowed the blue team to enter their reasoning. As the tree grew, the blue team was still able to see the scientific argumentation constructed by the green team.

In terms of the representation of multiple containers for counterarguments, we re-designed the counterargument container with a stack view (Figure 9). We also added a number indicator in the upper left corner and a switch button in the upper right corner. For example, in Figure 8, the green team could click on the small Counterargument icon in the lower-right corner of the blue team’s Claim box. Figure 9 indicates that the green team clicked on the Counterargument icon of the blue team’s Claim box three times, giving three counterarguments. They could click on the switch button to see the first, second or third counterargument in sequence.

To simplify the programming, we also studied the maximum number of containers needed for each scientific argumentation skill. The content expert reviewed the empirical data from prior research (Hsu et al., 2016; Hsu et al., 2018) and studied the maximum number of containers for each scientific argumentation skill that each student group developed. For instance, we determined that there were no more than ten pieces of evidence for each claim. There were no more than five counterarguments to each claim and no more than two rebuttals to each counterargument. Based on these numbers, we adopted a set form instead of a dynamic form for program expansion of each container, which means that space for each container could be preallocated on the screen.

We also re-examined the color scheme of each container and selected dark text in our new design (Figure 10). We also built in a function that allowed students to select a color for their container when they start the program (Figure 11). We completed the second iteration of the prototype in Spring 2021. The following section presents the users’ experiences with the second iteration of the prototype in Summer 2021.
Four adult participants who participated in the usability testing of the first iteration of the prototype had graduated. Only the science teacher continued to participate in alpha testing of the second iteration of the prototype. It was challenging to recruit participants during the pandemic. The first author was able to recruit one doctoral student to participate in the alpha testing. The science teacher was able to recruit ten students from her class to participate in beta testing. For usability testing, Nielsen (2012) suggested that while five people would be sufficient for usability testing, there could be as few as two participants for low budget projects.

Alpha Testing

The purpose of the alpha test was to examine whether the two adult participants could navigate the interface and identify the essential features that support representation of scientific argumentation. One participant is an instructional designer, and the other participant is a middle school science teacher. Due to the pandemic, we set up a Microsoft Teams Meeting with each individual participant and conducted online observations and interviews for one hour and a half. Each meeting was recorded and saved for later data analysis.

We began the performance measurement by providing an orientation to the concept of scientific argumentation and then a list of ten tasks for each adult to perform with the prototype (see Figure 8). We observed the performance and tracked whether each person did each task correctly and assigned one point to each task performed correctly. Each adult participant was prompted to choose a color (Figure 11) as he/she logged in to the program. The maximum score was ten and the minimum score was zero. The tasks included “Navigate down to the other team’s claim. Give a counterargument to it.” and “Switch back to your first counterargument.” We provided each adult participant a description of each task and asked them to perform each task one by one. Additionally, we also asked each of them to think aloud and provide comments.

The data indicated that the two adult participants completed all tasks except the ninth task correctly and received nine points for the performance measure. All participants had trouble switching back and forth among multiple counterarguments (Figure 9). We provided a hint about clicking directly on the arrows, so they were able to complete this task. The participants suggested having the whole button clickable instead of the arrows.

After they completed the performance measure, we followed up with an interview using these sample questions:
performed correctly. Each student participant was prompted to complete each task correctly, and assigned one point to each task. We observed their performance, tracked whether they did each task correctly, and assigned one point to each task performed correctly. Each student participant was prompted to choose a color (Figure 11) as he/she logged into the program. The maximum score was ten and the minimum score was zero. The tasks included “Navigate down to the other team’s claim. Give a counterargument to it.” and “Switch back to your first counterargument.” We provided each student participant a description of each task and asked them to perform each task one by one. Additionally, we also asked each of them to think aloud and provide comments.

The data indicated that all student participants completed all tasks correctly and received ten points for their performance measure except Marybeth (pseudonym). Marybeth, like both adult participants, had trouble switching back and forth among multiple counterarguments. After the students completed the performance measure, we followed up with an interview. Here are sample questions:

- Did you enjoy the program?
- What part of the program do you think needs improvement?
- What was the main method of navigation you used?
- After using this program, would you feel confident explaining how a scientist builds an argument?

Overall, all student participants enjoyed using the program and pointed out that it was easy to use. Mina (pseudonym) commented, “The program is a cool tool.” Eva (pseudonym) said, “It is easy to use.” Arthur (pseudonym) said, “It is very straightforward to navigate.” All students used either the touchpad or the touchscreen on their Chromebooks to navigate the program.

All students supported the use of the program to help them learn scientific argumentation. Sarah (pseudonym) said, “It helps explain the basics. I would use it as a template for practice.” Eva said, “It is easy to get a hang of it. I can use the program to explain scientific argumentation step-by-step.” Mina and Arthur commented that this tool can help students who struggled with scientific argumentation. Mimi (pseudonyms) shared similar comments. Mimi (pseudonyms) added, “It allows me to explain scientific argumentation to my friends. It is easier to add ideas. It helps with visualization and brainstorming with others.” Lisa and Mina also shared similar comments. Lisa said, “It shows you different components. It is user-friendly.” Arthur claimed the program has real-life applications. Alex said that he would use the program to provide examples.

However, the student participants also pointed out a few areas for improvement. Eva, Mina, Sarah, and Arthur suggested making the containers bigger because when they typed more text, the text in the containers shrank. All students also suggested providing instructions before they played the game. Sarah said, “You can provide instructions and directions on the main page.” Alex suggested adding a tutorial. All students suggested making the background

Beta Testing

The purpose of the beta test was to examine whether the student participants were able to navigate the interface and identify the essential features to support representation of scientific argumentation. We conducted the study in an online setting. We collaborated with a middle school science teacher who sent a Zoom link to ten volunteer students and us. We conducted online observations and interviews with each student for 20 minutes. Each meeting was recorded and saved for later data analysis. In this iteration we interviewed each student separately to make sure that information was collected from the quieter students.

We followed the same procedure as with the adult participants and began conducting performance measures with each student by providing orientation to the concept of scientific argumentation. Then we provided a list of ten tasks for each student to test with the prototype (see Figure 8). We observed their performance, tracked whether they did each task correctly, and assigned one point to each task performed correctly. Each student participant was prompted
more interesting. All students played online games (e.g., Minecraft, Roblox). With their experience, they suggested a number of gamification mechanics. Armor (pseudonym) and Mina suggested adding a challenge to the program by earning points. Sarah suggested allowing students to create their own characters.

**SUMMARY**

The objective of this project was to develop a gamified collaborative scientific argumentation program for middle school students. Overall, the results of usability testing suggested that the two iterations of the prototype effectively targeted representation of scientific argumentation. Additionally, the features in the prototype helped the students develop their scientific argumentation skills. We plan to fine-tune and add the features suggested by both the adult and student participants. In addition to game design elements such as stimulus and response, progression, and collaboration for scaffolding students’ development of scientific argumentation, we plan to add more gamification mechanics (e.g., a leaderboard) and conduct a study with students in another country. Currently we are working on integrating different parts of the system, including the front-end design shown in this study and the back-end database.

We learned a number of lessons in the design process which may offer insights for the design community. First, Agile methodology was effective for an interdisciplinary team and worth the time it takes. Weekly meetings allowed us to raise our questions, share design ideas and review ongoing design. Second, turnover of students in a team is inevitable. With that, it was critical to build a design knowledge base that could be passed along to other students who joined the team later. Third, conducting online meetings for design work was a challenge because it was difficult for us to draw or sketch. We had to scan our design documents and share those documents in the online meeting platform. With the advent of technology, the online meeting platform we used plans to incorporate an interactive Whiteboard and this issue could be resolved.

**ACKNOWLEDGEMENTS**

The authors gratefully acknowledge Gail Jacky for editing the manuscript, without which this work would not have been possible.

**REFERENCES**


