2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #34592

Evaluating an Intelligent Sketching Feedback Tool for Scalable Spatial Visualization Skill Training

Ms. Tiffany Wenting Li, University of Illinois at Urbana-Champaign

Tiffany Li is a Ph.D. student in computer science at the University of Illinois at Urbana-Champaign. Her research interest lies in human-computer interaction and educational technology.

Mr. Ziang Xiao, University of Illinois at Urbana-Champaign

Ziang Xiao is a graduate student from the computer science department at the University of Illinois at Urbana-Champaign. His primary research interest is in human-computer interaction.

Dr. Molly H. Goldstein, University of Illinois at Urbana - Champaign

Molly H. Goldstein is Teaching Assistant Professor in Industrial and Enterprise Systems Engineering at University of Illinois, Urbana-Champaign. She previously worked as an environmental engineer specializing in air quality influencing her focus in engineering design with environmental concerns. Her research interests center on engineering design in undergraduate and precollege settings. She obtained her BS in General Engineering (Systems and Design) and MS in Systems and Entrepreneurial Engineering from the University of Illinois and PhD in Engineering Education from Purdue University.

Prof. Michael L. Philpott PhD., University of Illinois at Urbana - Champaign

Mike Philpott is an Emeritus Associate Professor of Mechanical Science and Engineering at the University of Illinois at Urbana-Champaign (UIUC). He is also Founder and Chief Scientist of aPriori inc., an Enterprise Cost Management (ECM) software company offering the first CAD-integrated solution for real-time, predictive and precise cost estimating throughout the entire product development and product delivery process. The patented enabling technology for this was developed over 10 years of R&D at UIUC, funded primarily by John Deere. Mike has over 50 publications in this area, primarily in Design-to-Cost, Design-for-Manufacture, and Design-for-Assembly. He directed the research in process cost modeling and the development of the aPriori technology, and his graduate student team won the Harvard Business Plan competition in 2003. In 2004, the company achieved $4.5M venture capital 1st round funding from Bain Capital and Sigma Partners, and has since grown to 150+ employees. Mike received his MS in 1979 and PhD in 1986 from Cranfield Institute of Technology, England, and worked for a number of years as a design engineer in the British automotive and machine tool industry, specifically Land Rover, Jaguar, and Molins Machine Tool.

Dr. Brian Woodard, University of Illinois at Urbana - Champaign

Dr. Woodard received his Ph.D. in Aerospace Engineering from the University of Illinois at Urbana-Champaign in 2011. His Aerospace research interests currently focus on the effects of icing on the aerodynamics of swept-wing aircraft. In engineering education, he is also interested in project-based learning and spatial visualization. He teaches courses at the University of Illinois where he serves as the Director of Undergraduate Programs for the Department of Aerospace Engineering.

Evaluating an Intelligent Sketching Feedback Tool for Scalable Spatial Visualization Skill Training

Abstract

Spatial visualization skills are essential and fundamental to studying STEM subjects. The increasing need for STEM education poses scalability challenges in spatial visualization skill training. Many researchers and practitioners face a major challenge of supporting sketching, an essential component in spatial visualization training, at scale. Because of the enormous error space, traditionally, a significant amount of human effort is required to grade and provide individualized feedback for students' technical drawings. Our team leveraged data mining and unsupervised learning techniques to build an intelligent sketching feedback tool. The tool not only allowed students to practice their sketching skills in a scalable manner but also graded students' sketches and provided customized and actionable feedback based on the error patterns in real-time. We deployed our tool in a university-level spatial visualization class with about 60 students. Students interacted with our tool for eight weeks. We performed an interview study to understand students' experience and attitudes towards using such an automated feedback tool for practicing spatial visualization skills. Through a grounded theory approach, we identified themes that informed our future improvement of the tool. We discuss the future design of sketching feedback tools in spatial visualization training in general.

Introduction

Spatial visualization skills, the ability to understand the visual and spatial relationship among objects, are among the strongest predictors of future success in STEM coursework and STEM careers. Through longitudinal studies, Wei et al. [1] and Shea et al. [2] have shown that students who have stronger spatial visualization skills are more likely to achieve educational and vocational success in STEM fields.

The malleable nature of spatial visualization skills motivates researchers and educational practitioners to develop spatial visualization skill assessments and training programs to facilitate student improvement. For example, Sorby developed a workshop using choice-based and free-hand sketching questions to train students' spatial visualization skills and showed significant post-test improvement in a 6-year-long study [3]. However, the strong demand for STEM education poses a new challenge to such in-person workshops -- the ability to scale. In recent years, many computer-based training programs have been created in response to the challenges in scaling. Wauck et al. developed a game-based method that involved virtual environment exploration and object creation for training spatial visualization skills [4]. Xiao et al. transformed an in-person workshop into a web-based online platform that evaluated and trained college students' spatial visualization skills [5]. Computer-based methods make spatial visualization training more accessible and scalable.

Practicing free-hand sketching is one of the most effective approaches to improve spatial visualization skills [6]. Sketching exercises require a student to mentally manipulate and visualize a 3D object and then draw the object on graphing or grid paper. Those exercises also imitate real engineering drawing tasks. Many computer-based training programs have developed online sketching interfaces to incorporate sketching exercises into their curriculum.

Formative feedback is known to support self-regulated learning [7]. However, it is challenging for a computer-based training program to provide customized formative feedback for students' sketches automatically. Writing constructive feedback to students' sketches requires domain expertise to determine what type of mistakes a student made since sketching questions could have an enormous number of possible incorrect answers. Therefore, though time-consuming, sketching exercises are often manually graded by instructors. A recent computer-based training program, eGrove's Spatial Vis [8], either provides limited feedback on the percent correctness or gives out very detailed hints that show students all the correct, incorrect, and missing lines. The former cannot provide insights on the cause of the mistake or how students might improve their answers [8]. The latter is very close to revealing the answer key [8]. Li and Paquette proposed a data-driven method that categorized students' erroneous sketches based on both the number of mistakes and the type of mistakes in a sketch [9]. The features used in the data-driven method showed great promise in providing specific and actionable feedback that characterize the mistakes based on their nature and severity [9]. However, it is unknown how to construct feedback with the data-driven method in [9] to help students learn from their mistakes.

Motivated by the challenges of automatically providing customized formative feedback for sketching exercises in computer-based sketching tools, in our work-in-progress, we detail an intelligent sketching feedback tool that aims to detect students' mistakes in their sketches and provide specific and actionable formative feedback. The feedback tool extends Li and Paquette's data-driven model [9] and can automatically deliver customized feedback messages based on the number of mistakes and type of mistakes in a student's sketch.

We deployed our tool in a spatial visualization training course with about 60 students in Fall 2020 in a college of engineering at the University of Illinois at Urbana-Champaign to evaluate our sketching feedback. The training course used a computer-based training system that required students to complete eight weekly online modules on various spatial visualization training topics, such as isometric drawings, orthographic drawings, object rotations, and inclined and curved surfaces. Students, on average, improved significantly on their spatial visualization skills throughout the course as measured by the Purdue Spatial Visualization Test: Visualization of Rotations (PSVT:R) [10], shown in Figure 1. At the end of the course, we conducted an interview study with 12 student volunteers to learn about their experience with and suggestions for our intelligent sketching feedback tool. Through a grounded-theory-like approach, we identified themes that informed our future iterations of the feedback tool. Our results contribute to the future design of sketching feedback tools in spatial visualization training in general.

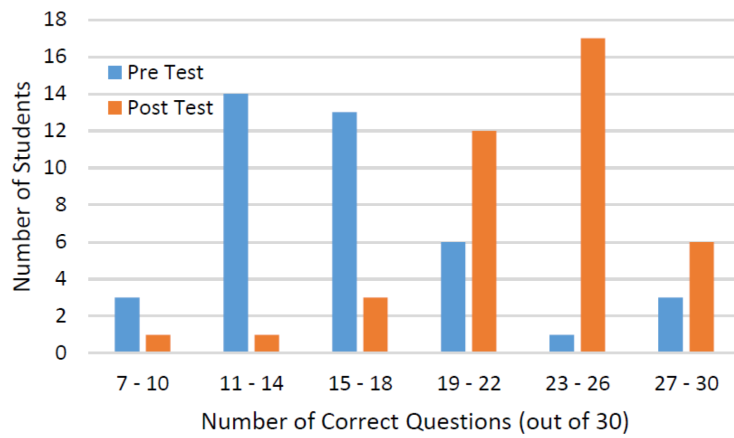

Figure 1. Number of students scoring in each bin range on both the PSVT:R test before and after the visualization training course.

Methods

Development and incorporation of the intelligent feedback tool: The intelligent sketching feedback tool was developed for educational purposes and incorporated in the Fall 2020 spatial visualization training course. During the spatial visualization course, six out of the eight weekly modules contained sketching exercises where the students drew their answers on our web-based sketching tool (see Figure 2). Students had two attempts per question. In the Fall 2019 course, many students expressed a need for more guidance via formative feedback between the two attempts. As a result, we trained and developed the intelligent sketching feedback tool to provide a customized feedback message in real-time upon an incorrect first submission. Here we introduce the feedback tool's salient features, but a complete description of the algorithm and its development process can be found elsewhere [9].

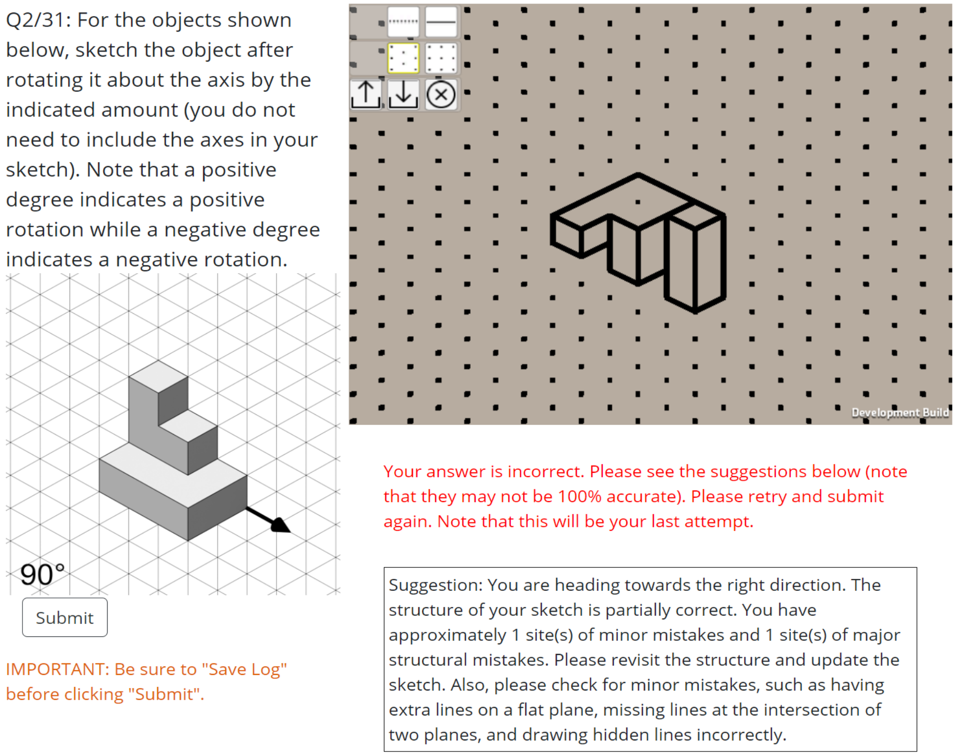

Figure 2. Screenshot of the sketching tool interface after an incorrect first attempt. The question prompt is on the left-hand side. The top right is the sketching pad. The message in red right below the sketching pad reports the auto-grading result. An example of the customized feedback message is in the box at the bottom right.

To generate customized feedback, the feedback tool first assigned a student's erroneous sketch into one of the five categories of incorrect answers (described in [9]). The categories were determined by the clustering outcome of a k-means clustering algorithm trained on students' erroneous sketches from Fall 2019. The five categories varied in the number of mistakes and severity of the mistakes in a sketch. Then, according to the category, the system chose the feedback message with the corresponding language and content. While the content of the messages differed, we structured each message in the same manner with three parts:

- **Overall correctness:** The first part of the message described how far the sketch was from the answer key. An answer got one of the three versions of the overall correctness message depending on its overall accuracy, including "almost there," "right direction," and "wrong direction."
- **Mistakes summary:** The next part of the message summarized the mistakes in the sketch. In most cases, it provided the number of minor and major mistakes that occurred in the sketch. The number of mistake sites was determined by the number of graph components in the undirected graph consisting of all the incorrect lines (both missing and extra lines). A minor mistake was a graph component with four or fewer unit lines; otherwise, it was a major mistake. In cases where most incorrect lines were clumped

together as one mistake over half the size of the sketch, the message did not include a mistakes summary.

- **Suggestions for next step:** At last, the message suggested potential next steps for the students. For minor mistakes, the message listed out prevalent causes for students to check. For major mistakes, we ask them to review the structure of the sketch. As for cases where there was one major mistake larger than half the sketch, we reminded the student of the possibility of the object facing in the wrong orientation.

At the beginning of the semester, the instructor briefly mentioned the existence of the feedback message. However, the class provided no information about how the message was generated or how to interpret it.

Participant recruitment: To learn about students' experience with the feedback messages and garner ideas for improvement, we conducted IRB-approved semi-structured interviews with students who used our sketching tool and received feedback for their submissions. To recruit participants, we sent emails to all 60 students who were enrolled in the spatial visualization training course in Fall 2020 after the final course grades were released. We recruited 12 students from the class, seven women and five men. Four interviewees were White, six were Asian or Pacific Islander, one was Hispanic or Latino, and one participant's race and ethnicity were unknown. All of them were in the first semester at our university. Eight of them had a final grade on or above average, and four performed below average. The sample was representative of the whole class in terms of race and ethnicity, and class performance. We had a higher percentage of females in the interview sample than in the class.

Interview procedure: We conducted the semi-structured interviews over Zoom. The average interview time was 33.16 minutes, ranging from 28 to 40 minutes. They received compensation at a rate of $15/hour. The interview focused on students' interpretation of the feedback messages as well as the strategies they employed and the challenges they faced when using the feedback to revise their submissions. In the interview, we first showed participants an example feedback message to remind them of the messages and set our interview context. Then the rest of the interview included four sections.

- In the first section, we showed three types of feedback messages students might encounter; we will refer to them as "almost there," "right direction," and "wrong direction" messages in this paper. All students mentioned they had seen all three kinds of feedback messages, though one student was unsure about one type of message. We asked what sort of incorrect answers they thought received this type of feedback and their subsequent strategies to revise their submission for each type of feedback message.
- In the second section, we asked students to label potential mistakes in a sketch (see Fig. 2). Then, students shared their interpretation of specific wordings in our feedback messages, namely "sites of mistake" and "major/minor mistakes."

- In the third section, we gauged student’s opinions on how valuable each part of a feedback message was and how they dealt with inaccuracy in each part of the feedback message.
- In the last section, we concluded our interview by asking about the student's overall experience with the feedback system, the degree of satisfaction, and ideas for improvement.

Analysis methods: We analyzed our data using a grounded theory approach, where we inductively generated codes that summarized each section of the interview with a reflective and iterative process. Before analyzing the data, we first transcribed the audio recordings into text (upon participant’s consent). First, two of the research team members picked three interviews and started with thematic coding independently. Then the two members came together, discussed, and revised the codes. With the agreement of a set of seed codes, the two coders then coded the remaining interviews. In the end, the two members adopted a semantic approach to iteratively refine the themes of students’ interpretation of the messages, strategies applied, and ideas for improvement.

Results and Discussion

In this section, we first summarize students’ satisfaction and perceived helpfulness of the feedback message. Next, we present three themes of improvements derived from the interview results and discuss the implications for future implementation of similar intelligent systems.

Satisfaction and perceived helpfulness: Among the 12 interviewees, nine reported they were very satisfied or pretty satisfied with the feedback message (N=9). The remaining three found the message satisfactory in some cases (“almost there” and “right direction”) but not others (“wrong direction”) (N=3).

Students provided different perspectives on how the feedback message helped them. Five students found the overall correctness part of the message assuring and effective in avoiding unnecessary rework (N=5). Four students mentioned that the message helped them realize the type of mistakes they should look for in the sketch (N=4). Three students said the message made it easier for them to count and find their errors, especially minor mistakes (N=3).

Besides helping students solve a particular problem, the feedback message also helped students learn in the long term. Three students commented that as they practiced identifying errors in a sketch per the feedback message, they learned to avoid similar mistakes in the future and improved the ability to review a sketch and fix the mistakes independently over time (N=3).

Increase system transparency to align message interpretation: As we elicited the interviewees' interpretation of the feedback message and their strategies towards each type of feedback message, we observed several discrepancies between what the message intended to convey and what students thought the message meant. The discrepancies in message interpretation included the definition of a "site" of mistake, the definition of a minor mistake versus that of a major mistake, and the perceived degree of mistake severity when a sketch received a "right direction" message. Here, we elaborate on the interpretation of what constitutes a "site" of mistake.

One of the most significant discrepancies lied in the definition of a "site" of mistake, which determined the number of mistakes reported in the feedback message. Half of the interviewees interpreted a site of mistake as a set of incorrect lines that shared the same cause (N=6). In the system, the algorithm grouped a set of lines into the same site of mistake by the proximity of their locations, i.e., whether they are interconnected. The algorithm was designed under the assumption that incorrect lines that were closer were more likely to share the same cause, so in cases in which incorrect lines caused by multiple reasons were interconnected, the algorithm could not separate the incorrect lines based on their causes.

As an example, in the interview, we asked students to locate all the sites of mistakes in the incorrect sketches in Figure 3. According to the system, this erroneous sketch consists of two mistakes, one in the front and one in the back (as shown in the left-most sub-figure in Figure 3). However, eight participants considered the sketch to have three sites of mistakes (N=8) (as shown in the middle sub-figure in Figure 3). They reasoned that they would separate the incorrect lines in the back into two distinct sites because even if one moved the pillar to the correct position, the extra line on the side (marked as mistake 2 in the middle sub-figure) would still be there as it is not caused by the pillar being in the wrong place.

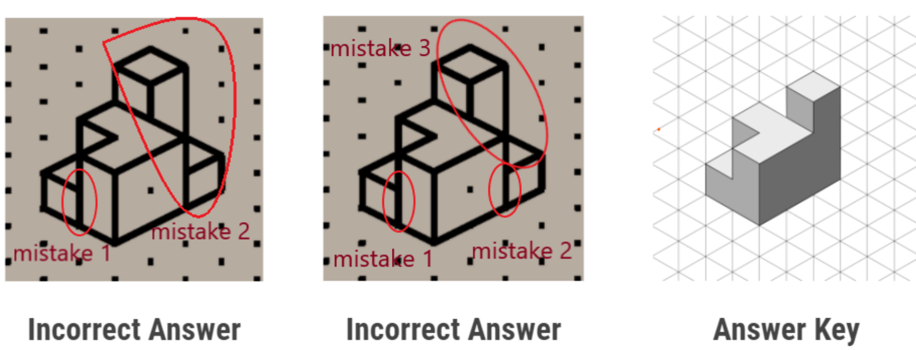

Figure 3. An example used in the interview to gauge students' interpretation of what constitutes one site of mistake by asking them to compare the incorrect answer with the answer key and identify all sites of mistake. Starting from the left, the first sub-figure shows how the system defines the sites of mistakes in this sketch by the proximity of location. The second sub-figure is a popular response from the interviewees who distinguished different sites of mistake by their cause.

While it makes sense conceptually to define a site of mistake based on the cause, it remains challenging for an algorithm to fully understand why a student draws a line incorrectly. As a result, students may perceive the feedback message as inaccurate due to a discrepancy in the definition. While the interviewees could not recall any instances of inaccurate feedback, when we asked them hypothetically how they would feel if the feedback message was inaccurate, seven interviewees reported that they might experience confusion, frustration, and a drop in trust towards the feedback message (N=7). Even more concerning, three students exhibited a tendency to overtrust the system and commented that they would doubt themselves if the number in the feedback message differed from their own counting (N=3).

Based on the above results, we observed a need to increase the transparency of the intelligent feedback tool to align students' mental models with how the system worked. The instructor in the Fall 2020 course did not provide any instruction about the feedback system to the students. A majority of interviewees reported that they formed their own definitions via finding patterns during practices over the semester. Four interviewees suggested including an introduction to the feedback tool may help them understand how the system defines the key terminologies (N=4). Another approach to align message understanding, proposed by five participants, is to show students the minor and major mistakes identified by the system after the final attempt so that students can more easily find the patterns of the system's decision-making process (N=5). This finding may inform the stakeholders who deploy other similar black-box algorithmic feedback systems in the future to carefully consider the degree of transparency needed for users to make sense of the result.

The tradeoff between feedback specificity and educational value: The current feedback message informed students of the types of mistakes and the number of mistakes. It did not give information about where the mistakes were or what the mistakes looked like, a critical difference between our system and other popular training platforms such as eGrove's Spatial Vis [8]. In addition, the message provided general suggestions for where to look for a minor or a major mistake, but we did not customize the suggestions for the specific mistake. There were two design considerations with these decisions. First, we kept the feedback less specific, intending to focus on training students' ability to reflect on their answers and solve the problem independently, confirmed and appreciated by three interviewees. Second, there were technical challenges to customize suggestions for major mistakes since it was hard for an algorithm to analyze why a structural error happened.

Seven interviewees reported difficulties in identifying all mistakes mentioned in the feedback message (N=7). Some of them struggled especially with major mistakes, and as a result, they ended up redrawing the sketch even when they were heading in the right direction. When we asked the interviewees how they would like to improve the feedback message, many expressed their preference for more specific and customized feedback. Three of them suggested marking the locations of the mistakes on the sketch as a part of the feedback (N=3). Two interviewees wanted to include a more detailed description of the mistakes in the message, such as a pillar being in the wrong place or a pillar being too tall (N=2). Another suggestion was to only include relevant suggestions in the message based on the mistakes in the sketch, as proposed by two participants (N=2).

While students may naturally desire more detailed hints, four interviewees recognized and understood the tradeoff instructors faced between the level of feedback specificity and the amount of learning that occurred in the process (N=4). Two of them thought the current level of detail was just right for effective learning (N=2).

Synthesizing interviewees' opinions above, as a next step to improve the feedback system, we will strive to increase specificity to the best of our ability without compromising the learning goal. For example, to help students gain confidence in identifying major mistakes when their answers are heading in the right direction, the feedback message can include hints for the general area where the major mistake is located, e.g., bottom vs. top, left vs. right. Hinting at a general area instead of the specific location preserves the need to problem-solve independently. For minor mistakes, the feedback can characterize the composition of the minor mistakes, e.g., extra lines vs. missing lines, solid vs. hidden lines. Giving away location information for minor mistakes may take away students' practice opportunities to carefully reflect on their own drawing. While these two improvement ideas are specific to our system, the implication of the interview results may generalize -- feedback system designers may need to engage in iterative design with instructors and students to find the optimal balance between achieving the learning goals and providing sufficiently detailed feedback.

When a feedback message alone does not help: In the interview results, one consistent theme was students being unsatisfied with the feedback message when it said their answers were heading in the wrong direction. Even though some interviewees found the feedback helpful in cases where their sketch was structurally correct but had the incorrect orientation, they felt stuck in scenarios where the sketch structure was significantly different from the answer key. As a result, ten interviewees always redrew the sketch when they received the "wrong direction" feedback without considering the number of mistakes mentioned in the message if given any (N=10).

While we understood students' frustration with unhelpful feedback messages when their objects were very far from the answer key structurally, we could not think of many improvements on the message itself to resolve the issue. We drew inspiration from some of the strategies students reported in the interviews and concluded that additional help from other channels might be a good alternative. One student would refer to course materials or past questions they have answered to improve their understanding of the current problem (N=1). Another student used the strategy of asking for external help from peers and instructors if they kept getting the "wrong direction" message (N=1). Both of these approaches made sense since students heading in the wrong direction likely had trouble understanding the concept or visualizing a specific kind of shape.

Therefore, in case of a very wrong answer, we should focus on helping students clear up misconceptions and learn the general approach to visualize a kind of shape rather than limiting the assistance to solving a specific question. Examples of additional assistance include offering easier practice problems that test skills similar to the current problem and encouraging students to ask for help, especially when they have repeatedly gone in the wrong direction. Future feedback system designers should keep in mind the limitations of a simple feedback message and be open to the use of other resources and channels.

Conclusions and future work

In summary, this paper presents a user study of an intelligent feedback tool that provides customized real-time formative feedback on sketching exercises for spatial visualization training in the field. We conducted semi-structured interviews with 12 students that experienced the feedback system in an 8-week online spatial visualization training course in Fall 2020. The majority of the interviewees were satisfied with the feedback messages they received and found them helpful in identifying errors in their sketches and improving their ability to reflect and solve spatial visualization problems independently over the semester. The interview results revealed three themes of improvements for the tool and implications for future designers of similar feedback systems:

1. Increase the transparency of the intelligent feedback tool to align users' interpretation of the messages with how the system defines the messages;
2. Find the optimal balance between achieving the learning goals and providing sufficiently detailed feedback; in our case, increase the level of details included in the feedback messages to the best of our ability without compromising the learning goal;
3. Utilize additional resources and channels besides feedback messages to help students who repeatedly head in the wrong direction with their sketches.

This work is presented as a Work in Progress because we have yet to incorporate the above improvements for the feedback tool and evaluate their effectiveness. The research team plans to continue the study by implementing the improvements and conducting user experiments in Fall 2021.

References

[1] J. Wai, D. Lubinski, and C. P. Benbow, “Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance.” *Journal of Educational Psychology*, vol. 101, no. 4, pp. 817–835, Nov. 2009

[2] D. L. Shea, D. Lubinski, and C. P. Benbow, “Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study.” *Journal of Educational Psychology*, vol. 93, no. 3, pp. 604–614, 2001

[3] S. A. Sorby, “Educational Research in Developing 3‐D Spatial Skills for Engineering Students,” International Journal of Science Education, vol. 31, no. 3, pp. 459–480, Feb. 2009

[4] H. Wauck, B. S. Woodard, Z. Xiao, T. W. Li, and B. P. Bailey, “A Data-Driven, Player-Centric Approach to Evaluating Spatial Skill Training Games,” in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, Virtual Event Canada, Nov. 2020, pp. 349–361

[5] Z. Xiao et al., “A Scalable Online Platform for Evaluating and Training Visuospatial Skills of Engineering Students,” in *2017 ASEE Annual Conference & Exposition Proceedings*, Columbus, Ohio, Jun. 2017.

[6] S. A. Sorby, "Developing 3-D spatial visualization skills." *Engineering Design Graphics Journal* 63, no. 2 (2009).

[7] N. David J., and D.Macfarlane‐Dick. "Formative assessment and self‐regulated learning: A model and seven principles of good feedback practice." *Studies in higher education* 31, no. 2 (2006): 199-218.

[8] E. Cowan, N. Delson, B. Mihelich, and L. Van Den Einde. “Improvement in Freehand Sketching Application for Spatial Visualization Training.” In *Inspiring Students with Digital Ink*, pp. 121-134. Springer, Cham, 2019.

[9] T. W. Li and L. Paquette. “Erroneous Answers Categorization for Sketching Questions in Spatial Visualization Training.” *International Educational Data Mining Society*, 2020.

[10] B. George M., and R. B. Guay. "The Purdue visualization of rotations test." *The chemical educator* 2, no. 4 (1997): 1-17.