2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #51267

Think-Aloud and Video Data to Improve a Spatial Skill App: The Journey of Learning and Discovery

Jamie Lynn Smith, Embry–Riddle Aeronautical University, Daytona Beach
Bianca Nicole Gerity
Dr. Heidi M Steinhauer, Embry–Riddle Aeronautical University, Daytona Beach

Heidi M. Steinhauer, Ph.D. is a Professor of Engineering Fundamentals, the former Inaugural Associate Dean for Undergraduate Studies, the former Inaugural Department Chair of the Engineering Fundamentals Department, co-advisor for the only all-women's Baja SAE Team, and Founding Member of FIRST (Female Initiative Reaching Success Together).

Lorraine Margot Acevedo, Embry-Riddle Aeronautical University - Daytona Beach
Juan Francisco Granizo, Embry-Riddle Aeronautical University - Daytona Beach
Kai Jun Chew, Embry-Riddle Aeronautical University - Daytona Beach

Kai Jun "KJ" Chew is an assistant professor in the Engineering Fundamentals department at Embry-Riddle Aeronautical University. He is passionate about teaching and research, and he strives to produce knowledge that informs better teaching. His research intersects assessment and evaluation, motivation, and power dynamics. His research goal is to promote improvement in engineering classrooms through these research intersections.

Using Think-Aloud and Video Data to Enhance a Spatial Skill App: The Journey of Learning and Discovery

1. Introduction

Spatial skills, defined as the capability to manipulate mentally 2D and 3D objects that can be acquired through formal training, are crucial in facilitating success among engineering students [1], [2]. Findings have shown that spatial skills can predict student success in STEM fields, with the effects being more prominent for beginner students while learning in STEM [1], [2], [3], [4], [5]. In engineering, especially developing spatial skills, such as the orthographic and isometric views, is quintessential due to these views being the primary form of communication in engineering graphics [2], [6]. Therefore, efforts to improve spatial skills are a worthy endeavor in engineering education. Research has shown that students' spatial skills can be improved through targeted instructional interventions, particularly when provided with visual, interactive, and progressive, scaffolded learning experiences [1], [7], [8], [9], [10], [11]. Recognizing that these skills are not innate or fully developed, current graphics education seeks tools that explicitly develop spatial reasoning during problem solving.

Examples of tools that have been introduced to support spatial skills development in graphics education are physical models, glass-box demonstrations, videos, animations, and computer-based visualization software [6], [12], [13], [14], [15], [16], [17]. While impactful, these tools are not without limitations including static, monochromatic, non-interactive, or restricted to specific physical locations such as laboratories or classrooms. Many of these tools present only the final 3D object, missing the opportunity to offer students the experience of scaffolding offered by step-by-step construction of the 3D model.

Our project contributes to these efforts by developing an app that can help supplement student learning of spatial visualization skills, with the focus on visualizing and manipulating components to build a part, step-by-step with color-coded components. This modality is to prepare students for a computer-aided design course where students are expected to learn the steps in modeling parts.

The app development process has involved the building and testing, with the focus of using think-aloud protocol as the preliminary testing of the app to identify improvements. The use of think-aloud, with qualitative data analysis, is appropriate because it can be beneficial in testing the app prototype through explicitly understanding the participant's thought process while using the app [18]. However, based on our literature review, using the think-aloud protocol to test a spatial skill learning app is not widely documented. We have found the lack of documentation of such app development and testing, with most works focus largely on the implementation as part of the curriculum [6], [16], [19], [20], [21]. Our work will contribute to this research gap by detailing our app testing process using a think-aloud protocol, providing future researchers and developers in the spatial skill learning technologies domain lessons and challenges for future work.

Our work focus on two objectives: 1) To test the content (or the thought process flow of the students) while using the app, and 2) to test the students' interaction with the app (or the user interface of the app), with Table 1 detailing the methods and intended outcome.

Table 1: Testing Study Design

Testing Objective	Method	Intended Outcome
To test the content (or the thought process flow of the students) while using the app	Think-aloud protocol with audio video recording, with qualitative coding on transcriptions	Emerging thought process on the flow in moving the blocks to create the complete part
To test the students' interaction with the app (or the user interface of the app)	Hand movement recordings with qualitative coding	Emerging interaction process and flow while using the app

2. Background of the SSTAR app

The SSTAR app has been in development over the last two years, with the goal of creating a supplemental app that helps students enhance their spatial skills [22], [23], [24]. The app is created and developed using the Unity platform. This app differentiates itself in that it provides step-by-step, color-coded, along with specific feedback about hidden lines to provide students with a more guided experience while improving their spatial skills.

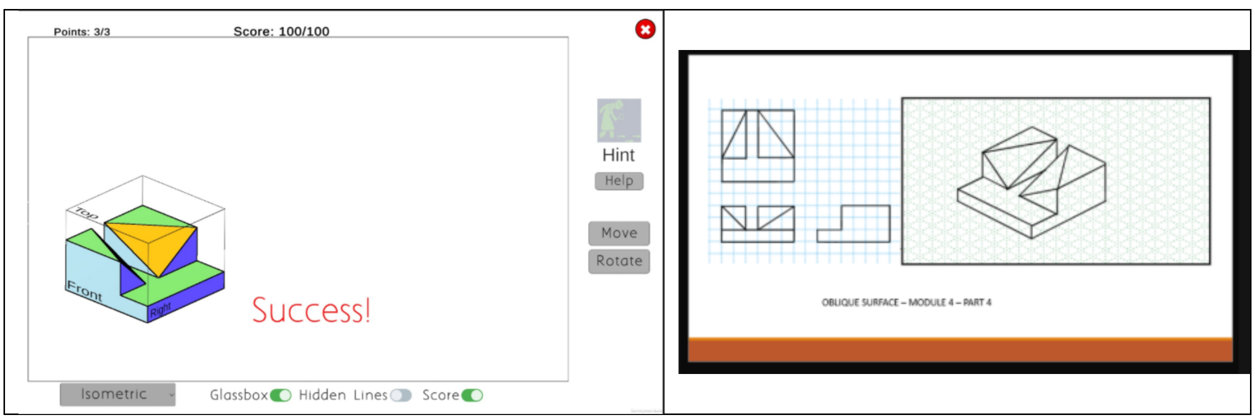


Figure 1: The latest iteration of the SSTAR app (left), with the idea that students would use this to assist them in completing a drawing exercise (right).

The initial goal was to create an augmented reality app as part of this research grant. However, it has proven to be challenging to export the visualization challenges into the Unity platform. Hence, our team's goal currently is to create a tablet-based app to first assist students with their spatial skills through learning and demonstrating their skills with drawing exercises, with future goals to translate this app into an augmented reality environment.

3. Literature Review

Using think-aloud protocol is appropriate for testing a learning app prototype in evaluating its pedagogical strategies. By understanding the thought process of user while using the app [18], one can have the capability to parse out the perceptions and experiences of the user and identify possible misalignment between these perceptions and experiences and the intended use of the learning app for improving spatial skills [25], [26].

There are three different forms that must be discussed and considered with think-aloud protocols. Alhadreti and Mayhew [27] classifies these as concurrent, retrospective, and hybrid. Cooke [28] looks at the concurrent method in more detail, and its overall usability. The concurrent method of the think-aloud protocol is the process where the participant is asked to voice their thoughts during the experiment. This can be seen as an issue as it can hinder the participants' thought process and overall performance. Olmsted-Hawala et al. [25] puts these protocol variations into comparison through testing website usability. This experiment looked at traditional, speech-communication, and coaching think-aloud protocols. The traditional think-aloud protocol is the same as the concurrent protocol, while the speech-communication protocol and the coaching protocol is an extension of the concurrent protocol. This is not the only study that has been done testing these methods, for example, Fischer et al. [29] used these methods to determine if their app could increase people's physical activity. However, both experiments found that the Concurrent/Ericsson and Simon method of the think-aloud protocol has some disadvantages. It was found that participants typically didn’t completely follow the protocol, because the idea of speaking all of your thoughts can be seen as uncomfortable. These methods have also been used in the medical field to understand the importance of computer software that oncologists used, which was discussed in Jaspers et al. [30]. This study also leaned toward the concurrent method of the think-aloud protocol, and it was found that this method was more successful due to their feedback coming from their working memory, rather than the long-term memory.

Each of these methods has their own advantages and disadvantages; these implications are analyzed in Brunn and Stage [31]. This article classifies each protocol as traditional (Concurrent/Ericsson and Simon), Coaching (extensions of traditional), and active listening (extension of traditional). The traditional method is used with the intent to alleviate any questions as to why a participant might be doing something during a study. This method, however, is only used when it will not distract the participant. It was also found during this study that this method is unnatural and can be difficult for the participants to actually verbalize all of their thoughts. An expansion of this protocol is the active listening protocol. Instead of directly asking the participant to continue, the interviewer would have the participant continue through active listening. The coaching method would be on the other side of the traditional method where the interviewer keeps the participant speaking through, what is described as empathetic encouragement, or asking questions. It was found that based on their testing, the traditional protocol had the best outcome.

However, these are not the only classifications for types of think-aloud protocols. Less popularly, Krahmer and Ummelen [32] arrange these into two categories, the Ericsson and Simon method, and the Boren and Ramey method. As previously mentioned, the Ericsson and Simon method is also known as the concurrent or traditional method. While the Boren and Ramey method requires two interviewers, one interviewer prompts questions for the participant, while the second interviewer is carrying out the active listening method described earlier. This research suggests that the Boren and Ramey method was generally more successful; however, it is acknowledged that this could be experiment-dependent. This method is a hybrid method between concurrent and active listening. Wang and Lien [33] look at the advantages of integrating these methods while also collecting video data.

4. Method

4.1 Equipment Setup

The app was tested using the configuration shown in Figure 2. A document camera was connected to a laptop, and Zoom was used to record the testing session to capture participants' interactions with the app and the drawing assignment. The document camera was elevated using stacked books to increase the visible workspace and ensure all interactions were recorded.

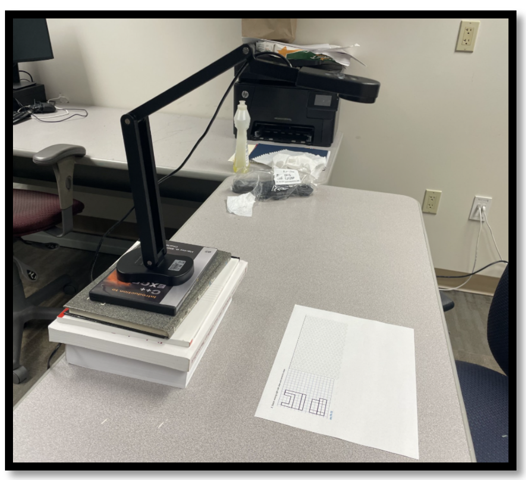


Figure 2: Document camera recording setup for testing.

4.2 Setup for the Think-Aloud

We used a hybrid think-aloud protocol where the participants expressed their thoughts while using the app, in addition to an interview after the testing is completed. Literature has shown that having a follow-up interview session could help with triangulation with retrospective data, considering that the think-aloud during a task could alter the participants' thought process and possibly shape how they would perform on a task [27].

To allow participants to become comfortable voicing their thoughts in the testing environment, various puzzles were introduced to the participants before introducing them to the app. Before participants were shown the puzzles, they were instructed to verbalize all thoughts that passed through their head and that we were not interested in if the thoughts were right or wrong, that we were more interested in their thought process. The puzzle was then introduced to the participants, and they began to solve the puzzle and verbalize their thoughts. If we noticed anything interesting, or the participant began to move puzzle pieces without verbalizing, we prompted the participant to verbalize their thoughts.

Once the participant was comfortable, the tutorial video and the app were introduced to the participants. We asked if the participant had any questions regarding how to use the app and reminded them to verbalize their thought process. While the participant was using the app, we would ask questions about their thought process or about problems they encounter when using the app.

Once the participant finished using the app, we asked for feedback regarding the tutorial video and their overall feelings using the app. We also asked for feedback on improving the app in future iterations.

4.3 Recruitment

The target participants for this study were engineering students who were currently enrolled in or had previously completed the computer-aided design course due to the app being under development. Participants were recruited through an informational flyer being disseminated via the various computer-aided design courses' Canvas pages, which included a QR code to receive the informed consent documentation to consent in the study. Participants in the study were rewarded with Sodexo vouchers totaling $15. A total of four students were recruited and participated in this study.

4.4 Data Analysis

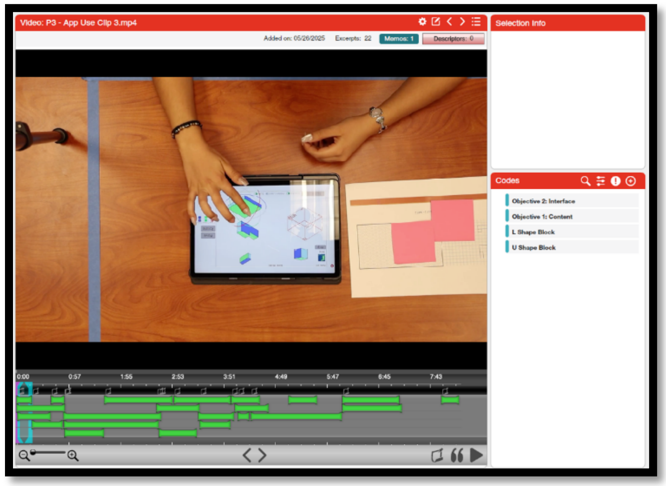


Figure 3: Using Dedoose to code recorded video and think-aloud

Based on existing think-aloud protocol literature, it was decided to conduct qualitative coding based on [34] a priori coding. We created two codes: Content and Interface, the two identified objectives of the research. Coding was conducted on Dedoose, with the videos being excerpted and categorized with both “content” and “interface” codes (Figure 3). Four researchers conducted the coding of the think-aloud data independently. Each researcher was assigned to code all four recordings. There were three iterations of coding. During the three iterations of coding, each of the other researchers would review the others' coding and add comments/or clarifications as needed, resulting in the final coding shown in Table 2.

Table 2: Final codebook for data analysis process.

Code	Definition	Code Type	What to look for with this code	Examples as boundary setting
Content	How students engage their spatial skills (knowledge they currently have on how they visualize) while using the app	A priori	Students are saying out-aloud their thought process (***verbalizations***) while they select, move, rotate, and place the blocks into the glass box (***hand movement***).	The students describing what they were doing at the moment while manipulating (move/rotate) a block to the correct placement and may include explanations on what they were thinking about how they were trying to manipulate and place the block into the glass box to form the part.
			We need to make judgment whether the students' thought process is affected by the app glitches or issues.	If a student is placing a block correctly but is facing challenges in getting the app to recognize the correct placement, then the coder can make the judgment that this is not a "content" excerpt, but an "interface" one (it is an interface problem).
Interface	How students engage and respond to the app (features, move/rotate, glitches, issues, and others)	A priori	Verbal cues and explanations (***verbalizations***) of the students while they move and rotate blocks, place the blocks into glass box, use the hint feature, use the different view feature (***hand movement***)	If a student is sighing or exasperated with sound, look for what the student was responding to. If is related to the interface (like the block jumping back after not recognizing the correct placement), then this must be coded as "Interface"
				If a student realizes the need to use hint, this should also be coded as an interface "excerpt"

The team of researchers met in person several times to review the coding of the recordings to ensure inter-rater reliability (IRR) of the coding process. We employed [35] method to assess IRR. Since there were four coders, [35] recommends having IRR calculated for each pair, meaning six pairs with six IRR calculations. The IRR score is a percentage calculated based on the number of excerpts coded to the same code (agreement) divided by the total number of excerpts coded (agreement and disagreement). Table 3 shows the IRR scores for each pair.

Table 3: IRR Calculations for the Coding Analysis

Pairing	IRR Score (%)
1 – 2	84
3 – 2	84
4 – 2	80
3 – 1	54
3 – 4	76
1 – 4	75

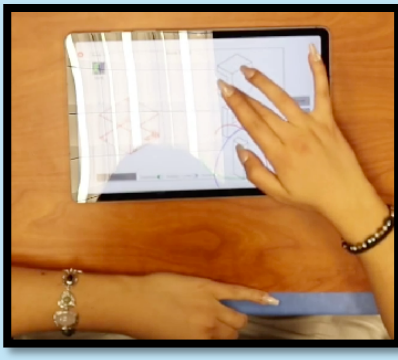
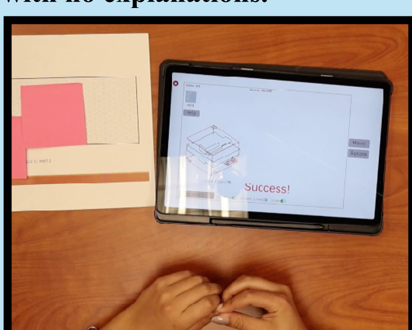
These scores were used to determine the different interpretations of the codebook. It is generally accepted that a score of 75% or higher is an indication of good interrater reliability [36]. While there is one rate pair (4 –1) score below this threshold, we believe the multiple group discussions addressed this and led to consistent agreement of the final interpretations of the codebook. This codebook was used to code the data for the final time before proceeding to the next steps.

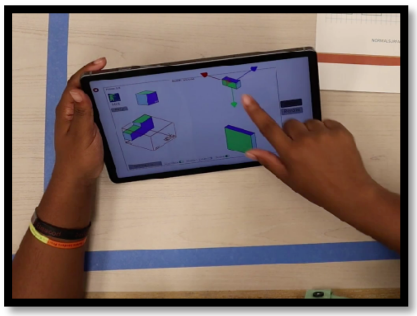
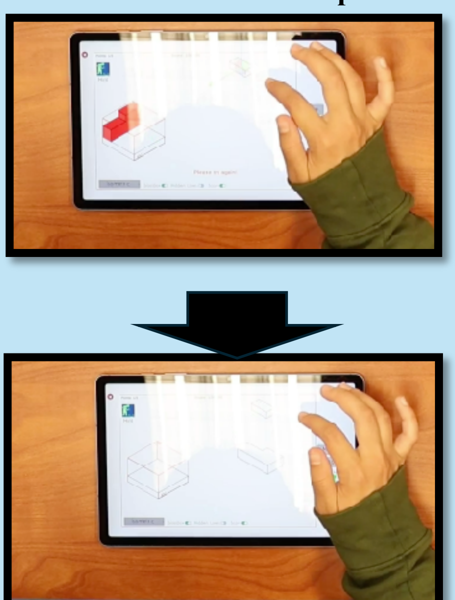
After the coding, we categorized all excerpts based on the codes to identify the areas of improvements by identifying patterns of what the participants were discussing and how they were using the app based on content and interface. These patterns became the outcome of our results.

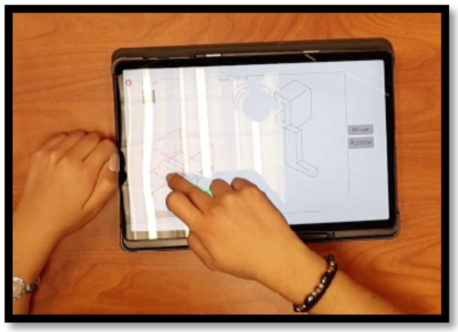
5. Results

Our research team has found the use of think-aloud protocol to be helpful in identifying areas of improvement. From the testing sessions with the four participants, we identified that the app needed several major addresses from the content and interface perspectives. Specifically, the use of think-aloud protocol to test the app yielded positive outcomes as our team was able to quickly identify these areas of improvements. When participants shared their thought process while using the app to complete the drawing exercises, we observed their behaviors while using the app with added contexts on why participants were experiencing the app in a certain way. This quickly allowed us to identify these preliminary areas to be addressed. The subsequent data analysis affirmed the following list of areas of improvements, with images and quotes selected from one participant from the testing as data to back up the claims. Table 4 summarizes the areas of improvement identified by each participant. These areas of improvement have appeared consistently among all four users in the data, except for the feedback as P2 did not raise or mention any discussion about the need for feedback.

Table 4: Areas of improvements for both content and interface identified by each participant from the testing.

Areas of Improvement	Comments			
	P1	P2	P3	P4
The app hardcoded certain solutions that imposed a specific way of visualizing users by only allowing them to pick certain blocks to get started. This could potentially harm students in imposing only a certain way of visualizing the problems on the students.  ***"Oh… I guess I can start with a different one…"*** **[P3 thought process while attempting to select a block where the app did not respond to their touch or movement]**	Y	Y	Y	Y
The app did not allow users to freely explore block movements without being informed whether their placement of the blocks was correct or incorrect. Providing students the flexibility to be able to choose can empower them to think about different ways of visualizing. **"*Be able to fully define geometry too…. I guess like a sandbox mode where you can like….if it gives you a part, and you get to choose which part to make it, I guess that could be fun…*" [P1, explaining the need for feature for students to choose how they could assemble the blocks into the part freely]**	Y	Y	Y	Y
The app should include more specific feedback if the blocks were placed at an incorrect position. The app only informed the user of correct or incorrect placement with no explanations. 	Y	N	Y	Y

<table>
<tr><th rowspan="2">Areas of Improvement</th><th colspan="4">Comments</th></tr>
<tr><th>P1</th><th>P2</th><th>P3</th><th>P4</th></tr>
<tr><td>"I think…because I wasn't putting it in the right orientation, I was confusing the two colors, so it would be kinda cool hinted on what side is at the right place." [P3 explaining why the need for feedback in addition to the hint as they were confused with the hint initially, and there was no feedback guiding her to think about placement.'</td><td></td><td></td><td></td><td></td></tr>
<tr><td>The color-codes for the different views should be the default when the app opens at the beginning instead of being a "hint."

"This piece is kinda like the bottom piece cause it has no no type of special shape, it's just kinda a block, so the only way you could solve this one is to turn on the hint cube…" [P2 when explaining why they needed to use the hint feature to solve the part].</td><td>Y</td><td>Y</td><td>Y</td><td>Y</td></tr>
<tr><td>The blocks, after being placed in an incorrect position, should not jump back to the original position. Users have found this to be frustrating, considering that most of them had the correct placements.

"[exasperated], I guess that wasn't right [it was the right placement based on the drawing]…[deep breathe] I just want to do it." [P4 while expressing her frustrations that the block kept jumping back if the app deemed the placement incorrect].</td><td>Y</td><td>Y</td><td>Y</td><td>Y</td></tr>
</table>

Areas of Improvement	Comments			
	P1	P2	P3	P4
The sensitivity where the blocks are being placed at the correct positions may need to be decreased as there were multiple attempts where the blocks were placed at the correct place, but it did fall within the range, resulting in "incorrect" designation  "***Alright, hopefully it would stay…***" **[P3 while hoping that the app received their placement after several times where the app deemed it incorrect even though the placement was correct].**	Y	Y	Y	Y

6. Future Work

The latest version of the app is being tested to identify more areas of improvement, with the goal of a full classroom implementation of the app in Fall 2026. Although data analysis has not commenced for the latest version of the app, our observations of the participants (think-aloud and hand movements) during the testing imply substantial improvements of students' user experiences with the app. Future work will include reporting out the outcome of data analysis to assess improvements, in addition to the testing of the app in classroom environment.

References

[1] S. A. Sorby, N. Veurink, and S. Streiner, "Does spatial skills instruction improve STEM outcomes? The answer is 'yes,'" *Learning and Individual Differences*, vol. 67, pp. 209–222, Oct. 2018, doi: 10.1016/j.lindif.2018.09.001.

[2] 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, 2009, doi: 10.1080/09500690802595839.

[3] J. Wai, D. Lubinski, and C. P. Benbow, "Spatial ability for STEM domains: Aligning cover 50 Years of cumulative psychological knowledge solidifies its importance," *Journal of Educational Psychology*, vol. 101, no. 4, pp. 817–835, Nov. 2009, doi: 10.1037/a0016127.

[4] L. E. Margulieux, "Spatial encoding strategy theory: The relationship between spatial skill and STEM achievement," *ACM Inroads*, vol. 11, no. 1, pp. 65–75, Oct. 2020, doi: 10.1145/2792790.

[5] D. H. Uttal and C. A. Cohen, "Spatial thinking and STEM education. When, why, and how?," in *Psychology of Learning and Motivation - Advances in Research and Theory*, vol. 57, 2012, pp. 147–181. doi: 10.1016/B978-0-12-394293-7.00004-2.
[6] N. Delson and L. V. Den Einde, "Sketching, assessment, and persistence in spatial visualization training on a touchscreen," in *Proceedings of the American Society for Engineering Education (ASEE) Annual Conference*, 2018.
[7] A. U. Gold *et al.*, "Spatial skills in undergraduate students-Influence of gender, motivation, academic training, and childhood play," *Geosphere*, vol. 14, no. 2, pp. 668–683, Feb. 2018, doi: 10.1130/GES01494.1.
[8] T. R. Lord, "Enhancing the visuo-spatial aptitude of students," *Journal of Research in Science Teaching*, vol. 22, no. 5.
[9] A. S. Alqahtani, L. F. Daghestani, and L. F. Ibrahim, "Techniques used to Improve Spatial Visualization Skills of Students in Engineering Graphics Course: A Survey," *ijacsa*, vol. 8, no. 3, 2017, doi: 10.14569/IJACSA.2017.080315.
[10] P. Connolly, L. V. A. Harris, and M. Sadowski, "Measuring And Enhancing Spatial Visualization In Engineering Technology Students," in *2009 Annual Conference & Exposition Proceedings*, Austin, Texas: ASEE Conferences, Jun. 2009, p. 14.868.1-14.868.9. doi: 10.18260/1-2--4712.
[11] S. Sorby and B. J. Baartmans, "The development and assessment of a course for enhancing the 3-D spatial visualization skills of first year engineering students," *Journal of Engineering Education*, vol. 89, no. 3.
[12] S. Tumkor and R. De Vries, "Enhancing spatial visualization skills in engineering drawing courses," in *2015 ASEE Annual Conference and Exposition Proceedings*, Seattle, Washington: ASEE Conferences, Jun. 2015, p. 26.663.1-26.663.12. doi: 10.18260/p.24001.
[13] L. Van Den Einde, N. Delson, and E. Cowan, "Freehand Sketching on Smartphones for Teaching Spatial Visualization," in *2019 ASEE Annual Conference & Exposition Proceedings*, Tampa, Florida: ASEE Conferences, Jun. 2019, p. 32859. doi: 10.18260/1-2--32859.
[14] G. Frey and D. Baird, "Does rapid prototyping improve student visualization skills," *Journal of Industrial Technology*, vol. 16, no. 4, 2000.
[15] Zheng Jian, "Teaching of engineering drawing in the 21st century," in *2011 Second International Conference on Mechanic Automation and Control Engineering*, Inner Mongolia, China: IEEE, Jul. 2011, pp. 1713–1715. doi: 10.1109/MACE.2011.5987287.
[16] S. W. Crown, "Improving visualization skills of engineering graphics students using simple javascript web based games," *Journal of Engineering Education*, vol. 90, no. 3, pp. 347–355, 2001, doi: 10.1002/j.2168-9830.2001.tb00613.x.
[17] N. Delson and L. V. Den Einde, "Tracking student engagement with a touchscreen app for spatial visualization training and freehand sketching," in *Proceedings of the American Society of Engineering Education (ASEE)*, Seattle, WA, 2015.
[18] K. A. Ericsson and H. A. Simon, "Effects of verbalizaton," in *Protocol Analysis*, The MIT Press, 1993. doi: 10.7551/mitpress/5657.001.0001.
[19] R. Tóth, M. Zichar, and M. Hoffmann, "Improving and measuring spatial skills with augmented reality and gamification," in *Proceedings of the 19th International Conference on*

Geometry and Graphics, Springer Science and Business Media Deutschland GmbH, 2021, pp. 755–764. doi: 10.1007/978-3-030-63403-2_68.
[20] W. Li, E. Hilton, T. Hammond, and J. Linsey, "Persketchtivity: An Intelligent Pen-Based Online Education Platform for Sketching Instruction," presented at the Electronic Visualisation and the Arts, Jul. 2016. doi: 10.14236/ewic/EVA2016.28.
[21] D. Boari, M. Fraser, D. Stanton Fraser, and K. Cater, "Augmenting spatial skills with mobile devices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Austin Texas USA: ACM, May 2012, pp. 1611–1620. doi: 10.1145/2207676.2208284.
[22] K. J. Chew *et al.*, "Work-in-Progress: Testing Content of a Spatial Skill Learning Supplementary App," in *American Society for Engineering Education (ASEE) Annual Conference & Exposition*, Montreal, Canada, 2025.
[23] J. F. Granizo, L. M. Acevedo, M. Chandramouli, K. J. Chew, and L. Sun, "Spatial Skills with Augmented Reality: The Journey of Integration," in *Proceedings of the American Society for Engineering Education (ASEE) Annual Conference*, Portland,OR, 2024.
[24] L. Sun, M. Chandramouli, and K. J. Chew, "Augmented reality: Bridging the inclusivity gap in engineering graphics education through spatial skills enhancement," in *Proceedings of the American Society for Engineering Education (ASEE) Southeastern Section Conference*, 2024.
[25] E. L. Olmsted-Hawala, E. D. Murphy, S. Hawala, and K. T. Ashenfelter, "Think-aloud protocols: A comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2010, pp. 2381–2390.
[26] S. N. Abdo, J. L. Hsu, C. Kapetanakis, D. L. Newman, L. K. Wright, and J. Bailey, "An Exploration of Spatial Visualization Skills: Investigating Students' Use of 3D Models in Science Problems during Think-Aloud Interviews," *Journal of Chemical Education*, Sep. 2024, doi: 10.1021/acs.jchemed.3c01355.
[27] O. Alhadreti and P. Mayhew, "Rethinking Thinking Aloud: A Comparison of Three Think-Aloud Protocols," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC Canada: ACM, Apr. 2018, pp. 1–12. doi: 10.1145/3173574.3173618.
[28] L. Cooke, "Assessing Concurrent Think-Aloud Protocol as a Usability Test Method: A Technical Communication Approach," *IEEE Trans. Profess. Commun.*, vol. 53, no. 3, pp. 202–215, Sep. 2010, doi: 10.1109/TPC.2010.2052859.
[29] T. Fischer, G. Reinhardt, P. Stumpf, P. E. H. Schwarz, and L. Harst, "Think-aloud usability testing of an app to increase physical activity," Dec. 13, 2019, *In Review*. doi: 10.21203/rs.2.17034/v2.
[30] M. Jaspers, T. Steen, C. Bos, and M. Geenen, "The think aloud method: a guide to user interface design," *International Journal of Medical Informatics*, vol. 73, no. 11–12, pp. 781–795, Nov. 2004, doi: 10.1016/j.ijmedinf.2004.08.003.
[31] A. Bruun and J. Stage, "An Empirical Study of the Effects of Three Think-Aloud Protocols on Identification of Usability Problems," in *Human-Computer Interaction – INTERACT 2015*, vol. 9297, J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, and M.

Winckler, Eds., in Lecture Notes in Computer Science, vol. 9297. , Cham: Springer International Publishing, 2015, pp. 159–176. doi: 10.1007/978-3-319-22668-2_14.

[32] E. Krahmer and N. Ummelen, “Thinking About Thinking Aloud: A Comparison of TwoVerbal Protocols for Usability Testing,” *IEEE Trans. Profess. Commun.*, vol. 47, no. 2, pp. 105–117, Jun. 2004, doi: 10.1109/TPC.2004.828205.

[33] T.-L. Wang and Y.-H. B. Lien, “The power of using video data,” *Qual Quant*, vol. 47, no. 5, pp. 2933–2941, Aug. 2013, doi: 10.1007/s11135-012-9717-0.

[34] M. B. Miles, A. M. Huberman, and J. Saldana, “Fundamentals of qualitative data analysis,” in *Qualitative Data Analysis*, 2014, pp. 69–103.

[35] A. McAlister, D. Lee, K. Ehlert, R. Kajfez, C. Faber, and M. Kennedy, “Qualitative Coding: An Approach to Assess Inter-Rater Reliability,” in *2017 ASEE Annual Conference & Exposition Proceedings*, Columbus, Ohio: ASEE Conferences, Jun. 2017, p. 28777. doi: 10.18260/1-2--28777.

[36] K. A. Hallgren, “Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial,” *TQMP*, vol. 8, no. 1, pp. 23–34, Feb. 2012, doi: 10.20982/tqmp.08.1.p023.