Preliminary findings for ProveIt

1. Evaluation Sessions Summary

We conducted evaluation sessions with three participants (referred to as Participant 1, 2, and 3 to maintain privacy) who formed an ideal focus group. All participants had been introduced to mathematical induction previously but had limited recent experience applying it to solve actual math problems, making them representative of our target demographic. Each session was designed to last approximately 45 minutes, and Sessions 2 and 3 met this timeframe. However, Session 1 extended beyond the planned duration due to technical setup adjustments and because this participant had weaker foundational knowledge compared to the others, requiring additional support during the evaluation.

2. Data Collection and Analysis Methods

We obtained consent to video record all sessions, with participants encouraged to think aloud and share their thoughts throughout the process. These video recordings, combined with screen recordings, provided valuable insights into participants' reactions to various features of our website. All recordings were conducted using Microsoft Teams, with OBS Studio used additionally for Participant 3. We also conducted pre-interviews to assess participants' previous knowledge and experience with mathematical induction, as well as their initial confidence levels in solving induction problems. Post-interview questions measured confidence level changes and gathered participants' impressions of the tool after using it. The post-interview phase provided our most valuable data, including detailed feedback and suggestions for system improvements.

3. Results

Tutorial Section

All participants showed positive engagement with the tool's gamification elements at first sight in the Tutorial section, particularly expressing excitement about Dr. Cube and its tips delivered through visual cloud elements, as noted by Participant 1. The Tutorial section presented minimal difficulties for participants, with the evaluation proceeding smoothly through this initial stage.

Exercise 1

Exercise 1 also presented minimal problems for participants, maintaining the smooth progression from the Tutorial section. The drag-and-drop feature was initially perceived as overly simplistic by Participant 2 at the start of Exercise 1, though this perception changed as participants progressed to more advanced exercises.

Exercise 2

Several usability issues emerged consistently across all participants in Exercise 2, where we observed participants really struggle compared to the earlier sections. The most common difficulty involved inputting Big Sigma Notation, where participants reported confusion due to the extensive array of buttons and options available in the mathematical notation interface. Another widespread issue was the requirement to press "Enter" to confirm input accuracy after entering statements, with participants suggesting that Dr. Cube should provide hints to clarify this feature. System limitations began to appear in accepting alternative correct answer formats. One participant reported decreased focus during sections where calculation processes were automatically revealed rather than user-generated, suggesting confusion when all calculations appeared simultaneously.

Final Exercise

The Final Exercise presented the most significant challenges for participants. The Big Sigma Notation input difficulties intensified, compounded by insufficient input field size for complex mathematical notation. The "Enter" confirmation requirement became more problematic as mathematical expressions grew more complex. Most critically, system limitations in accepting alternative correct answer formats became severe barriers, preventing all participants from finishing the final exercise despite correct mathematical reasoning.

A critical concern raised by Participant 1 was the difficulty gap between the exercises, which he identified as too significant, causing users to lose concentration and focus. He specifically suggested adding more practice exercises before progressing to the final exercise to better prepare users for the increased complexity.

Overall Findings

A significant strength highlighted by all participants, especially Participant 2 who had a strong mathematical background, was the tool's unique structural approach to explaining each step of mathematical induction, which they felt would be particularly beneficial for students encountering this concept for the first time. The evaluation demonstrated positive learning outcomes, with measurable improvements in confidence levels: Participant 1 increased from 0 to 1-2, Participant 2 improved from 3-4 to a solid 4, and Participant 3 showed similar improvement. Additionally, all participants indicated that the interactive tool was superior to traditional learning methods such as textbooks, YouTube videos, or standard university tutorials. Noticeable excitement reactions have been observed by all participants to the detail of the square sign at the end of the proof problem.

4. Discussion and Key Takeaways

The evaluation revealed several interesting findings that exceeded our expectations. All participants demonstrated enthusiasm for the gamification and interactive elements of our website, particularly those features embodied by Dr. Cube. A surprising discovery was the unanimous appreciation for our scaffolding methodology, where task difficulty progressively increases with each level, culminating in a final exercise requiring complete independence once users have filled the independence bar. The drag-and-drop feature evolved from being perceived as overly simplistic in Exercise 1 to being recognized as appropriately challenging in Exercise 2, demonstrating the effectiveness of our progressive difficulty design.

The evaluation pattern revealed a clear technical escalation issue, where problems that emerged in Exercise 2 became critical barriers in the Final Exercise. The system's inability to save user progress when attempting to review previous sections compounded these difficulties, resulting in unnecessary time consumption during evaluation sessions. These technical problems were primarily attributed to time and resource constraints, which can be addressed in future iterations. The identified difficulty progression gap represents a pedagogical concern that requires careful attention to maintain user engagement and learning effectiveness.

Regarding our research question, "Did the web-based tool (ProveIt) support students' understanding and ability to solve problems that require mathematical induction?" We can confidently conclude that the tool provided substantial support for learning, though significant room for improvement remains.

5. Future Recommendations

Based on our evaluation findings, we recommend several improvements to our evaluation approach and tool development. For future evaluation protocols, we suggest conducting sessions with participants encountering mathematical induction for the first time, as our study revealed that the tool may not be sufficiently engaging for those with prior knowledge, as evidenced by Participant 2's experience.

Regarding technical improvements, adding animations on Dr.Cube and its dialog box is also important to attract more user attention to Dr.Cube. Also adding information on how to interact with the input components should be added to address user’s struggles to interact with the app. Another improvement can be made on making specific Dr.Cube expressions that show availability of a hint. This is important since most participants are not sure whether Dr.Cube has a hint or not.

The pedagogical structure requires refinement to address the difficulty gap identified between exercises, potentially through the addition of intermediate practice problems as suggested by Participant 1. In the future development, a dashboard to showcase user achievement can be added to the app. Training on certain knowledge components of the proof is also another way to scaffold into more reachable chunks of skill, such that users can train the skill they need individually for final exercise.