Lessons Learned using *FullProof*, a Digital Proof Platform, in a Geometry for Teachers Course

Orly Buchbinder, University of New Hampshire
Sharon Vestal, South Dakota State University
Tuyin An, Georgia Southern University

Proof-writing is a core practice of mathematicians and a crucial skill every future teacher and mathematics major needs to acquire. A proof-based geometry course is a fruitful ground for developing proof-writing skills. Despite recent pedagogical and technological advances, proof remains challenging to learn and teach. We conducted a pilot study to explore potential advantages of an innovative digital proof platform, FullProof, which aims to advance students' proof-writing skills in Euclidean Geometry. We integrated FullProof in three Geometry for Teachers (GeT) courses at three universities. We report on the findings pertaining to students' interactions with FullProof, their perceptions of the usefulness of this tool to their proof-writing skills, and the lessons learned from the FullProof integration from an instructor perspective.

Keywords: Geometry, Proof, Prospective Secondary Teachers, Automated Proof Writing

Introduction and Theoretical Perspectives

Proof-writing is a core disciplinary practice of mathematicians, and a crucial skill all mathematics majors and future mathematics teachers need to acquire. Developing undergraduate students' facility and comfort with proofs are among the main objectives of many mathematics courses, including geometry. While geometry courses may be either required or elected for mathematics and other STEM majors, they are usually required for mathematics education majors. Hereafter we refer to such courses as GeT- Geometry for Teachers (An et al., in press; Grover & Connor, 2000). GeT courses are essential for cultivating robust and flexible knowledge and skills such as writing, analyzing, and critiquing proofs that will allow prospective secondary teachers (PSTs) to support their students' learning of geometry (AMTE, 2017; González & Herbst, 2006; NCTM, 2000; NGA & CCSSO, 2010).

Technology plays an important role in supporting PSTs' engagement with proof. Dynamic Geometry Environments (DGE) like *GeoGebra, Geometer's Sketchpad* and others have been essential in providing opportunities to explore geometrical properties, make and test conjectures (Jones, 2000; Mariotti & Baccaglini-Frank, 2018). These tools have also been beneficial for fostering PSTs' attitudes towards proof and comfort with proof and teaching it to secondary students (Abdelfatah, 2011; Kardelen & Menekse, 2017). However, DGEs offer little support for the most critical component of mathematics exploration - the *writing* of a deductive proof of a conjecture or a theorem. Recent advancements in artificial intelligence and machine learning offer new tools that support automated and interactive proof writing and verification (e.g., Lodder et al., 2021; Matsuda & VanLehn, 2005; Wang & Su, 2017), and give new opportunities for innovative research in the educational applications of these tools (Hanna et al., 2019).

FullProof is a digital platform that was developed to advance students' proof-writing skills in Euclidean Geometry. In this tool, students write a complete step-by-step proof, using an interactive diagram, an equation editor, and a library of theorems and definitions. Students can choose their own proof paths. The software checks the proof and provides feedback and/or hints to improve their work. Our project aimed to explore the potential advantages of the *FullProof* digital tool for GeT students. Our overarching research question was: *How and to what extent does engaging GeT students with the FullProof software affect their confidence with and attitudes toward geometry proofs?*

Research Setting and Methods

In Fall 2021, we conducted an exploratory study in which we integrated *FullProof* in three GeT courses in three US public universities in the Southeast, Northeast, and Midwest (See Table 1 for course and participant info). In each course, a large portion of the curriculum was devoted to Euclidean Geometry from a synthetic perspective. All instructors had experience with teaching the GeT course. To familiarize students with the platform, the instructors introduced *FullProof* during class and worked together with students to solve several proofs.

Table 1. Course and Participant Information

University Type	University A Southeast, large public	University B Northeast, large public	University C Midwest, medium public
Student population	~ 20 students: PSTs and STEM majors	~15 students: PSTs and STEM majors	~ 6-18 students: PSTs only
Course focus and modality	Content focused Hybrid	Content focused Face-to-face	Mixed content and pedagogy. Face-to-face
No. of students	14	10	6

The instructors integrated the *FullProof* platform into the courses as a support system for writing proofs, both in class and in homework. Each instructor used about 20 proof problems, from *FullProof* 's vast collection of problems on various topics from Euclidean geometry, at three levels of difficulty: high, medium, and low. Figure 1 shows a screen capture of a sample problem (medium difficulty) about triangle midsegments. Given DE is a midsegment of the triangle ABC. F is a point of BC such that BF=(1/2)FC. The segment AF intersects DE at point G and the segment BE at point O. Prove that BO≅EO.

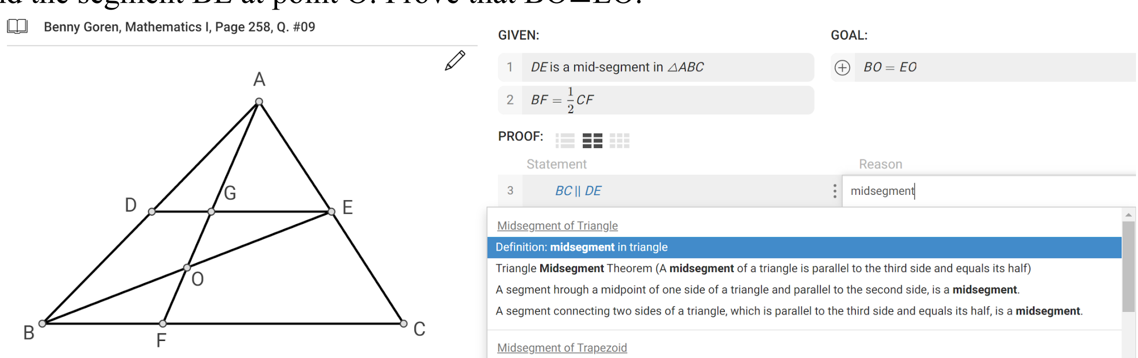


Figure 1. Sample proof problem

While the figure itself cannot be dynamically manipulated, each segment, point, and angle in the figure become highlighted when pointed to. Angles and segments can be marked, color-coded, and named as shown in Fig. 2. The right side of the screen shows the givens and the statement to be proven. The students write the proof by typing statements in the numbered lines and justifying them. The justifications can be searched by keywords, as shown in Fig. 1, or by browsing through the *FullProof* justification library. Despite the two-column format, *FullProof* allows for variation in proof strategies and the order of steps. For example, the problem in Fig. 1 can be solved with congruent triangles or properties of centroid.

The students may ask for a hint by clicking the hint button. The system will produce hints in order of increased specificity from a vague "try using triangle congruence," to suggesting a certain theorem to use, or even proposing a particular step, like "try proving ΔBFO ≅ ΔEGO". Once finished, the students click the "check" button, and in a few seconds, the algorithm checks the proof and provides feedback. Correct proof lines get a green check mark. Incorrect or partially correct lines are marked down with an explanation of the mistake. Fig. 2 shows a solution with two mistakes – a missing proof step and a missing justification for the last step. Clicking on these notifications will show a student what the missing step was (here ΔBFO ≅ ΔEGO by angle-side-angle theorem).

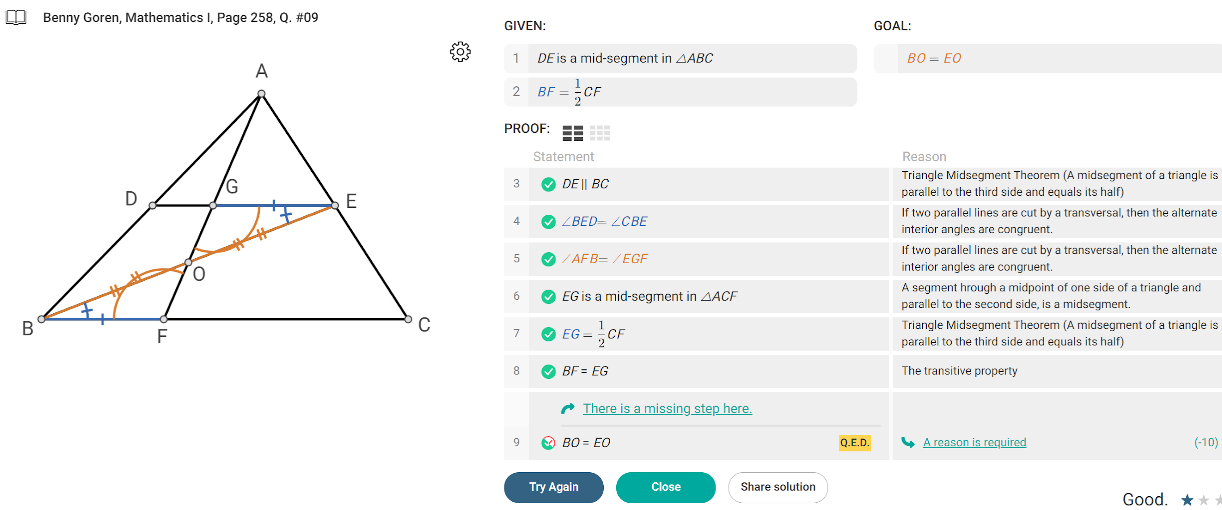


Figure 2: Sample solution in FullProof and the feedback provided by the software.

Clicking the "try again" button allows a student to improve their work, submit a corrected solution and receive more feedback. The number of hints and submissions is unlimited, allowing the students to achieve a perfect score eventually. The information about the number of hints and submission attempts is stored in the system. The instructor's interface contains information about all students in the class – which of the assigned problems they solved, how many hints and attempts they used, and can see the last submitted solution (Fig. 3).

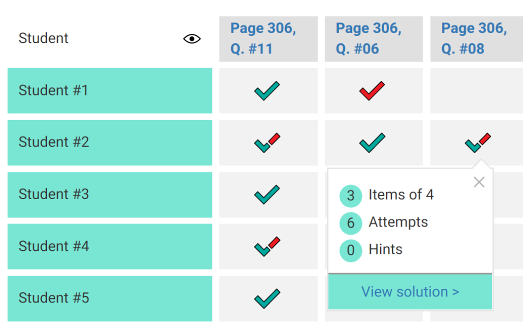


Figure 3: Excerpt of instructor interface in FullProof.

Data Collection and Analysis

To respond to our research question, we administered pre- and post-survey via Qualtrics. We focused on the results of two sets of Likert-type questions about attitudes toward proof, and comfort level with writing proofs (see Fig. 4 for sample items). The items were adapted from the relevant instruments in the literature (e.g., Kaspersen & Ytterhaug, 2020; Stylianou et al., 2015). The Likert-scale for proof-related attitudes was: *Never/Almost never (1), Sometimes (2), Often (3), Always/almost always (4)*; and for comfort-level was: *Strongly Disagree (1), Disagree (2), Agree (3), Strongly Agree (4).*

Attitudes towards Proof questions (20 items). Examples:

1. I take the initiative to learn more about proof writing than what is required in this course.
3. **When I learn a new proof, I try to think of situations when it wouldn't work.**
13. **When I face a proof problem, I consider different possible ways I can prove it.**
18. **I can explain why my solutions are correct.**
20. If I immediately do not understand what to do, I keep trying.

Comfort with proof (10 items). Examples:

1. I feel comfortable when I prove mathematical theorems in the *FullProof* platform.
2. **I feel confident in my ability to prove mathematical results using the *FullProof* platform.**
3. **I am familiar with different proof techniques using the *FullProof* platform.**
4. **I am confident about my ability to teach proof using the *FullProof* platform.**
6. I believe I can learn to read and write proofs if I put enough effort into it.

Figure 4. Selected Likert Items from the FullProof *Survey.*

Twenty-nine students responded to the pre-survey while only 20 responded to the post-survey. To analyze Likert-scale data, we initially used the chi-squared test of independence, which is appropriate for categorical data, but since some of the categories had less than five data points, the chi-square test is unreliable. To address this challenge and to avoid discarding unpaired observations by using a paired sample *t*-test, we used a new statistical test–developed by Ben Derrick, University of the West of England, which allows analyzing observations taken at two points in time, where the population membership changes over time but retains some common members (Derrick, 2020).

The post-survey also contained seven open-ended questions, developed by the researchers, about student perceptions of the *FullProof* platform (Fig. 5). We analyzed the data qualitatively, using open coding and thematic analysis (Patton, 2002) to reveal recurring themes in students' responses, specifically the types of positive and negative appraisals of *FullProof*. Each appraisal is defined as a statement that has a relatively complete and independent meaning. To capture all themes, comments that included multiple ideas were coded multiple times.

1. How much experience did you have writing proofs before using this tool?
2. **Has this tool changed the way you write proofs? If yes, explain in what way.**
3. **What are some of the successes/challenges you have had in proof writing when using this tool?**
4. **How has this tool changed your understanding of reasoning and proof?**
5. **Would you use this tool in your future geometry classroom? Why? How? If not, explain why not.**
6. What other features would you like to see in this tool?
7. **What is your overall impression of the FullProof tool?**

Figure 5. Open-ended Items from the FullProof *Survey.*

In addition to the survey, we analyzed student solutions to a few problems assigned in all three GeT courses to examine the types of proof approaches utilized by the students, and the types of mistakes they made. Due to space constraints, we do not report on this analysis here.

Results

Survey Results

Attitudes toward proof. Fig. 6 shows the statistical test results for three of the survey questions related to proof attitudes. These questions were selected since their Derrick test *p*-values were relatively close to 0.05, indicating that the results are statistically significant.

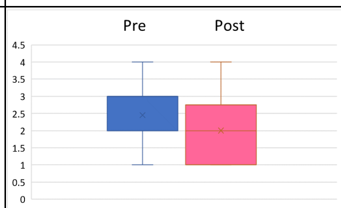
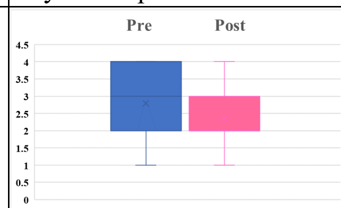
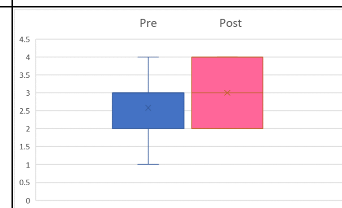
Question	When I learn a new proof, I try to think of situations when it wouldn't work.	When I face a proof problem, I consider different possible ways I can prove it.	I can explain why my solutions are correct.
Box plot			
Derrick test	0.0431	0.0713	0.0438

Figure 6. Sample Survey Results about Mathematical Identity.

While the pre-post changes for the questions in Fig. 6 are significant, the results are not necessarily in the intended direction. For example, the boxplots for the first two questions show a decrease in participants' tendency to think about situations in which a proof wouldn't work or consider different proof paths. With respect to the first question, the results may be reflective of the nature of geometry proof problems, like the one in Fig. 1, as opposed to proofs of theorems; and the static nature of the FullProof figures, as opposed to DGE. These features of *FullProof* are indeed less likely to invoke one's thinking about situations where the proof is inapplicable. The decrease in considering various solution methods was surprising to us since *FullProof* allows for much flexibility in proof approaches that instructors always encouraged. However, the hints from the software may lead students to a certain approach. One explanation for this result is that students' responses at the end of the semester reflect a pragmatic preference for one correct solution that guarantees a good grade, rather than exploring multiple solutions. It is also possible that students' confidence at the beginning of the semester decreased due to the course being harder than they anticipated. The results for the third question are encouraging, with students' increased confidence in their ability to explain their correct solutions at the end of the semester.

Comfort-level with proof. Fig. 7 shows results related to using the *FullProof* platform.

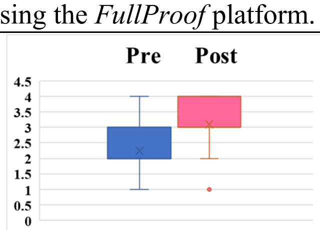
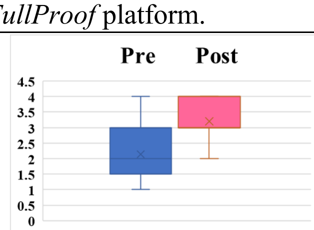
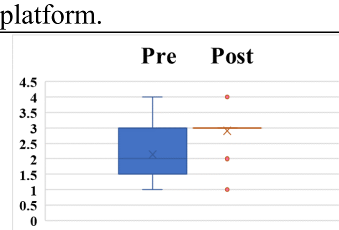
Question	I feel confident in my ability to prove mathematical results using the *FullProof* platform.	I am familiar with different proof techniques using the *FullProof* platform.	I am confident in my ability to teach proof using the *FullProof* platform.
Box plot			
Derrick test:	0.0009	0.000136	0.003014

Figure 7. Sample Survey Results about Comfort Level with Proof.

These questions had the lowest *p*-values, indicating significant pre-to-post changes, which is not surprising since the students had never used *FullProof* prior to the course. Still, we are encouraged by the positive direction and amount of change. For example, in the first two questions in Fig. 7, 75% of the post-survey responses were higher than the pre-survey responses. These results indicate significant increases in students' familiarity with various proof techniques and increased confidence in their ability to write proofs and even teach proofs with the *FullProof* platform. The increased familiarity with different proof techniques in *FullProof* (question 2, Fig. 7) suggests that students' lack of consideration of multiple proving approaches (question 2, Fig. 6) occurred despite increased knowledge of these techniques. This may support our assumption about pragmatic reasons for not considering multiple ways of proving. The analysis of open-ended questions sheds additional light on these results.

Student Appraisals of the *FullProof* Platform

Nineteen students responded to the six open questions about *FullProof*. There was a total of 84 appraisals, 61 (73%) positive and 23 (27%) negative. "Appraisal" is a thematic unit, such that a single written comment may contain both positive and negative appraisals, or more than one appraisal of the same type. Six out of 19 students provided positive appraisals only, others shared a combination of both. One interesting pattern in the data was: when asked to describe some of the successes and challenges they had with *FullProof* (#3, Fig. 5), 65% of comments described challenges as opposed to 35% of successes. However, the overall impressions of *FullProof* (# 7, Fig. 5) were very positive with 73% of comments containing positive reviews.

Three main themes emerged among the positive appraisals. The largest category (41%) described the *affordances of FullProof for supporting students' writing and understanding of proofs*. These included the searchability of reasons for proof steps, clear structure that supports communication and comprehension, interactive feedback, and the ability to pursue multiple solution paths. One student wrote, "*FullProof* has helped me a lot when writing proofs. I like how I can search for reasons if I am not completely sure about a reason/theorem." Another student wrote: "*FullProof* made writing proofs easier not only to write, but also to understand."

The second theme (26%) described the *advantageous technical features of FullProof,* such as hints which help them move forward if one is stuck, interactive feedback pointing to errors, visual clarity, and color-coded elements of a diagram. The third theme (33%) described the affordances of *FullProof* as a pedagogical tool for teaching others. In this theme, the participants highlighted the elements of the software that would support them as teachers in a geometry classroom. Students wrote that *FullProof* presents "a good instructional strategy to implement in the classroom," and the platform "Makes it easier for the students to see what their errors were."

The 23 negative appraisals were distributed rather uniformly across five themes. The main critique (7 out of 23 appraisals, 30%) concerned the discrepancy between the wording of the theorems in the *FullProof* platform and in GeT course. For example, *FullProof* does not have the angle-angle-side triangle congruence theorem, so students need to complete an extra proof step to use the angle-side-angle congruence theorem. Another category of critiques (22%) described the variation in standards of rigor employed by *FullProof* vs. GeT instructors. The students wrote that the software allowed them to skip steps and earn full points for their solution, while the students were aware that their instructor would have likely deducted points. One student wrote: "It allows some things to slide, that should be wrong." Other types of critiques concerned some minor technical issues like the lack of an undo button (17%). Three students admitted having personal struggles with technology, in general (13%). Finally, some students wished to break out from the two-column structure altogether to write paragraph-style proofs (17%).

Additional patterns emerged from the analysis of specific questions. When asked “Has *FullProof* changed the way you write proofs?” the responses split almost evenly between *yes* and *no*. But 83% of the accompanying written comments for this question were positive regarding *FullProof*. A similar pattern was observed in a question: “How has *FullProof* changed your understanding of reasoning and proof?” Sixty-one percent of students wrote that *FullProof* positively influenced their understanding, but interestingly, 91% of their comments contained some specific description of the positive effect of *FullProof*. In response to the question “Would you use *FullProof* in your future classrooms,” 14 out of 17 students who answered this question responded positively, citing the advantages of the platform.

Discussion

Overall, the quantitative analysis shows little impact of the use of *FullProof* on student attitude towards proof and comfort level with proof writing, since pre-post changes on most survey questions were not statistically significant at the 0.05 level (as indicated by large *p*-values on Derrick’s test). Yet, there were some questions showing significant changes using the Derrick test. While there is still some risk of a Type I error when using the partially overlapping sample *t*-test, Derrick’s research indicates that this is “the recommended test for analyzing data recorded on an ordinal scale when partially overlapping samples are present and is particularly useful when sample sizes are small” (Derrick, p. 147).

We are also aware of the limitation of our quasi-experimental research design. Due to the lack of randomization and a single experimental group in the research setting, potential confounding variables were not controlled (Johnson & Christensen, 2012). Each course was taught by a different instructor, with no common curriculum or textbook across the institutions. There was also some variation in how instructors used *FullProof* in their courses. Finally, because of the Covid-19 pandemic, one institution required the instructor to accommodate hybrid teaching modality, which affected course pedagogy, including the use of the *FullProof* platform.

Because of these variations, our study design emulates natural conditions of how instructors might use any technological tool within the unique constraints of their institutional environments. Against this backdrop, we find it encouraging that the qualitative analysis showed more positive appraisals than negative appraisals about using the platform, which indicates the overall positive impact of using *FullProof* in GeT courses. The significance of this result is due to our analyses not distinguishing between students of different majors - mathematics education or not - meaning that potential advantages of *FullProof* apply to all GeT students.

As instructors, we note several advantages to using *FullProof* in our GeT courses. Due to the availability of multiple proof problems at various difficulty levels, and the feedback provided by the *FullProof* platform, we were able to assign a greater number of problems and more challenging problems, compared to the past. Since *FullProof* offers hints and multiple submissions, students were able to get assistance from the software itself. However, this also presented a challenge, when some students abused this functionality to get hints on every step of the proof. In the future, we plan to limit the number of hints and submission attempts. Another limitation of *FullProof,* also noted by the students, is that there are some discrepancies in the standards of rigor with the GeT courses. However, this can be leveraged to engage students in analyzing and critiquing proofs, which is an important learning objective of GeT courses (An, et al., in press). Based on the results of our exploratory study, we revised our research instruments and protocols, to use in the next round of the study. Future studies should expand the scope of research questions to include an examination of cognitive and meta-cognitive processes afforded by *FullProof*.

References

Abdelfatah, H. (2011). A story-based dynamic geometry approach to improve attitudes toward geometry and geometric proof. *ZDM-Mathematics Education* 43, 441–450. doi.org/10.1007/s11858-011-0341-6

Association of Mathematics Teacher Educators. (2017). *Standards for Preparing Teachers of Mathematics*. https://amte.net/standards

An, T., Boyce, S., Brown, A., Buchbinder, O., Cohen, S., Dumitrascu, D., Escuadro, H., Herbst, P., Ion, M., Krupa, E., Miller, N., Pyzdrowski, L., J., Sears, R., St. Goar, J., Szydlik, S., Vestal, S. (in press). (Toward) Essential student learning objectives for teaching geometry to secondary pre-service teachers. *AMTE Professional Book Series, Volume 5: Reflection on Past, Present and Future: Paving the Way for the Future of Mathematics Teacher Education.*

Derrick, B. (2020). Statistical inference with paired observations and independent observations in two samples. https://www.researchgate.net/publication/354031181

González, G. and Herbst, P. (2006). Competing arguments for the geometry course: Why were American high school students supposed to study geometry in the twentieth century? *International Journal for the History of Mathematics Education, 1*(1), 7-33.

Grover, B., & Connor, J. (2000). Characteristics of the College Geometry Course. *Journal of Mathematics Teacher Education*, *3*, 47–67.

Hanna, G., Reid, D. A., & De Villiers, M. (Eds.). (2019). *Proof Technology in Mathematics Research and Teaching*. Springer International Publishing.

Johnson, B., & Christensen, L. (2012). *Educational research: Quantitative, qualitative, and mixed approaches* (4th ed.). SAGE Publications, Inc.

Jones, K. (2000). Providing a foundation for deductive reasoning: Students' interpretations when using dynamic geometry software and their evolving mathematical explanations. *Educational Studies in Mathematics, 44*(1), 55–85.

Kardelen, G., & Menekse, S. T. B. (2017). Views of pre-service teachers following teaching experience on use of dynamic geometry software. *Educational Research and Reviews*, *12*(24), 1208-1219.

Kaspersen, E., & Ytterhaug, B. O. (2020). Measuring mathematical identity in lower secondary school. *International Journal of Educational Research, 103*, 101620. https://doi.org/10.1016/j.ijer.2020.101620

Lodder, J., Heeren, B., Jeuring, J., & Neijenhuis, W. (2021). Generation and Use of Hints and Feedback in a Hilbert-Style Axiomatic Proof Tutor. *International Journal of Artificial Intelligence in Education*, *31*(1), 99-133.

Mariotti, M. A., & Baccaglini-Frank, A. (2018). Developing the Mathematical Eye Through Problem-Solving in a Dynamic Geometry Environment. In *Broadening the Scope of Research on Mathematical Problem Solving* (pp. 153-176). Springer.

Matsuda, N., & VanLehn, K. (2005). Advanced Geometry Tutor: An Intelligent Tutoring System for Proof-Writing with Construction. *Proceedings of the Japan National Conference on Information and Systems in Education*, (Vol. 125, pp. 443-450).

National Council of Teachers of Mathematics [NCTM]. (2000). *Principles and standards for school mathematics*. National Council of Teachers of Mathematics.

National Governors Association Center for Best Practices [NGA], & Council of Chief State School Officers [CCSSO]. (2010). *Common Core State Standards for Mathematics*. http://www.corestandards.org/Math/

Patton, M. Q. (2002). *Qualitative research & evaluation methods*. Sage.

Stylianou, D. A., Blanton, M. L., & Rotou, O. (2015). Undergraduate students' understanding of proof: Relationships between proof conceptions, beliefs, and classroom experiences with learning proof. *International Journal of Research in Undergraduate Mathematics Education, 1*(1), 91–134.

Wang, K., & Su, Z. (2017). Interactive, intelligent tutoring for auxiliary constructions in geometry proofs. *arXiv:1711.07154v1*