

E-assessment of understanding of geometric proofs using interactive diagrams

Yael Luz and Michal Yerushalmy

University of Haifa, Faculty of Education, Haifa, Israel, yael.luz.halevi@gmail.com

The present study is part of a larger one examining the design principles of e-assessment of understanding of geometric proofs. The interactive assessment tasks are checked automatically and feedback is generated. We reviewed various proving task design studies, looking for a template that incorporates interactive sketching and that can be checked automatically. Our findings suggest that examples sketched in a dynamic geometry environment in order to complete the assessment task are a challenging target for e-assessment and are relevant for eliciting the students' understanding of geometric terms, geometric statements, and their validity.

Keywords: Assessment, DGE, examples, validity.

BACKGROUND


The present paper focuses on the challenge of automated assessment of knowledge of geometric proofs. Specifically, we propose to assess processes of conjecturing and argumentation. To this end, we study the roles of examples (designed to be checked automatically) demonstrating the logical validity of geometric statements.

Assessing understanding of geometric proofs

For decades an extensive effort has been made to automate the process of checking proofs. But a deductive proof is a complex process, and most of the proofs produced by students are subject to their teachers' evaluation. We are interested in a facet that is not often checked in a school setting (neither by humans nor by machines), although it is an inherent part of mathematicians' work on proofs: conjectures and argumentation. In the process of conjecturing, mathematicians try to generate different examples, including extreme and boundary examples, in order to create a rich example space from which a pattern can emerge. When mathematicians investigate the

validity of a new conjecture, they usually do not only look for the proof but try to construct counter-examples by means of quasi-empirical testing, because such testing can expose hidden contradictions, errors, or unstated assumptions (De Villiers, 1990). Proving often involves understanding how the proof relates to specific examples and how these examples can illustrate it. Being able to follow a sequence of inferences in a proof based on a specific example has been considered by mathematicians to be an indispensable tool for understanding a proof. Mejia-Ramos and colleagues (2012) constructed a model for assessing the comprehension of aspects of a proof. They discovered four main facets: summarizing the main idea of the proof, understanding the components or modules of the proof, applying the method of the proof in other contexts, and illustrating the proof with examples or diagrams. Kuzniak (2013) presented a new framework for the didactics of geometry work. The framework encapsulates instrumental processes. It transforms artefacts, such as dynamic geometry software, into tools in the construction process and into a discursive process of the proof that confers meaning on the properties used within the mathematical reasoning. The assessment template we designed follows Ramos's (2012) assessment model and is inspired by the framework suggested by Kuzniak (2013).

Buchbinder and Zaslavsky (2009) offered a mathematical framework for designing tasks that question the students' understanding of the role of examples in determining the validity of mathematical statements. This framework provides a basis for constructing tasks that assess and support students' understanding of the logical connections between examples and statements. Any mathematical statement can be reduced to two sets of mathematical objects: the "if" part of the statement, which is the domain, the set of all mathematical objects to which the statement refers (e.g., isosceles triangles), and the "then" part of the

statement, the proposition that defines the set of all mathematical objects that exhibit a certain property (e.g., right angle triangles). In this context, the status of the example object can be defined as follows: a supporting example is an object in domain D that exhibits P (e.g., an isosceles triangle that is a right-angle triangle \triangle). A counter-example is an object in domain D that does not exhibit P (e.g., an isosceles triangle that is not a right-angle triangle .

Interactive diagrams

Herbst and Arbor (2004) found that building reasoned conjectures or using deductive reasoning to find out what could or should be true can be supported by tasks that engage students in generative interactions with diagrams. This study uses interactive diagrams as a tool to create examples that either support or refute conjectures. The interactive diagrams we use are designed to describe the domain and context of the task (an example of a geometric figure or drawing) and to support autonomous guided inquiry, by enabling direct manipulations of the example and by providing feedback that reflects the process of inquiry (Yerushalmy, 2005). Students interact with the diagram while exploring the conditions under which a version of the problem can be solved (Herbst & Arbor, 2004). By dragging a dynamic diagram students generate an example that demonstrates specific properties. The use of dragging allows students to experience kinematic dependence that can be interpreted as logical dependence within the dynamic environment, but also within the geometric context (Mariotti, 2006). According to Mariotti, the dragging interaction can be compared with formal arguments used in mathematical proofs (Mariotti, 2006).

e-Assessment

Whereas traditional assessment focused primarily on testing factual knowledge, new technologies gave rise to the need to assess new skills, such as problem solving, creativity, critical thinking, and risk-taking. To nurture and develop these skills, the assessment strategy should aim beyond testing factual knowledge and capture the less tangible themes that underlie these skills. Although technology offers a rich learning experience, studies show limitations when it comes to assessing solutions for complex mathematical problems. First-generation e-assessment was limited to multiple choice questions, subsequently enhanced by short verbal or numeric answers (Scalise & Gifford, 2006). Studies show that assessments based

exclusively on questions of this sort lead to limited learning and incorrect inferences of purpose of the assessment, such as “there is only one right answer,” “the right answer resides in the head of the teacher or test maker,” and “the role of the student is to get the answer by guessing” (Bennett, 1993). To assess the complex processes involved in proofs we must understand the role that examples play in proofs and the links between these examples (in our case, dynamic figures) and logical argumentation.

THE STUDY

Objectives

The present paper presents results derived from a broader research study for which we developed a tool to assess students' skills in geometry proofs. The tool provides immediate multiple representation feedback and analysis at the level of the single student, group, or class, based on the teacher's choice. The challenge of the innovative development is to study strategies followed by students in investigating the validity of a geometric statement and in generating examples. We proposed to determine which aspects of the comprehension of a geometry proof are demonstrated by the automatically generated evaluation presented either as a personal solution or as a visually comparative collection of answers, and which aspects are demonstrated in the accumulated visual feedback.

Study setup

Sixty three middle school students participated in the study; 33 were in the 8th and 30 in 9th grade. The geometry classrooms featured mixed abilities, including gifted students and students with special needs. We use the terms “high-, medium-, and low-skill students” for students who showed high, medium, and low achievement in the subject of the examination before the experiment.

The experiment included a practice session, an examination session, and a whole-class discussion period. Before the examination, students had a practice session aimed at presenting the instructions and the technological interface (personal tablets) using a similar template task. The examination included three problem-solving tasks with a repeated five-item template [1]. In this paper we report the results of the work on two items of the template appearing in three tasks. We also report on the class discussion that took place a few days after the examination and on task-

In the diagram ABC is an isosceles triangle ($AB=AC$)
 DA, DB and DC are the bisectors of angles A, B and C correspondently

1 Gil conjectures that the bisectors always divide the triangle into three congruent triangles
 Generate a counter example to Gil's conjecture and click
 or click if there is no counter example

2 Shai conjectures that one of the triangles created by the bisectors can be an acute triangle
 Generate a supporting example to Shai's conjecture and click
 or click if there is no supporting example

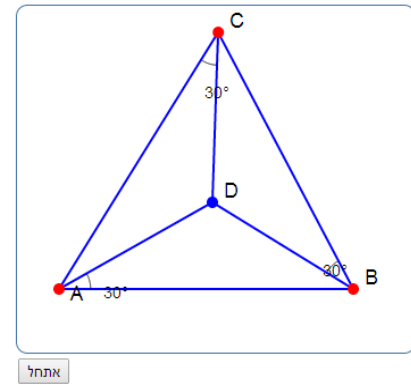


Figure 1: Page setup of paired items B1 and B2 (points A, B, C are draggable)

based interviews with three pairs of students (two high-skilled and one low-skilled) solving the same examination items a few weeks after the exam. The group feedback used in the whole-class discussion and the task-based interviews were conducted for the purpose of triangulation.

Design of the research tools

Task: The examination included three tasks that are instances of the same five-item template. Three items focused on the comprehension of terms and the reading of the proof. Here we discuss the other two items, which focused on assessing students' understanding of the validity of a geometric statement and generating supporting or counter-examples. The students

were asked to provide a counter-example for the universal statement and a supporting example for the existential statement (Figure 1). The students could also choose "none," indicating that such an example does not exist. The domain of both items was identical in each task. The interactive diagram was a figure, therefore robustly in the class of objects described by the domain, and the properties defining the domain were invariant under dragging. The initial orientation of the figure does not represent the correct answer, and students must actively drag the diagram to generate the correct answer. The items of the three tasks are described in Table 1.

Objectives	Domain (if)	Proposition (then)	Diagram
A1. Counter-example	$AB = AC$ $BD = CE$	ΔEFD cannot be a square	
A2. Supporting example		ΔFDB can be an isosceles triangle	
B1. Counter-example	$AB = AC$ DA, DB and DC are bisectors	The bisectors always divide the triangle into three congruent triangles	
B2. Supporting example		One of the triangles created by the bisectors can be an acute triangle	
C1. Counter-example	$AB \perp CD$ $\angle A = \angle D$	$\Delta AMC \sim \Delta DMB$	
C2. Supporting example		$\Delta AMC \cong \Delta DMB$	

Table 1: The logical structure of the three pairs items

In each diagram, some measures of lengths and angles were displayed. Other measures that were required for constructing the example were such that could be derived from the displayed measures and the properties of the domain (e.g., the measure of only one of the equal sides of an isosceles triangle is displayed). The use of measurements in interactive diagrams confronts students with the issue of the relationship between measurement and proof (Chazan, 1988).

Rubric: A rubric was developed for each task based on a literature review of relevant misconceptions and on pilot trials of the tasks conducted with students (mostly pre-service mathematics teachers).

Feedback: We designed two feedback sheets. A personal feedback sheet displayed the student's examination answers and personal evaluation feedback [2]. The feedback was not provided to the students immediately after completion of the examination in order to support an unbiased class discussion. The group feedback sheet (Figure 6) was designed to display the collective answers and results of all students for a single item.

Data sources and analysis

Data analysis included the personal and group feedback sheets, the class discussion videotape, and videotapes and transcripts of the interviews. We verified the automatic check by manually checking the results. Next, we used the group feedback to look for patterns of mistakes. Expected pre-configured misconceptions were captured by the automatic check, and other mistakes were identified manually. We conjectured about the source of the mistake and later verified our conjectures during the interviews.

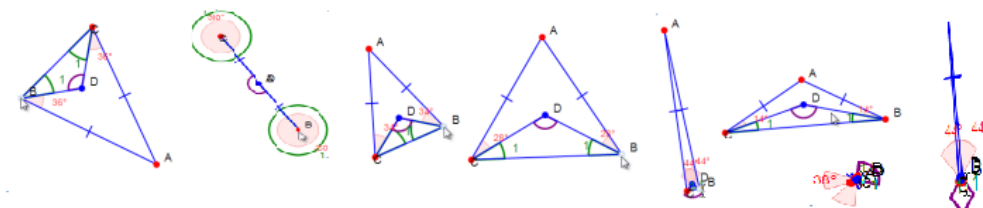


Figure 2: Example space of item B2 created by the high-skill pair

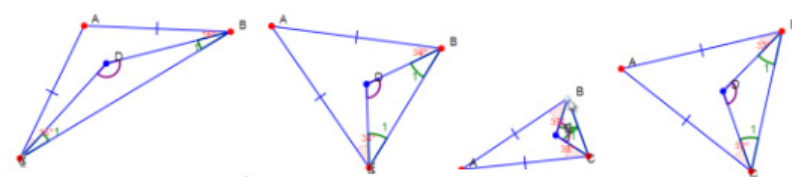


Figure 3: Example space of item B2 created by the low-skill pair

FINDINGS AND RESULTS

The findings are classified into three categories: strategies, misinterpretation of terms, and understanding statement validity according to the three questions formulated in the objectives of the study. Each category includes the triangulation of the results obtained by analysis of the three research tools.

Strategies used to create examples

During the examination we found that highly skilled students demonstrated active and fearless engagement with the interactive diagrams, dragging points to the extreme and collapsing them to a single point.

This impression was verified during the interviews, when the two high-skill pairs intensively dragged the diagram from an initial state to extremes, constantly expanding the example space (Figure 2), whereas the low-skill pair showed hesitating and careful dragging, making small changes to the initial diagram (Figure 3).

Some students (mostly high-skill) submitted extreme examples even when standard examples were sufficient to provide a counter-example, as shown in Figure 4.

Another strategy was the use of symmetrical diagrams, with as regular as possible shapes. This strategy helped students to avoid common error such as matching non-corresponding sides, trying to create congruent triangles (see Figure 5) or examining isosceles triangle as a possible counter example to the statement "the median divides the triangle into two equal area triangles".

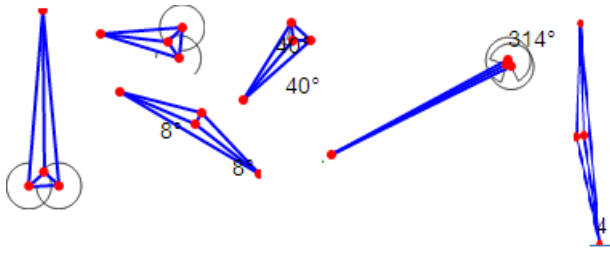


Figure 4: Counter-examples for item B1

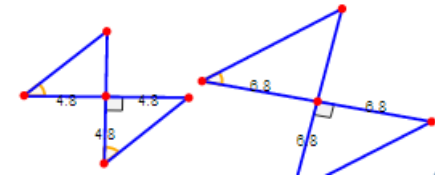


Figure 5: Examples of common strategy in item C2 (using isosceles triangles)

Analyzing terms in the given statement

The study uses the group feedback sheet to demonstrate the affordance of the assessment tool to inform the teacher in regard to students' interpretation of the terms in the given geometric statement. The examination feedback provided a view of the

interpretation of terms and concepts in class. In the feedback sheet for the task “if $AB \perp CD$ and $\angle A = \angle D$, and M is the intersection point of AB and CD , then $\triangle ACM \cong \triangle DBM$ ” (Figure 6), we were able to identify misinterpretations of the concept of “corresponding parts in congruent triangles.” Although most students generated triangles with equal angles and one pair of equal sides, a substantial number of them (marked in orange) matched non-corresponding sides ($AM = CM$, instead of $AM = DM$).

Another common misinterpretation was found on the group feedback sheet of another task (B2), where students were asked to provide a supporting example for the incorrect existential statement: “the bisectors of an isosceles triangle create an acute triangle” (Figure 7). Although the example on the left is an extreme one and the closest to an acute triangle, the bisectors in the other two diagrams create clearly obtuse triangles. We conjectured that students misinterpreted the term “acute triangle.” This assumption was verified in the class discussion following the ex-

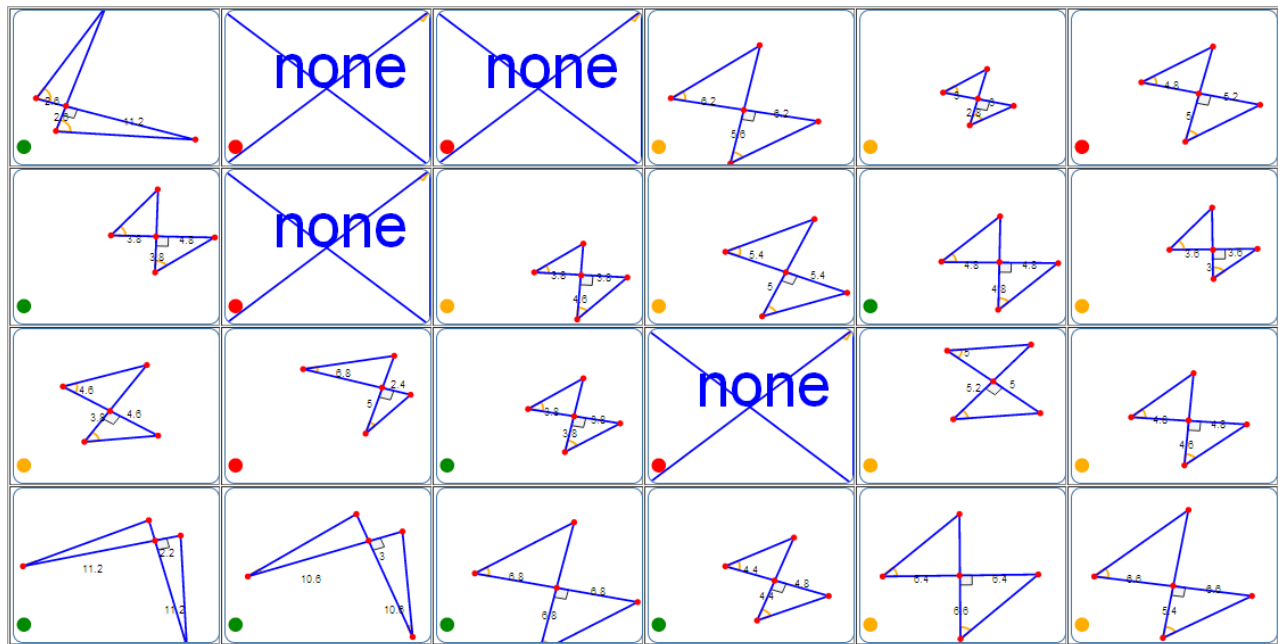


Figure 6: Group feedback for item C2

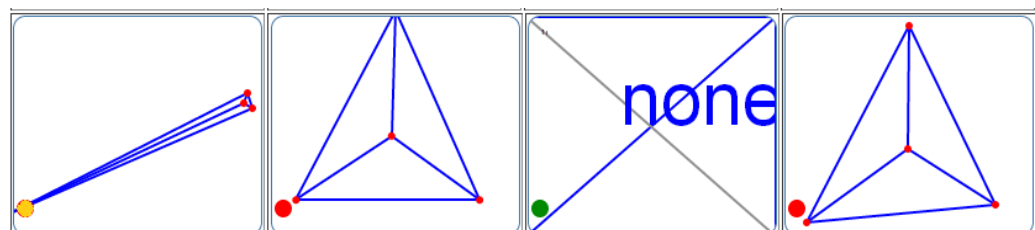


Figure 7: Partial feedback sheet for item B2 supporting example

	Counter-example		Supporting example	
	Validity	% Correct	Validity	% Correct
A	Invalid	84%	Valid	83%
B	Valid	79%	Invalid	59%
C	Valid	68%	Invalid	76%

Table 2: Correct responses in regard to validity of statements

amination with students who defined an acute triangle as one that includes a single acute angle.

Determining the validity of the statement

In general, students were able to determine the validity of the statement, as can be seen in Table 2. The lower scores were in task B2, where the term “acute triangle” was misinterpreted by many.

Although most students correctly identified the validity of the statement, they did not necessarily provide a correct example figure (e.g. correspondence in congruent triangles). In a class discussion following the examination, students were asked to justify the “none” answers presented in the aggregated feedback sheet. All students based their justifications on empirical evidence, which means that they were not able to find any examples. This phenomenon was confirmed in the interviews. But during the interviews the high-skill pairs were able to produce a deductive explanation relatively quickly.

DISCUSSION

Based on the partial findings presented above we can draw several conclusions with regard to the objectives of the study: (a) we were able to learn what strategies students used, and take pedagogical action, such as encouraging low-skill students to use more active and fearless dragging in order to investigate extreme examples; (b) we were able to use the group feedback sheet as a tool to identify misinterpretations of terms; (c) the design of the items and the automatic checking provided immediate quantitative and personal data about the validity of the understanding of the geometric statement, but not about the presence of a justification when examples were not available; students simply based their answers on empirical evidence.

The visual feedback sheet helped us quickly identify central patterns of knowledge in class. In a discussion that took place in class after the examination, we displayed the group feedback sheet and found that it

encouraged active discussion, as students were eager to participate in a conversation based on their generated examples. Successful brain-storming, involving many members of the class, was conducted about the existence of multiple correct answers, the importance of a large-scale example space and the use of active and fearless dragging strategy for expanding it, the role of extreme examples, and the need for students to justify their “no such example” answer. Future research on the role, use, and format of personal and group visual feedback should be a fruitful field for studying the use of the ample and immediate data produced by e-assessment.

ACKNOWLEDGEMENT

This research was supported by the Israel Science Foundation (grant 522/13)

REFERENCES

- Bennett, R. E. (1993). On the meaning of constructed response. In R. . Bennett & W. C. Ward (Eds.), *Construction versus Choice in Cognitive Measurement: Issues in Constructed Response, Performance Testing, and Portfolio Assessment* (pp. 1–27). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Buchbinder, O., & Zaslavsky, O. (2009). A framework for understand the status of examples in establishing the validity of mathematical statements. In M. Tzekaki, M. Kaldrimidou, & H. Sakonidis (Eds.), *Proceedings of the 33rd Conference of the International International Group for the Psychology of Math* (Vol. 1, pp. 225–232). Thessaloniki, Greece: PME.
- Chazan, D. (1988). Proof and measurements: an unexpected misconception. In *Proceedings of the 12th PME Conference* (pp. 207–214). Veszprem, Hungary.
- De Villiers, M. (1990). The role and function of proof in Mathematics. *Pythagoras*, 24, 17–23.
- Herbst, P., & Arbor, A. (2004). Interactions with diagrams and the making of reasoned conjectures in geometry. *ZDM*, 36(5), 129–139.

- Kuzniak, A. (2013). Teaching and learning geometry and beyond... In B. Ubuz, Ç. Haser, & M. A. Mariotti (Eds.), *Proceedings of CERME 8* (pp. 33–49). Antalya, Turkey.
- Mariotti, M. A. (2006). Proof and proving in mathematics education. In *Handbook of research on the psychology of mathematics education: Past, present and future* (pp. 173–204). The Netherlands: Sense Publishers.
- Mejia-Ramos, J. P., Fuller, E., Weber, K., Rhoads, K., & Samkoff, A. (2012). An assessment model for proof comprehension in undergraduate mathematics. *Educational Studies in Mathematics*, 79(1), 3–18.
- Scalise, K., & Gifford, B. (2006). A Framework for Constructing “Intermediate Constraint” Questions and Tasks for Technology Platforms Computer-Based Assessment in E-Learning : *The Journal of Technology, Learning, and Assessment*, 4(6), 45.
- Yerushalmy, M. (2005). Functions of interactive visual representations in interactive mathematical textbooks. *International Journal of Computers for Mathematical Learning*, 10(3), 217–249.

ENDNOTES

1. Third test task can be viewed at: <http://geo.gigaclass.com/tasks/en/task2bEn.html>
2. Third task feedback is available at: <http://geo.gigaclass.com/tasks/en/fb2b2En.html>