Online assessment of students' reasoning when solving example-eliciting tasks: using conjunction and disjunction to increase the power of examples

Michal Yerushalmy[1] · Shai Olsher[1]

Abstract
We argue that examples can do more than serve the purpose of illustrating the truth of an existential statement or disconfirming the truth of a universal statement. Our argument is relevant to the use of technology in classroom assessment. A central challenge of computer-assisted assessment is to develop ways of collecting rich and complex data that can nevertheless be analyzed automatically. We report here on a study concerning a dedicated design pattern of a special kind of task for assessing student's reasoning when establishing the validity of geometry statements that go beyond a single case, concerning the similarity of triangles. In each task students are given three relations that exist either in every triangle or in special types. Their task is to verify, by creating an example, claims that argue for logically compounded claims built out of the given relations. 50 students, aged 15–16, were asked to verify or disprove claims. Their submissions were automatically characterized along categories based on the correctness of the claims they chose and the examples they used to support the claims. We focus on characterizing the properties of the conjunction/disjunction design for automatically assessing conceptions related to examples generated by the learner with interactive diagrams. Our analysis shows that our automated scoring environment, which supports interactive example-eliciting-tasks, and the design principles of conjunction and disjunction of geometric relations, enable one to assess students' exploration of the logic of universal claims, characterize successful and partial answers, and differentiate between students according to their work.

Keywords Examples of universal statements · Example eliciting tasks · Automatic online assessment

1 Introduction and rationale: examples in conceptual assessment

Interactive environments in mathematics offer opportunities for argumentation and substantiation. We have studied how existing assessment items, known to be useful for informing about students' mathematical skills, can be redesigned for computer-based assessment. Cycles of research-based design have led to the development of an innovative assessment research platform. The principles of the automatic assessment that we articulated and tested provide a basis for online analysis of open-ended tasks that are automatically assessed (Olsher et al. 2016). A central design challenge for this environment involves tasks that explore and assess mathematical argumentation involving aspects that go beyond correctness. Two main types of design patterns (DPs) formulating example-eliciting tasks (EET), which apply to a variety of curricular content and schooling contexts, led our design of assessment activities with an automated scoring platform, STEP (Olsher et al. 2016), and were found to provide essential information about students' knowledge: (a) analysis of constructed examples to support or contradict a claim, and (b) construction of examples conforming to a given definition (Yerushalmy et al. 2017). Our current study explores automatic assessment of students' mathematical logical reasoning when establishing the validity of geometry statements concerning the similarity of triangles. Research has demonstrated the crucial role of understanding the connection between examples and proving. Buchbinder and Zaslavsky (2019) offered a comprehensive review and

✉ Shai Olsher
  olshers@edu.haifa.ac.il

  Michal Yerushalmy
  michalyr@edu.haifa.ac.il

[1] Department of Mathematics Education, University of Haifa, Haifa, Israel

analysis of research on the logical role of examples and counterexamples in different tasks, contexts, and settings. Studies had documented the interplay between exemplifying and proving in students' problem solving, and looked into processes involving students gaining insight into the differences between deductive arguments and empirical arguments (Chazan 1993). The understanding of argumentation by examples in relation to universal statements is especially challenging. We question the known use of examples to confirm existential statements and to contradict universal ones. We offer expressly designed tasks that foster the use of an interactive environment to exemplify claims concerning the logical conjunction and disjunction of subsets of a domain satisfying universal and existential statements. We seek to contribute to the study of automatic analysis of tasks that can inform teachers and researchers about students' understanding of universal and existential statements.

In this paper, we report on a study of the design of tasks that foster the construction of examples using a given interactive diagram to exemplify conjunctions and disjunctions of given statements. By analyzing the student-generated example spaces, we explored the opportunities of the environment and the specific task design pattern to automatically provide feedback and assess students' mathematical skills, based on logical relations between examples and universal statements. We asked, (a) What are the conjunctions and combinations of conditions that can inform about understanding of universal and existential statements? and (b) What are the characteristics of students' use of examples in EETs that were designed based on logical conjunction?

2 Theoretical background

2.1 Connecting examples and universal statements: the interaction between proving and exemplifying

A key affordance of technology is that it allows the easy creation of many different examples. Example generation, which lies at the heart of mathematical reasoning, is the cornerstone on which our study is based. Using geometry interactive diagrams, students are expected to assemble a repertoire of available examples and develop methods of example construction for their own personal use. Watson and Mason (2005) referred to this repertoire as a personal example space, and inferred that personal example spaces may contain a range of example types, including central or obvious examples that come to mind. According to Sinclair et al. (2011), a personal example space is not simply a list of remembered examples, but a structured space that could provide access to the structures in which examples may be constructed systematically. In determining the validity of

mathematical statements, a personal example space can be used to reveal reasoning processes related to the logical connections between examples and statements.

Proving that a mathematical statement is correct is a key reasoning process of mathematical activity. Generating examples, generalizing, conjecturing, and drawing diagrams are types of semantic reasoning that have been shown to inform proof generation (Zazkis and Zazkis 2016; Jeannotte and Kieran 2017). Examples could serve as inductive of general example-based arguments (Dreyfus et al. 2012), providing an initial step in the proving process. Mills (2014), who studied the types of uses of examples by instructors, offered a review of a wide range of proof-related skills supported by exemplification. Mejia-Ramos et al. (2012) found illustrating with examples to be a measure of understanding of a proof at the undergraduate level. These uses of examples are also typical of exploratory reasoning processes facilitated by technology-based learning environments introduced for learning school geometry.

Attempts have been made to study the interplay between exploration, based on exemplification with diagrams and measurement tools, and aspects that relate to engagement with proving mathematical statements. Stylianides and Stylianides (2010) sought to identify the "scientific argumentation practices" that students tend to apply in the context of dynamic geometry. DGE-based research focuses on argumentation practices and on helping students shift from local arguments formulated while working with the tools, to general convincing arguments (Jones 2000). Chazan (1993) studied students' understanding of similarities and differences between measurement of examples and proof. Chazan's analysis of interviews with students who experienced empirical methods in geometry using interactive construction software, focused on students' reasons for viewing empirical evidence as proof and mathematical proof as evidence. Through analysis of successful student responses Healy and Hoyles (2002) showed how dynamic software tools can help students move from argumentation to logical deduction. Analyzing the process of instrumental deconstruction in the case of interactive diagram-based tasks, Mithalal and Balacheff (2019) argued that abstract reasoning involves a mix of observation and deduction (p. 162). Arzarello and Soldano (2019) introduced the notion of cognitive (dis)continuity between argumentation and proof, and discussed what they call "the basic gap" that can arise in the classroom. "We call it the basic gap between (formal) proofs and (intuitive) arguments: whatever definition of proof is given, even the most open and inclusive, the basic gap is behind it and can make any approach to the proof in the classroom problematic" (p. 2). They suggested that DGEs can support forms of cognitive continuity, and they introduced a technology-based dynamic geometry activity to illustrate the gap. Based on the work of Arzarello et al.

(2012), they argued and illustrated by analysis that abduction bridges the gap between the explorative and the proving phases.

Buchbinder and Zaslavsky's studies (2009, 2019) focused on the construction of a framework that supports the analysis of the use of examples in proving mathematical statements. Within this framework, 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, and the 'then' part, the proposition that defines the set of all mathematical objects that exhibit a certain property. Introducing the use of examples as a way of determining the validity of mathematical statements, they suggested a collection of tasks for examining students' understanding of the roles of examples in proving and in determining the validity of mathematical statements. This framework proved useful in designing tasks that assess the logical connections between statements and learner generated examples in school geometry (Olsher and Yerushalmy 2017). The uses of example for demonstrating, confirming, or refuting existential statements are often found to be non-trivial. Komatsu and Jones (2018) reported that students often encounter difficulties in formulating counter-example diagrams. They found that DGEs were highly useful in overcoming these difficulties and helping students produce counter-example diagrams.

2.2 Task design principles for supporting the logic of universal statements

Tasks that elicit student production of multiple examples of a mathematical object have been used in clinical interview settings as a mechanism to explore learners' concept images and how those depart from curricular concept definitions (Zaslavsky and Zodik 2014; Watson and Mason 2005; Stylianides 2008). A similar kind of task asks students to submit multiple examples that illustrate the truth of a claim about the existence of mathematical objects with certain characteristics. Such a task contains a claim (e.g., 'There exist functions for which the same line can be tangent to the graph of the function in two different points') and asks students, if they agree with it, to submit multiple examples (e.g., to either make a sketch of the graph of such functions or to write the defining expression for such functions), and if they do not agree with the claim, to explain why such an example cannot exist. One characteristic of EETs is that they often have more than one right answer, and as a result, can reveal the divergent thinking of different students. Another quality of EETs is that they shift the attention of students from the single correct final answer to the learner-generated example space (Watson and Mason 2005). The presence of the claim and of multiple examples also creates opportunities for engagement with the logic of proofs and refutations, and the

use of mathematical argumentation (Lakatos 1976), and as a result, new opportunities for assessment of students' reasoning within dynamic environments (Sangwin et al. 2010). Thus, the coordination of examples and mathematical claims can provide important insights into both the learners' understanding of a concept and their understandings of the logic that is central to mathematical argumentation (Olsher and Yerushalmy 2017). EETs motivate processes that may reveal (mis)conceptions, limitations, and strengths in being able to generate appropriate examples, and the ability to reason with examples and verify that an example satisfies the given conditions (Zaslavsky and Zodik 2014).

To be useful in instruction, EETs must be focused on mathematically important conceptual structures, whose characteristics can be revealed by the examination of multiple examples and how they relate one to another. Such tasks can be especially powerful when they build on existing research that has identified persistent alternative conceptions and errors, whose presence can be identified in examples submitted by the student. For the present study, we chose the topic of the similarities of triangles. The topic is related to school geometry and proving in school mathematics. It is highly visual, which provides an opportunity for the use of a DGE and its affordances, and it is based on studies that enable research-informed task design. Chazan (1988) identified four areas of difficulty that students seem to have with the topic of similarity, each of which was observed in previous research to be an obstacle for high school students. The obstacle that is highly related to our research concerns the common misunderstanding of proportions in right triangles, which often leads to incorrect conclusions, based on the prototypic view that does not take into account the right correspondence between angles and sides. Chazan stressed that often the conclusions are relevant only to special cases (e.g., an isosceles triangle), where the central altitude is also the angle bisector. The prototypic view, the difficulty of appropriately identifying shapes within triangles, and the difficulty of correctly transforming them to support identification of the proportions, was reviewed by Yerushalmy (1993). Based on cognitive research (e.g., Hoz 1981; Anderson 1985), Yerushalmy identified similar obstacles when characterizing visual difficulties students had when solving problems with the Geometric Supposer, a geometry construction tool designed in the 1980s. The interactive diagrams that the present study used to support the online analysis of students' competence in the logic of existential and universal statements, captures these visual elements.

Various technology tools have particular characteristics that may be useful in supporting online analysis of EETs, for example, tools that help researchers collect and analyze differences in students' work, such as those used by Clark-Wilson (2010), who explored TI-Nspire, and Lee

et al. ([2006](#)), who offered learning and assessment dragging tasks in dynamic geometry environments. The STEP platform is built around the GeoGebra tool that students use to produce and submit their examples and to create their personal example spaces. The interactive diagrams and tools of GeoGebra offer means for the exploration of tasks with multiple solutions. The STEP platform works with the mathematical characteristics of examples that are identified by developers or by end users. Users (teachers, researchers, or developers) can specify the mathematical characteristics required of examples that are deemed correct, as well as the mathematical characteristics expected in typical incorrect examples. Using these specifications, STEP can indicate whether the examples meet task requirements: Do they represent familiar misconceptions? What other characteristics do they have, beyond the correctness? How are generated examples (of the same student or of different students) similar to or different from one another? STEP makes possible the online monitoring of work on rich EETs.

STEP was designed and developed to support EETs that are formulated to be implemented as online assessment tasks along specific design patterns (DPs). The DPs are all based on one of two requirements: analysis of constructed examples to support or reject a claim, or construction of examples complying with a given definition. Following Mislevy et al. ([2017](#)), the DPs developed offer a variety of approaches that can be used to obtain evidence about reasoning processes. Cusi and Olsher ([2019](#)) introduced a task DP that promotes the use of a certain type of example, which they refer to as *limit confirming examples*. Nagari Haddif and Yerushalmy ([2015](#)) studied a DP that supports the development of tasks in which students construct and submit examples for refuting or supporting a statement in a multiple linked representations environment, in which they can construct any supporting example, without any constraints. By contrast, Olsher and Yerushalmy ([2017](#)) studied a DP in which the examples can be constructed only by dragging and altering an already constructed diagram, by zooming, translating, or rotating it.

In the present study, we explored opportunities for automatically assessing students' dealing with logical relations between examples and universal statements. We asked what the characteristics are of students' use of examples in geometry EETs designed based on logical conjunction, and what the challenges are in using the logical operands for online automatic analysis. In reporting on this research, we seek to further contribute to the emerging field of online assessment by the study of student-generated example spaces when using interactive diagrams to support compound claims concerning the similarity of triangles.

3 Analytical framework

The activity comprised two multiple-selection with supporting examples tasks (Olsher and Yerushalmy [2017](#)). In this DP, the context of the task was a dynamic figure in multiple representations: a draggable digital geometry environment (DGE, in this case GeoGebra) construction (Fig. [1](#)), a symbolic representation of the constraints embedded in the DGE construction, and a verbal description of the task, including several relationships between lengths of segments in the figure. Measurement tools and numeric feedback were not supported in this task. Students were given three relationships (i, ii, iii in Fig. [1](#)) to consider. Finally, for each context described in the interactive diagram, three statements were provided (1, 2, and 3 in Fig. [1](#)). Participants needed to determine which of the statements were correct (potentially, more than one). In addition to selecting the correct statements, participants were asked to construct an example with the applet in the given context that exemplified each of the statements they had marked. They could drag different points of the figure, and use a pen tool to add graphic annotations on the dynamic diagram.

Our working definitions are adapted from those of Buchbinder and Zaslavsky ([2009](#)), when referring to universal mathematical statements, as follows: "Every mathematical statement can be characterised by the domain (D) of objects (x) to which it refers to and a proposition (P(x)) that specifies some property. A Universal statement states that a proposition is true for all the objects in the domain: $\forall x \in D, P(x)$. An Existential statement asserts that there exists an object in the domain for which the proposition is true: $\exists x \in D, P(x)$" (p. 28). For example, given domain D: "all triangles DEF similar to triangle ABC" and the property P: "AB/DE = BC/EF," we can form two types of statements: (a) a universal statement, "In all of the triangles DEF similar to triangle ABC, AB/DE = BC/EF;" and (b) an existential statement, "There is a triangle DEF similar to triangle ABC, for which AB/DE = BC/EF".

It is clear that an example is sufficient to refute a universal claim: $x \in D, \neg P(x)$. For example, if one can provide a triangle DEF that is similar to triangle ABC, and which fails to satisfy the statement AB/DE = BC/EF, the universal statement mentioned above is refuted. Furthermore, an example is sufficient to confirm an existential statement: $x \in D, P(x)$. For example, if one can provide a triangle DEF that is similar to triangle ABC, in which AB/DE = BC/EF, the existential statement mentioned above is confirmed. Yet, while defining $x \in D, P(x)$ as a "confirming example" (Buchbinder and Zaslavsky [2009](#)) in the broad context of both universal and existential statements, Buchbinder et al. ([2017](#)) referred to $x \in D, P(x)$ as "supporting"

Fig. 1 Multiple selection with supporting example task 1

In the dynamic figure below, triangle ABC is a right triangle with the right angle at C. DC is the altitude of triangle ABC. Explore the dynamic figure and consider the following relationships between the lengths of segments in the figure:

i) $CD^2 = AD \times DB$
ii) $BC^2 = AB \times DB$
iii) $AC:CB = AD:DB$

For each statement below, if you think a statement is true, please save an example from the dynamic figure that shows the statement is true. If you think a statement is false, leave the statement as is.

1. A triangle ABC exists for which all three relationships (i, ii, and iii) are true.
   Save

2. A triangle ABC exists for which only relationships i and ii are true.
   Save

3. A triangle ABC exists for which only relationships i and iii are true.
   Save

AC⊥CB
CD⊥AB

in the context of universal statements only: "A supporting example is an element of D which satisfies the property P ($x \in D, P(x)$). Although it supports the universal statement, it is insufficient for proving it, since in order for a universal statement to be true the property has to hold for *all* objects in the domain" (p. 220). This refinement is possible only when taking the existential statement out of the discussion, as a supporting example is a confirming example in the case of an existential statement.

In the present study, we examined a possible extension of the role of the properties $x \in D, P(x)$ in relation to the universal statement $\forall x \in D, P(x)$, in a certain domain D, as constructed in an interactive diagram. We proposed a certain setting: the setting in Fig. 1, where ABC is a right triangle, and CD is perpendicular to AB. Specifically, students are given a set of relations to examine within the provided domain. Some of the given relations are true universal statements for the domain D (i.e., they are always true), and some

are existential statements in domain D (i.e., there are cases for which they are true). Table 1 shows the different relations, and expressions comprised of these relations, and for each one, it states whether it is a universal statement under D meaning that it is always true, or whether it is an existential statement under D, meaning that is true for a subset of examples included in D.

The rationale for building these relations is that we can use the "and" operand to create a variety of logical conjunction types. Mainly, we designed the relations and the expressions in ways that require attention to the geometric attributes, but more important, to the domain of existence. For example, there is a logical conjunction between sets satisfying a universal statement over D (e.g., R1) and an existential statement over D (e.g., R3), to obtain an expression that is an *existential statement* (E1), which is true for the same subset *that is true for the granular existential statement* (R3). Similarly, a logical conjunction between sets satisfying the universal statement with

Table 1 Domain, relations, expressions, statements, and truth values for task 1

Relation/expression	Definition	Validity in the domain D of right triangle ABC with CD $\perp$ AB		Universal, existential, or false statement in domain D
		Subdomain of D: isosceles triangle ABC	Subdomain of D: scalene triangle ABC	
Relation 1 (R1)	$CD^2 = AD \times DB$	True	True	$\forall x \in D, R1(x)$
Relation 2 (R2)	$BC^2 = AB \times DB$	True	True	$\forall x \in D, R2(x)$
Relation 3 (R3)	$AC:CB = AD:DB$	True	False	$\exists x \in D, R3(x)$
Expression 1 (E1)	$R1 \wedge R2 \wedge R3$	True	False	$\exists x \in D, E1(x)$
Expression 2 (E2)	$R1 \wedge R2 \wedge (\neg R3)$	False	True	$\exists x \in D, E2(x)$
Expression 3 (E3)	$R1 \wedge (\neg R2) \wedge R3$	False	False	$\forall x \in D, \neg E3(x)$
Expression 4 (E4)	$(\neg R1) \wedge R2 \wedge R3$	False	False	$\forall x \in D, \neg E4(x)$

the complementary subset (not complying with the existential statement) gives us an expression that is an *existential statement* over D, which is satisfied by a subset that *does not satisfy the granular existential statement* (R3). In addition, a logical conjunction between sets satisfying one universal statement with the complementary of another universal statement (R1, ¬R2) results in a false universal statement in the expression that is satisfied by the empty set, which means that it cannot be satisfied over D.

These expressions are means to assess students' logical reasoning on two dimensions: (a) identification and argumentation about universal statements in a certain domain, and (b) the conditions required to satisfy an existential statement in that domain. Yet, although providing an example is sufficient only to prove an existential statement or disprove a universal statement, we suggest the use of examples also in cases where expressions include the logical conjunction of subsets of a domain satisfying universal and existential statements: one in which all the relations need to be satisfied (e.g., expression 1), and one in which the existential statement is refuted (e.g., expression 2). These two expressions can provide us with examples that suggest that the student takes into account the conditions under D that satisfy the existential statement only, demonstrating that the universal statements do not matter because they are always true over D. We can also use statements that include both a universal statement joined with a refutation of a universal statement (expressions 3 and 4 in Table 1), expecting students not to choose this statement as true because no example in D can satisfy it. This design enables us to work with automatically analyzable examples also when supporting universal statements.

4 Method

The study used design-based research principles (Cobb et al. 2003) to explore design principles of EETs in the context of universal statements. The two-cycle study focused on tasks concerning the topic of triangle similarity, and the various representations of ratios between different corresponding sides. The first cycle was conducted in two high-school classes by the developers of the tasks. The second cycle was conducted by the teachers of the classes described below as part of their geometry class. We report here the results of the second cycle.

Participants were 50 high school students from three 10th grade classes (ages 15–16) in two high schools in Israel. The students were familiar with the STEP platform and completed an online activity as part of Euclidean geometry classes on the similarity of triangles. The activity was conducted after the students finished learning the topic of similarity of triangles and Thales theorem.

4.1 The tasks

Task 1, shown in Fig. 1 and described in Table 1 (Expressions 1, 2, 4 in Table 1), was designed to provide an opportunity for students to demonstrate whether they recognized what relations were always true (R1, R2 in Table 1), and which were true only in special cases (R3 in Table 1). The logical conjunctions between the relationships that formed the various expressions were designed to provide an opportunity to differentiate between the complementary subdomains in which each expression could be satisfied: isosceles triangle compared to scalene triangle.

Task 2, shown in Fig. 2 (its components are described in Table 2) was constructed using the same DP as task 1. Yet, to add another layer of validation to the answers submitted to task 1, this task used expressions 1, 2, and 3 (Table 2) as answer choices. Expression 1 is similar to the one in task 1, including a conjunction of all of the relations that need to be satisfied. Expressions 2 and 3, however, include both a universal statement joined with a refutation of a universal statement (statements 2 and 3 in Table 2). Later, it was possible to compare these choices with the choices students made on task 1, to suggest patterns of work when constructing

Fig. 2 Multiple selection with supporting example task 2

In the dynamic figure below, triangle ABC is a right triangle with the right angle at B. D is the midpoint of side AC and DE is perpendicular to AC.

Explore the dynamic figure and consider the following proportions:
i) BC:AC = ED:AE
ii) AB:ED = AC:BC
iii) DE:BC = AD:AB

For each statement below, if you think a statement is true, please save an example from the dynamic figure that shows the statement is true. If you think a statement is false, leave the statement as is.

1   A triangle ABC exists for which the three proportions (i, ii, and iii) are true.   [Save] [🖼]

2   A triangle ABC exists for which only proportions i and ii are true.   [Save] [🖼]

3   A triangle ABC exists for which only proportions ii and iii are true.   [Save] [🖼]

∠CBA=90⁰
AD=DC
DE⊥AC

Table 2 Domain, relations, statements, and truth values for task 2

Relation/statement	Definition	Validity in domain D of right triangle ABC with ED ⊥ AC, AD = DC		Universal, existential, or false statement over D
		Isosceles triangle ABC	Scalene triangle ABC	
Relation 1(R1)	BC:AC = ED:AE	True	True	$\forall x \in D, R1(x)$
Relation 2 (R2)	AB:ED = AC:BC	True	False	$\forall x \in D, R2(x)$
Relation 3 (R3)	DE:BC = AD:AB	True	True	$\exists x \in D, R3(x)$
Expression 1 (E1)	R1 ∧ R2 ∧ R3	True	False	$\exists x \in D, E1(x)$
Expression 2 (E2)	R1 ∧ R2 ∧ (¬R3)	False	False	$\forall x \in D, \neg E2(x)$
Expression 3 (E3)	(¬R1) ∧ R2 ∧ R3	False	False	$\forall x \in D, \neg E3(x)$

examples or providing justifications, in a way that is close to task 1 in the DP, but not similar. Similarly to task 1, task 2 also provides automatically analyzable examples also when supporting universal statements.

4.2 Automatic checking of tasks' solutions

The tasks in this study were designed for the system to indicate whether the corresponding example is part of the subdomain satisfying the relevant expression (in this case, equal or unequal length of triangle sides indicates whether the example is of a scalene or an isosceles triangle). A margin of accuracy was determined by the teacher, within which solutions are considered sufficiently accurate.

4.3 Data sources and analysis

Our unit of analysis was the student work, consisting of answers submitted in STEP to the two tasks. Data sources included the submitted answers. The first stage included identifying discrepancies between a correct choice and the attached example. For each participant, we checked the correctness of marked statements and the presence of examples, aggregating the number of participants who chose the two correct statements as the answer to task 1 and the number of correct statements for task 2. In the second stage, we refined the analysis: for each of the answers, we checked its automatic assessment and coded the different types of evidence for a correct answer and justification that was not checked automatically. The third stage included the coding of incorrect answers and examples (e.g., *familiar mistakes* or *additional reasoning*), to study the characteristics that could be automatically assessed relating to both the topic (similarity of triangles) and the logical conjunction of different types of logical statements (universal, existential, and the refutation of either).

5 Results

We report the results in four parts, based on automatic analysis on the STEP platform, and on detailed manual analysis of specific work. The first part includes the description of patterns of success that typify the work of the sample on each of the two tasks. The other three parts describe various characteristics, which in most cases were analyzed automatically, and provide potential insight into patterns that could be looked into when assessing student work: the second part includes various successful answers, the third part includes explainable mistakes, and the fourth part includes evidence of students' work characteristics on the two tasks.

5.1 Patterns of success

As noted, the two tasks were designed along the same pattern. The results show that 16 out of 50 submitted answers to task 1 (32%) and 13 out of the 46 submitted answers to task 2 (28%) were fully correct.

The fully correct answer for task 1 required examples for two statements (expressions 1 and 2 in Table 1). Eighteen out of 50 students did not select both correct expressions and did not attach correct examples. In addition to the 16 fully correct answers for task 1, we identified 16 partially correct answers, which generally consisted of a correct example for statement 2.

The complete answer to task 2 consists of choosing one statement and an example of an isosceles triangle. Having only one correct expression, leaves fewer options for partially correct answers. Thirteen out of 46 students (28%) provided an example of an isosceles triangle for statement 1. Task 2 resembles the relative complexity of task 1 regarding the logical conjunction of universal statements with the special case that complements what seems to be incorrect from a universal perspective.

5.2 Characteristics of evidence of successful answers for task 1

We present the evidence about successful answers to task 1 in two parts: first, we analyzed examples that can be categorized automatically as correct; second, we manually checked answers that were flagged for human assessment.

5.2.1 Automatic identification of distinct examples

For automatic assessment by STEP, correct answers for task 1 are defined as the selection of statements 1 and 2, accompanied by an isosceles triangle ABC as an example for statement 1, and a diagram that does not have an isosceles triangle ABC exemplifying statement 2. Although this definition for automatic identification seems to be straightforward, we must explain why the rich assessment we tried to support is more challenging than is a merely algorithmic one, but provides even more information about the student's work, in addition to the correctness of the answer.

We automatically assessed as correct answers all submissions to task 1 that (a) consisted of two automatically distinguishable figures: an isosceles triangle attached to statement 1, "A triangle ABC exists for which all three relationships are true," and a figure that can be identified as not an isosceles triangle attached to statement 2, "A triangle ABC exists for which only relationships i and ii are true," and (b) did not include any selection or example attached to statement 3 "A triangle ABC exists for which only relationships i and iii are true.". We assume that the choice of two such examples,

as shown in Fig. 3, is not incidental, and demonstrates consideration of the difference between universality and existence. To strengthen the evidence, STEP also automatically assessed whether the answer to task 2 of the same student was correct.

Some of the submissions of two distinct figures to statements 1 and 2 included marked notations or textual explanation and calculations. The figures that contained notations were automatically identified by STEP as such. When manually assessing the various notations added to correct examples, we found that these either marked congruent segments (as in the case of the student whose submission is shown in Fig. 4), contained computations of all the angles involved, or contained text demonstrating a process of drawing conclusions about the special case of two congruent angles (of 45°). All of the above provided additional justifications for the submitted example.

5.2.2 Identical figure distinguished by semantic identifications

Figure 5 shows a submission for task 1 that was automatically assessed as partially correct. The reason for this assessment is the fact that the two attached figures are indistinguishable when analyzing only the constructed triangles. The figure for statement 2 was assessed as correct because it did not include an isosceles triangle ABC. The submission was also characterized as containing a notation, and was manually assessed. Manual analysis of the notations suggested that although it was geometrically identified as not an isosceles triangle ABC, it was furnished with notations that are specific to isosceles triangles for statement 1 (Fig. 5a), and not for statement 2 (Fig. 5b). The student appears to have used the same constructed figure, rather than providing a unique correct geometric construction for each statement. The notations make the identical figures distinguishable constructions. The first one (Fig. 5a) includes the evaluated angle value as 45°, and the second (Fig. 5b) does not. As in the previous case (Fig. 4), the justification was provided on a drawing that resembles the position of the traditional textbook figures demonstrating similarity within right triangles. Further analyzing the submitted figures, the student appears to have been working toward a formal justification that may resemble the work performed with the horizontally aligned figures that traditionally populate figures of tasks relating to the similarity of triangles in textbooks.
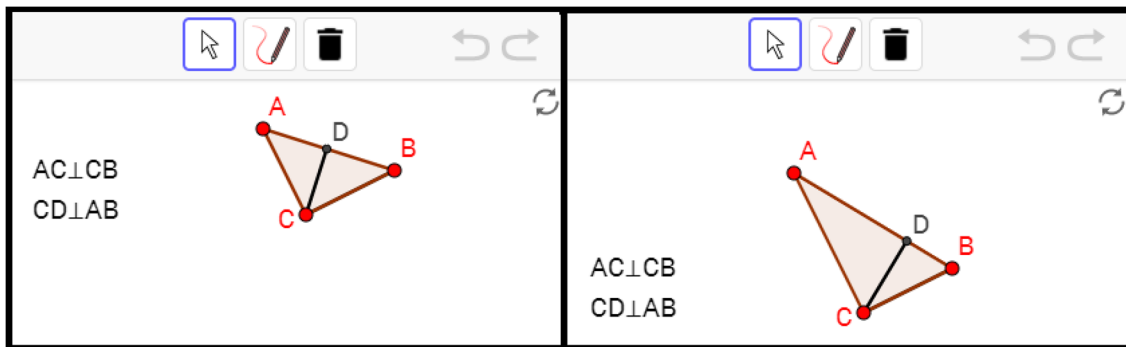

Fig. 3 Correct examples attached to statement choices of a student for task 1

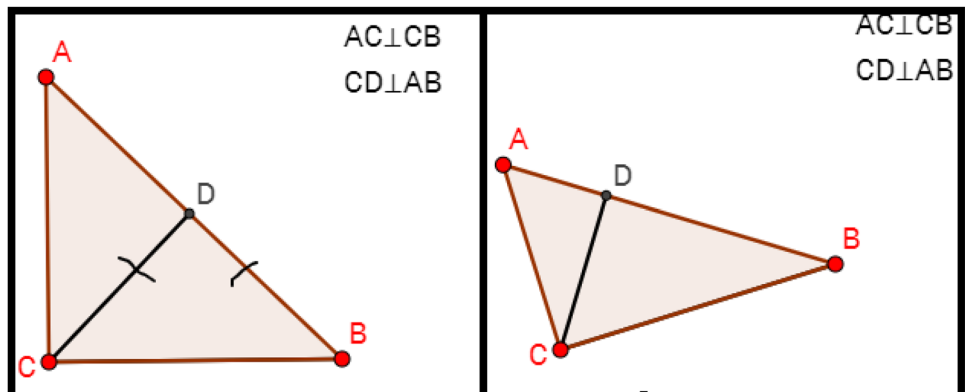**Fig. 4** Correct examples with notations (left) attached to task 1

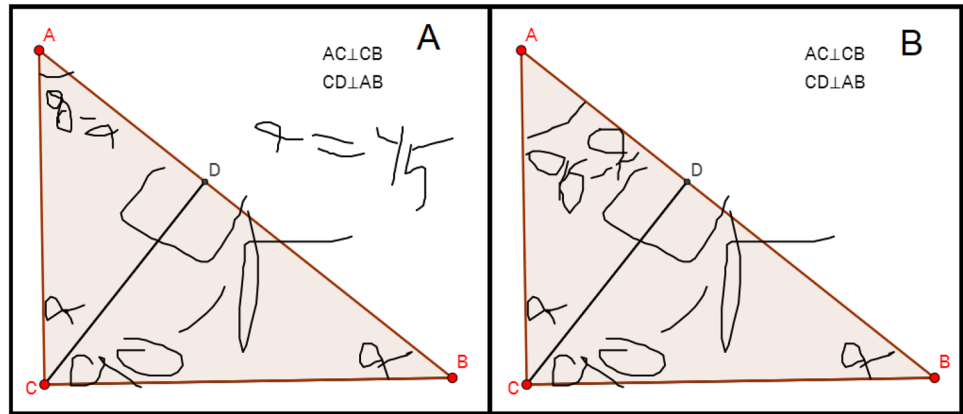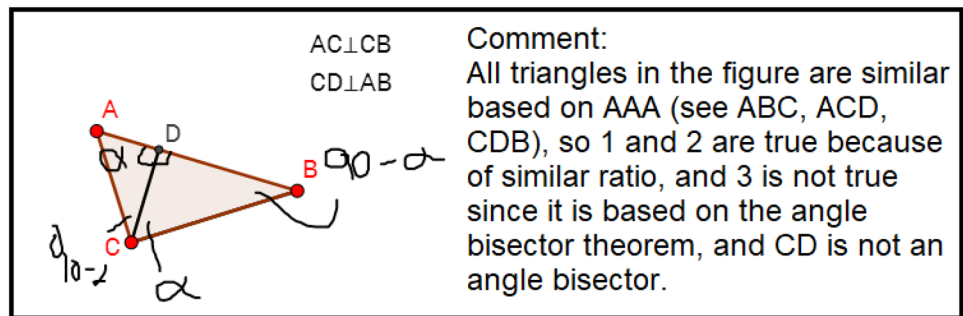**Fig. 5** Similar figures with different notations submitted as examples for task 1

Fig. 6 Partially correct answer, with verbal explanation, for task 1

AC⊥CB
CD⊥AB

Comment:
All triangles in the figure are similar based on AAA (see ABC, ACD, CDB), so 1 and 2 are true because of similar ratio, and 3 is not true since it is based on the angle bisector theorem, and CD is not an angle bisector.

5.3 Identification of explainable mistakes

Apparently, students were challenged by an existential statement that is true only for a subdomain of special cases, exemplifying R1∧R2∧R3, of which two relations are universally true for the domain (R1 and R2 in task 1 and R1 and R3 in task 2), and either R3 or R2 are true for a subset of the domain. In this section, we present evidence of student work concerning the logical conjunction of various parts of the domain of examples presented in relation to tasks 1 and 2.

5.3.1 "IS" means "always IS:" confusing "always" true with "sometimes" true

Some of the answers contained a textual argument, analyzed manually, that excluded the possibility of exemplifying statements that are true in cases that exist only in a subdomain. Figure 6 shows a submission that includes a textual explanation (all student comments were translated from Hebrew, with slight modifications to correct grammatical errors), together with a partially correct answer for task 2. The student, who chose and exemplified statement 2, argued that R3 is not true because CD is not an angle bisector. Because that was not a given, the relation is "false." The student did not acknowledge the possibility of a special case, represented by an isosceles ABC, which satisfies R3, in which the given
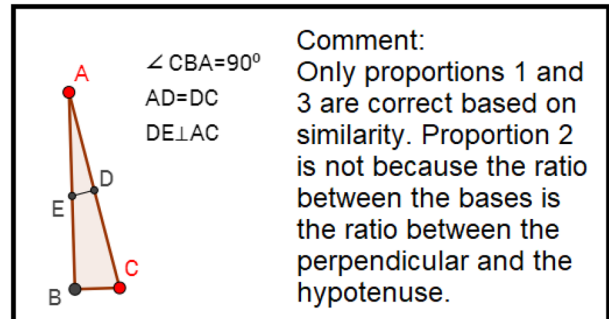

∠CBA=90⁰
AD=DC
DE⊥AC

Comment:
Only proportions 1 and 3 are correct based on similarity. Proportion 2 is not because the ratio between the bases is the ratio between the perpendicular and the hypotenuse.

Fig. 7 Partially correct answer, with verbal explanation, for statement 1 of task 2

altitude is also an angle bisector. This submission indicates confusion between "always true" and "sometimes true" cases. The difference between these can be demonstrated by a special type of example belonging to a subdomain of the domain of all possible examples.

Figure 7 shows a submission of an example and comment exemplifying statement 1 of task 2. The student provided an incorrect example. The initial state of the figure represents a "generic" figure, not an isosceles triangle that would satisfy statement 1. The answer includes a verbal explanation of why the submitted figure is a
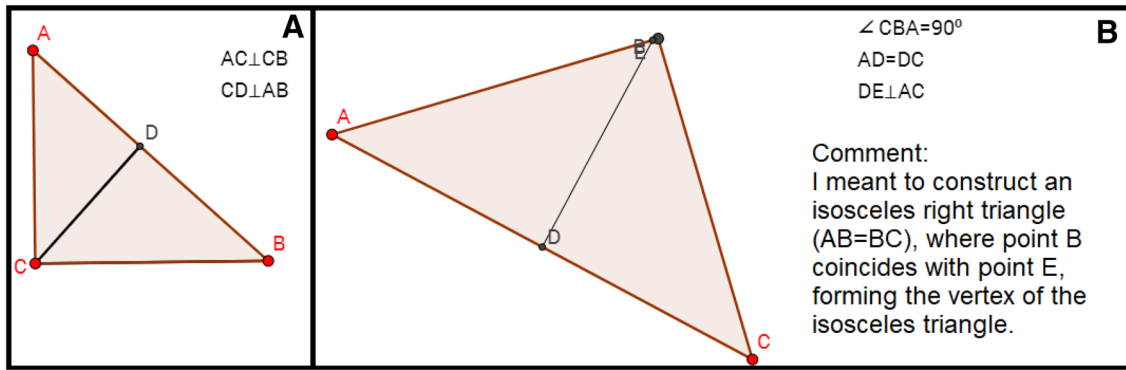
Fig. 8 Examples and textual explanation satisfying all given relations

non-example. The explanation indicates that only R1 and R3 are correct (these relations are universally true), but R2 is false because the proportion refers to non-corresponding segments.

5.3.2 Finding an example that satisfies a conjunction of all given relations

In some cases, the automatic assessment indicated correct submissions for statement 1 of both tasks, and no examples for statement 2 (in task 1). Further analysis of the verbal explanations suggest that students, as illustrated in Fig. 8, considered examples only for statements that represent conjunction of all the given relations (R1 ∧ R2 ∧ R3), and did not exemplify statements that exclude one of the three (R1 ∧ R2 ∧ (¬R3)).

5.3.3 Disregarding negation

Automatic assessment of answers shows that some examples of general cases were exemplified by isosceles triangles. This may have to do with the fact that students did not take into consideration that one of the relations was negated, as implied by "only." This type of answer may suggest that students misunderstood or ignored the fact that exactly two of the relations should be satisfied, and the third should not. In Fig. 9, a student submitted an example containing an isosceles ABC for task 1, statement 2. The student did not attend to the constraint ¬R3. R3 is satisfied for an isosceles ABC, and therefore ¬R3 is false for an isosceles ABC, so that statement 2 excludes the subdomain of examples in which ABC is isosceles.

The automatic assessment also identified a few submissions of examples containing an isosceles ABC for statement 3 of task 1: R1 ∧(¬ R2)∧ R3. In the example shown in Fig. 10, R1 and R3 are universally true in the domain of the given diagram. R2 is true for the subdomain of examples in which ABC is isosceles, representing a special case. Thus,
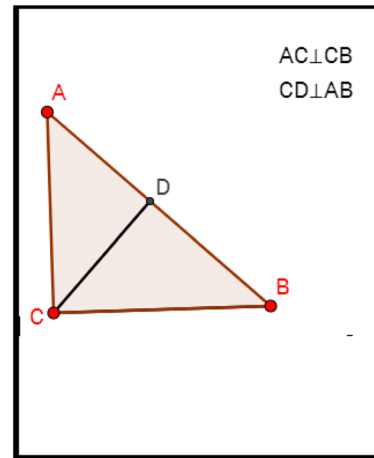

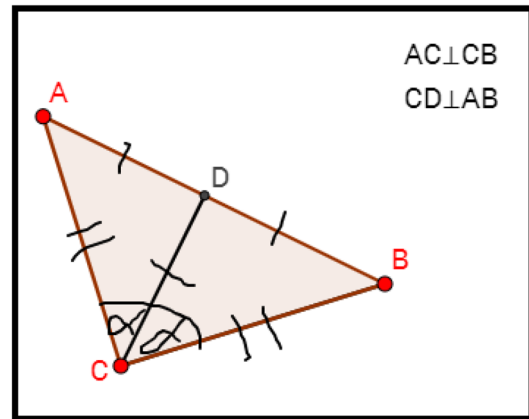Fig. 9 Example disregarding the negation of an existential statement

Fig. 10 Example disregarding the negation of a universal statement

¬R2 is false for an isosceles ABC, and the example provided is incorrect for statement 3. This could be a result of the same type of pattern that does not take into account the

negation of R2, but only the fact that R1 and R3 are satisfied by this example.

5.4 Identification of work characteristics across tasks

The common DP of the two tasks, and the same mathematical topic addressed in them, enabled us to assess sets of answers submitted by a student to find initial evidence of work strategies or response patterns that appeared in both tasks. Assessing sets of answers for both tasks that resulted in similar patterns based on success, is yet another way of enhancing the reliability of the analysis. Identifying both success across tasks and inconsistent answers for the two tasks serve as methods designed to achieve a deeper analysis of learning by individual students.
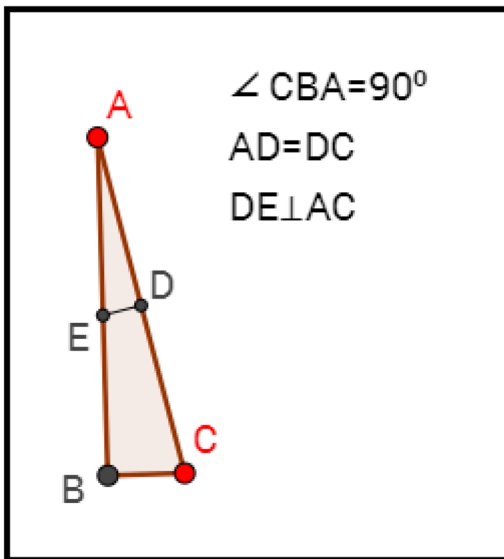
5.4.1 Identifying success across tasks

Five students provided fully correct answers to both tasks. Another three answered task 2 correctly, and task 1 partially correctly. These results suggest that the students can use supporting and confirming examples. The submissions also suggest that some students can distinguish between universally true conjunctions and existential logical expressions, and are able to exclude false expressions.

5.4.2 Consistent mistakes across tasks

In several cases, students submitted what we identified as consistent mistakes in both tasks. In Fig. 11, we show an example submitted for task 2 that is similar to the initial state of the diagram. The student also submitted the initial state (only one example) to exemplify a single statement in task 1. We found this type of response pattern in several cases for one task. When a student submits the initial state as an example multiple times, our confidence in the assumption that this is a response pattern increases.

One student chose and exemplified statement 2 of task 1 as a conjunction of two universally true relations, as shown in Fig. 12a. The student wrote that only these two relations were "correct," and submitted a supporting example of a generic diagram (a non-isosceles triangle ABC) for task 2.

In Fig. 12b, the student chose and exemplified statement 3 only, writing that R2 and R3 are the only correct relations, providing an isosceles triangle ABC as an example. The two expressions R1 ∧ R2 ∧ (¬R3) in task 1, and R2 ∧ R3 ∧ (¬R1) in task 2 were mistakenly considered coherent (the analogous expression should have been R1 ∧ R3 ∧ (¬R2), but it did not appear in task 2 as part of the design considerations stated above).

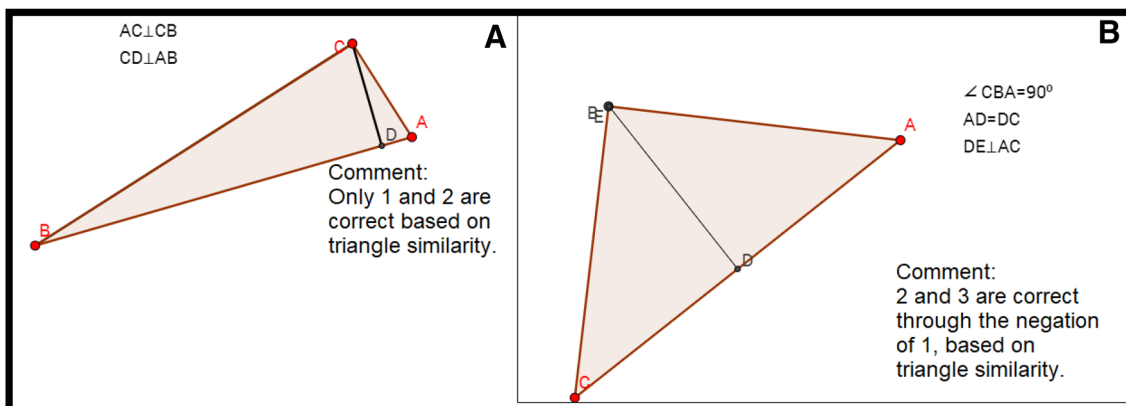**Fig. 11** Example using the initial state of the diagram

Fig. 12 Examples and textual explanations including universally true relations

5.4.3 Inconsistent answers across tasks

The two tasks are similar but not identical. In designing the two tasks, we sought commonalities and differences between answers. As shown above (Sect. 5.1), in general, the answers reflect the commonality but at the same time STEP automatically suggests differences by identifying an inconsistency between correct answers to one task and incorrect answers to the other. We present an instance of structural difference that may have led to the identified inconsistencies.

It demonstrates the fact that the two tasks are logically identical but structurally different, as the list of the relations does not follow the same order of validity. The student made a wrong decision, the intentionality of which cannot be assessed, to use the same strategy for answering both tasks, not noticing the differences between the two. The three relations in the two tasks have two relations that are universally true in the domain of possible examples, and one that is true only for a subdomain of examples in which ABC is isosceles, but they appear in a different order, and so do the logical expressions. We suspect that part of the mistakes that suggest inconsistent performance were the result of the tendency to apply the answering strategy that worked for the first task in the case of the second task as well, mistakenly assuming that the two tasks are logically identical. Figures 13a, b show such examples submitted for statement 1 and statement 2, which answered task 1 correctly. Then, reproducing the same example types for task 2, the student attached the examples that appear in Figs. 13c, d for the first two statements in task 2, which is incorrect.

The second case of identified inconsistency appears to relate to the fact that statement 1 of task 1 was constructed upon the same pattern as statement 1 of task 2, and was based upon upon similar required geometric transformations. It is therefore puzzling why students who presented a confirming example for statement 1 of task 2 failed to do so in task 1. Figure 14 shows the submission of a student who chose one statement for each task. Figure 14a shows the single example submitted for statement 2 of task 1, and 14B the example for statement 1 of task 2. Both examples are correct, but task 1 is only partially answered. We hypothesize that students who exhibit this response pattern might have operated under the assumption that only a single answer is required for each task, as in traditional multiple-choice tasks, and would have been able to answer task 2, which has only a single confirming example, correctly. We already have evidence of this in our previous research (Olsher and Yerushalmy 2017). Students checked the correct answer to statement 2 of task 1 using the given default figure, and ignored the request to check and exemplify each of the three statements.

Both tasks were designed along similar patterns, which can be used to design additional tasks in other contexts. Reusing these patterns in future studies may be useful in confirming or refuting the present findings.

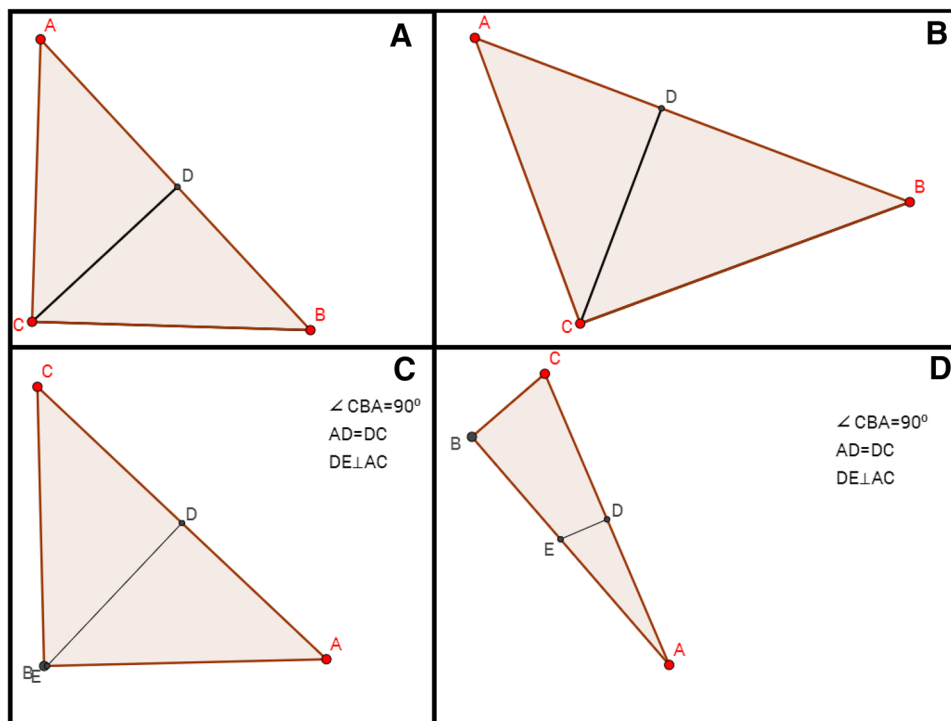**Fig. 13** Special and general case with similar positioning submitted for both tasks by a student
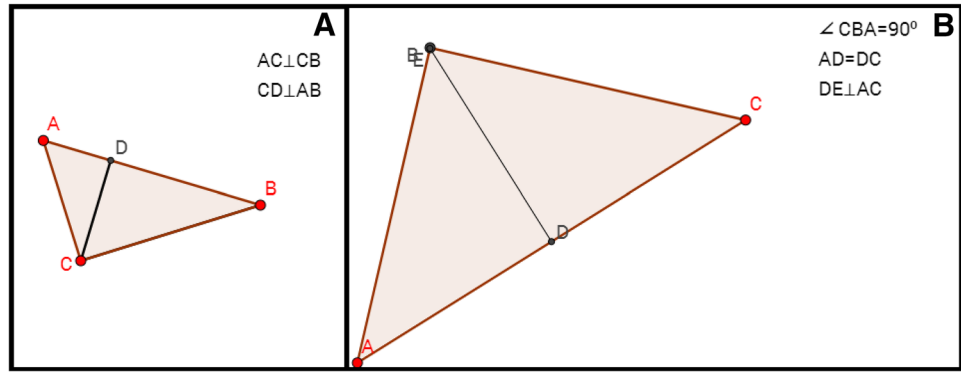
Fig. 14 Single correct examples submitted for tasks 1 and 2 by one student

5.4.4 Strategic decision: preference for using the given diagram

Some students used the initial state of the diagram as the example. This phenomenon appeared mainly in response to statement 2 of task 1, where the statement exists for the domain, except for statement 1 of task 2, which should be an isosceles triangle. Figure 15 shows a submission in which, in both cases, the figures are the initial states of the given diagram.

6 Discussion and conclusions

In this study, we explored opportunities for automatic assessment of the interplay between (a) students' logical reasoning with regard to relations between examples and universal statements, and (b) students' geometric skills in dealing with the similarity of triangles. The findings show that the suggested DP of tasks enables automatic assessment and differentiation between various levels of correctness.

Our initial findings support those of Olsher and Yerushalmy (2017), which showed that because the students are required to formulate their claims, and also need to construct an example within a given domain to support their claims, their answers can further distinguish between

possible guessing and competence. We argue that students are probably able to distinguish between universal and existential statements in a given domain if they submit a supporting example for the universal case, and a proving example for an existential statement in the same domain. It is quite likely that students would provide the 'default' example to exemplify that they believe a statement is universal in a certain domain. Yet, when a special case is added, exemplifying another statement that happens to be an existential one provides another layer of validation to the initial, default example. Furthermore, if this student provides another special case in another context, also identifying that there cannot be examples for expressions containing a conjunction between a universal statement and a negation of a universal statement in the given context, yet another layer of validation of this student's competences is obtained. An example of a universal statement could provide mere support, but this study demonstrates that expressions containing a conjunction of universal and existential statements, and their negation, could be automatically assessed, and could serve to distinguish between the students who submitted the answers. Using the task DP presented in this study, we were able to better assess different students' work and use of examples of universal and existential statements.

The automatic analysis detected response patterns of successful answers that showed two distinguishable examples

Fig. 15 Initial states submitted as examples for tasks 1 and 2
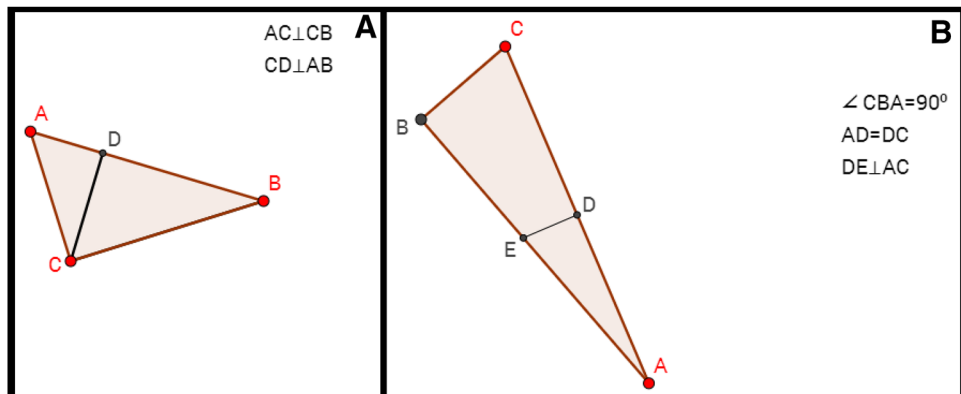
demonstrating different sub-domains of the constructed interactive diagram. Yet, some of the students did not submit different examples, but rather chose (or were only able) to substantiate the chosen statement by providing examples that were annotated graphically or with added verbal and symbolic explanations. This required paying further attention to verbal answers and graphic annotations, which could be automatically identified, even if not completely analyzed, bringing them to the teacher's attention.

The suggested DP revealed different types of explainable mistakes in student answers. One phenomenon, which appeared in several presentations, was that of students having difficulty distinguishing universal from existential statements. Some students considered a statement to be true only if it was true on the full domain (mixing universal and existential statements), and provided examples that did not consider conjunction with negation, which is an indication of an example of a specific subdomain. The partial answers identified in the study consisted mostly of a correct example for a statement that required students to exemplify existence of 'any' triangle that is not isosceles. These partial answers helped us to address the second research question, regarding the characteristics of students' use of examples in EETs that were designed based on logical conjunctions.

The results suggest that examples of a statement describing the conjunction of existence of all three relations were less common. A possible explanation is that students begin their solving process by checking the given relations. They examine the context, realizing that some relations are always true in the given context, but the remaining relation not always. Thus, when considering statements, they recognize the universality of relations that are always true, and, the 'almost universality' of the remaining relation not being true, they mark it as correct, providing an example of a figure, which is often identical or similar to the originally given figure. It is not possible to make a strong argument that the evidence suggests awareness of the negation of the remaining relation. As many answers provided successful examples only for the statement that required the 'general', non-isosceles, triangle, we can infer that they are likely to provide the correct answer only by considering the universality of the relations that are always correct. They assumed that the original figure demonstrates exactly this statement, and they provided a correct example. Because our automatic analysis does not follow the order of work that led to the final submissions (in STEP, students may go back and forth, keep their own portfolio of saved choices and examples, and determine the timing and order in which they submit their answers), we can only conjecture that this particular statement was the first they picked as a correct one, which fits the given figure. Future investigations should examine this conjecture using tasks that follow this DP. For this future research, our results suggests a sequence of operations beginning with consideration of the easily available figure, followed by choice of a statement, rather than the expected process of first analyzing a statement and then constructing a supporting example. The statement required students to consider the conjunction of the relations that are always true, and exclude all cases in which the remaining relation was incorrect. For a similar reason, we assume that many students did not consider to be true, a statement that violates the universal truth of one of the relations that is always true.

Some of the student responses across the tasks suggested patterns that incorporate strategic considerations. These response patterns are indifferent to the content of the tasks, therefore at times resulted in mistakes. We interpret such mistakes as stemming from an incoherent concept image, resulting in inconsistent reasoning in similar logical situations (Buchbinder and Zaslavsky 2019). The reason may have to do with a contextual difference between the tasks. Our findings show that the confirming example for the existential statement satisfying all the relations in the first task was more frequently correct than the example of an isosceles triangle in the analogous statement in the second task. Another hypothesis suggests that some mistakes were the result of faulty strategic assumptions of similarity between the two tasks with respect to (a) the number of true statements (2 in task 1 and 1 in task 2), and (b) the inconsistency between the indices of the given relations (in task 1 R1 and R2 are universally true, whereas in task 2, R1 and R3 are universally true). Additional studies and a larger body of submissions are needed in order to verify this hypothesis.

We argue that examples can do more than just illustrating the truth of existential statement or disconfirming the truth of a universal statement. We suspect that the mathematics we analyzed, which involved justifying and exemplifying, is ultimately related and central to the understanding of deductive proof (Dreyfus et al. 2012). A central challenge of computer-assisted assessment is to develop ways of collecting rich and complex data that can nevertheless be analyzed automatically. The rich assessment shown here provides information about the student's work, in addition to the correctness of the answer. The use of example elicitation, and not only monitoring of the solution process and final answer, provides evidence based on student work that could be analyzed automatically, thus broadening the dimensions of online assessment. Our argument about automatic assessment is relevant more broadly to the use of technology in classroom assessment; We broaden existing use of examples and contribute to the design and research of automatic analysis of tasks that can inform teachers and researchers about students' understanding of universal and existential statements.

References

Anderson, J. R. (1985). *Cognitive psychology and its implications*. San Francisco: Freeman.

Arzarello, F., & Soldano, C. (2019). Approaching proof in the classroom through the logic of inquiry. In G. Kaiser & N. Presmeg (Eds.), *Compendium for early career researchers in mathematics education* (pp. 221–243). Cham: Springer.

Arzarello, F., Bartolini Bussi, M. G., Leung, A., Mariotti, M. A., & Stevenson, I. (2012). Experimental approaches to theoretical thinking the mathematics classroom: Artefacts and pro proofs. In G. Hanna & M. de Villiers (Eds.), *Proof and proving in mathematics education: The 19th ICMI study* (Vol. 15, pp. 97–143). New York: Springer.

Buchbinder, O. & Zaslavsky, O. (2009). A framework for understanding 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 group for the psychology of mathematics education* (Vol. 2, pp. 225–232).

Buchbinder, O., & Zaslavsky, O. (2019). Strengths and inconsistencies in students' understanding of the roles of examples in proving. *The Journal of Mathematical Behavior, 53*, 129–147. https://doi.org/10.1016/j.jmathb.2018.06.010.

Buchbinder, O., Zodik, I., Ron, G., & Cook, A. L. (2017). What can you infer from this example? Applications of online, rich-media tasks for enhancing pre-service teachers' knowledge of the roles of examples in proving. In A. Leung & A. Baccaglini-Frank (Eds.), *Digital technologies in designing mathematics education tasks* (pp. 215–235). Cham: Springer.

Chazan, D. (1988). *Similarity: Exploring the understanding of a geometric concept* (Technical Report 88–15). Cambridge, MA: Educational Technology Center, Harvard Graduate School of Education. https://eric.ed.gov/?id=ED295822.

Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics, 24*(4), 359–387.

Clark-Wilson, A. (2010). Emergent pedagogies and the changing role of the teacher in the TI-Nspire Navigator-networked mathematics classroom. *ZDM - The International Journal on Mathematics Education, 42*(7), 747–761.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in education research. *Educational Researcher, 32*(1), 9–13.

Cusi, A., & Olsher, S. (2019). Task design fostering construction of limit confirming examples as means of argumentation. In U. T. Jankvist, M. Van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the eleventh congress of the European society for research in mathematics education* (CERME11, February 6–10, 2019). Utrecht: Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.

Dreyfus, T., Nardi, E., & Leikin, R. (2012). Forms of proof and proving. In G. Hanna & M. de Villiers (Eds.), *Proof and proving in mathematics education—The 19th International Commission for Mathematics Instruction study* (pp. 111–120). New York: Springer.

Healy, L., & Hoyles, C. (2002). Software tools for geometrical problem solving: Potentials and pitfalls. *International Journal of Computers for Mathematical Learning, 6*(3), 235–256.

Hoz, R. (1981). The effects of rigidity on school geometry learning. *Educational Studies in Mathematics, 12*(2), 171–190.

Jeannotte, D., & Kieran, C. (2017). A conceptual model of mathematical reasoning for school mathematics. *Educational Studies in Mathematics, 96*(1), 1–16.

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*(1–2), 55–85.

Komatsu, K., & Jones, K. (2018). Task design principles for heuristic refutation in dynamic geometry environments. *International Journal of Science and Mathematics Education, 17*(4), 801–824.

Lakatos, I. (1976). *Proofs and refutations*. Cambridge: Cambridge University Press.

Lee, A., Wong, K. L., & Leung, A. (2006). Developing learning and assessment tasks in a dynamic geometry environment. In C. Hoyles, J. B. Lagrange, L. H. Son, & N. Sinclair (Eds.), *17th study conference of the international commission on mathematical instruction* (pp. 334–341).

Mejia-Ramos, J., 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.

Mills, M. (2014). A framework for example usage in proof presentation. *Journal of Mathematical Behavior, 33*, 106–118.

Mislevy, R. J., Haertel, G., Riconscente, M., Rutstein, D. W., & Ziker, C. (2017). *Assessing model-based reasoning using evidence-centered design: A suite of research-based design patterns*. New York: Springer.

Mithalal, J., & Balacheff, N. (2019). The instrumental deconstruction as a link between drawing and geometrical figure. *Educational Studies in Mathematics, 100*(2), 161–176.

Nagari Haddif, G. & Yerushalmy, M. (2015). Digital interactive assessment in mathematics: The case of construction e-tasks. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the ninth congress of the European society for research in mathematics education* (CERME9, 4–8 February 2015) (pp. 2501–2508). Prague: Charles University in Prague, Faculty of Education and ERME.

Olsher, S. & Yerushalmy, M., (2017). What can we learn about mathematics multiple choice problems from attached supporting examples? In T. Dooley, & G. Gueudet (Eds.), *Proceedings of the tenth congress of the European society for research in mathematics education* (CERME10, February 1–5, 2017) (pp. 2437–2445). Dublin: DCU Institute of Education & ERME (**hal-01942129**).

Olsher, S., Yerushalmy, M., & Chazan, D. (2016). How might the use of technology in formative assessment support changes in mathematics teaching? *For the Learning of Mathematics, 36*(3), 11–18.

Sangwin, C., Cazes, C., Lee, A., & Wong, K. L. (2010). Micro-level automatic assessment supported by digital technologies. In C. Hoyles, & J. B. Lagrange (Eds.), *Mathematics education and technology—Rethinking the terrain. The 17th ICMI study.* (Vol. 13, New ICMI Study Series, pp. 227–250). New York: Springer. https://doi.org/10.1007/978-1-4419-0146-0_10

Sinclair, N., Watson, A., Zazkis, R., & Mason, J. (2011). The structuring of personal example spaces. *The Journal of Mathematical Behavior, 30*(4), 291–303.

Stylianides, G. J. (2008). An analytic framework of reasoning-and-proving. *For the Learning of Mathematics, 28*(1), 9–16.

Stylianides, G. J., & Stylianides, A. J. (2010). Mathematics for teaching: A form of applied mathematics. *Teaching and Teacher Education, 26*(2), 161–172.

Watson, A., & Mason, J. (2005). *Mathematics as a constructive activity: Learners generating examples*. Mahwah, NJ: Lawrence Erlbaum Associates.

Yerushalmy, M. (1993). Generalizations in geometry. In J. L. Schwartz, M. Yerushalmy, & B. Wilson (Eds.), *The geometric supposer: What it is a case of?* (pp. 57–84). Mahwah, NJ: Erlbaum.

Yerushalmy, M., Nagari-Haddif, G., & Olsher, S. (2017). Design of tasks for online assessment that supports understanding of students' conceptions. *ZDM Mathematics Education, 49*(5), 701–716. https://doi.org/10.1007/s11858-017-0871-7.

Zaslavsky, O., & Zodik, I. (2014). Example-generation as indicator and catalyst of mathematical and pedagogical understandings. In Y. Li, E. A. Silver, & S. Li (Eds.), *Transforming mathematics instruction* (pp. 525–546). Cham: Springer.

Zazkis, D., & Zazkis, R. (2016). Prospective teachers' conceptions of proof comprehension: Revisiting a proof of the Pythagorean theorem. *International Journal of Science and Mathematics Education, 14*(4), 777–803.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.