ORIGINAL ARTICLE

Reasoning and proof in geometry: effects of a learning environment based on heuristic worked-out examples

**Kristina Maria Reiss · Aiso Heinze · Alexander Renkl ·
Christian Groß**

Abstract We analyze heuristic worked-out examples as a tool for learning argumentation and proof. Their use in the mathematics classroom was motivated by findings on traditional worked-out examples, which turned out to be efficient for learning algorithmic problem solving. The basic idea of heuristic worked-out examples is that they encourage explorative processes and thus reflect explicitly different phases while performing a proof. We tested the hypotheses that teaching with heuristic examples is more effective than usual classroom instruction in an experimental classroom study with 243 grade 8 students. The results suggest that heuristic worked-out examples were more effective than the usual mathematics instruction. In particular, students with an insufficient understanding of proof were able to benefit from this learning environment.

K. M. Reiss (✉)
Mathematisches Institut, Ludwig-Maximilians-Universität München, Theresienstrasse 39, 80333 Munich, Germany
e-mail: kristina.reiss@math.lmu.de

A. Heinze
Abteilung Didaktik der Mathematik, Leibniz-Institut für die Pädagogik der Naturwissenschaften, Olshausenstrasse 62, 24098 Kiel, Germany

A. Renkl
Institut für Psychologie, Universität Freiburg, Engelbergerstrasse 41, 79085 Freiburg, Germany

C. Groß
Mathematisches Institut, Universität Augsburg, Universitätsstrasse 10, 86159 Augsburg, Germany

1 Proof competencies in the mathematics classroom

Enhancing students' abilities to reason correctly and argue coherently is regarded an important aim of instruction at school. Reasoning and argumentation skills are important for many different domains and play a special role in mathematics. Accordingly, reasoning, argumentation, and mathematical proof should be integrated in the mathematics classroom through all grades (National Council of Teachers of Mathematics, 2000). However, many students face serious difficulties with consistent reasoning and argumentation, and in particular with mathematical proof (e.g., Balacheff, 1982; Harel & Sowder, 1998; Healy & Hoyles, 1998; Reiss, Klieme, & Heinze, 2001).

In the last decades research in mathematics education has analyzed the field of mathematical argumentation, reasoning and proof from various perspectives. In particular, students' problems with learning mathematical proof lead to a profound investigation how the concepts of argumentation, reasoning and proof in mathematics might be distinguished. This conceptual distinction is important in order to discuss possible implications for the teaching and learning of mathematics. For example, Hanna and de Villiers (2008) define argumentation as "a reasoned discourse that is not necessarily deductive but uses arguments of plausibility". They discuss two different viewpoints in the mathematics education community, one group regarding argumentation and proof as a dichotomy and another group seeing both as two poles of a continuum. They argue that these viewpoints are related to specific didactical implications. In the first case argumentation can be seen as

an epistemological obstacle to the learning of mathematical proof (cf. Balacheff, 1999) and the teaching of proof should focus mainly on the logical organization of statements in a proof and on a conceptual framework that builds proof independent of problem solving. The second group would focus primarily on the production of arguments in the context of problem solving, experimentation and exploration, but would expect these arguments to be logically organized only later in order to form a valid mathematical proof (Hanna & de Villiers, 2008). For our own research we take the second viewpoint and postulate that an extension or transformation of students' argumentation competency to mathematical proof competency by mathematical instruction is possible. We consider mathematical proof as the combination of reasoning, which is the ability to think logically, and argumentation, which is the ability to deduce propositions from given arguments. Nevertheless, we think that students' need a conceptual framework which help them to understand the specific character of a proof in mathematics.

Empirical research indicates that the ability to argue in a mathematically correct way and to generate a proof depends on certain prerequisites, including the knowledge of mathematical concepts and heuristic strategies, their application in a problem situation, the use of metacognitive control strategies, as well as an adequate understanding of the nature of proof in mathematics (e.g., Schoenfeld, 1983). Several empirical studies from different countries and cultures suggest that many students lack one or more of these facets of proof competency. Healy and Hoyles (1998) report on a survey with high-achieving grade-10 students. The authors identified difficulties of these students with the use of implications and elaborate that many of them rather prefer to rely on empirical-inductive arguments. Reiss, Hellmich, and Reiss (2002) as well as Klieme, Reiss, and Heinze (2003) show similar results for students from grades 7 and 8 and students from grades 12 and 13, respectively. Many students approach a proof task by searching for empirical evidence, for example by analyzing one or two examples or, particularly in geometry, by measuring angles and lines. Sometimes they use case-based reasoning that can encompass adequate ideas for a proof. However, most students particularly have difficulties bridging the gap between inductive and deductive reasoning in mathematics. They lack strategies that enable them to identify mathematical arguments supporting their empirical ideas and to generate mathematical evidence.

Using empirical arguments like measuring or generalizing from a few examples might be a deficit, which is typical for students from western countries. Lin (2000) argues that students from Asian countries generally are encouraged to use deductive arguments from the very beginning of a proving task. In these countries the teachers might take the viewpoint that argumentation and proof are

different aspects (in the sense of Hanna & de Villiers, 2008, see above) and their teaching might accordingly focus on a conceptual framework of proof independent from problem solving. Thus, students' arguments may include errors but they can be corrected by using logical reasoning and will probably provide an easier way to the valid proof, whereas empirical arguments are mostly inappropriate in order to identify the steps of a proof. Harel and Sowder (1998) argue in a similar direction and suggest that students might have an inadequate understanding of the nature of mathematical proof. According to them, even university students will not necessarily be able to establish an adequate scheme for mathematical proof. Their performance in proof tasks is probably not exclusively based on an empirical understanding; however, an empirical proof scheme and other types of proof schemes might coexist.

Studies on proof and argumentation do not only show that students have difficulties in proving, but provide evidence what the nature of proving competency is. Moreover they reveal significant interindividual differences in proving competencies. For example, Reiss, Hellmich, and Thomas (2002) asked seventh grade students to solve mathematical problems in the context of proving. They allocated their test items to three levels demanding (1) basic knowledge of facts (level I problems; e.g., calculating the size of the third angle if two angles of a triangle are given), (2) proof by making use of a single deduction (level II problems; e.g., proving that opposite angles are identical), and (3) complex proof involving multiple deductive steps (level III problems; e.g., proving that the angles of a triangle add up to 180°). They found that the proving competencies of the students could be clearly analyzed with respect to these three levels. In particular, low-achieving students (with respect to the test score) were hardly able to deal successfully with the highly complex items on level III whereas high-achieving students (with respect to the test score) performed well on level I and level II items and satisfactorily on level III tasks. Furthermore, there were significant achievement differences between classrooms despite the fact that they used identical curricula, were situated in similar social surroundings, and recruited students with comparable prerequisites (Reiss, Hellmich, & Reiss, 2002; Heinze, Reiss, & Rudolph, 2005). It can be assumed that not only individual characteristics but also the classroom has a strong impact on a student's learning. Unfortunately, there is little evidence available about the specific factors that contribute to these classroom differences in proving competencies.

2 The process of proving in mathematics

In order to design an effective learning environment for mathematical proof, i.e. a classroom-based teaching-

learning-framework for mathematical proof which offers well organized learning opportunities to the students, it is essential to characterize not only the expected result (the mathematical proof), but also the process of getting to this result (mathematical proving). The process of proving a theorem can take a long time and may include a sudden progress as well as unexpected setbacks. The example of Fermat's last theorem has demonstrated this to a broad public (Singh, 1997). The process of proving may comprise various attempts. Although the final proof consists of coherent arguments organized in a deductive chain, such a published version will hardly reflect its generation. Similarly, proofs in textbooks are also presented as a sequence of consistent arguments that provide directly logical evidence for a proposition. However, and this is important to note, this does not show how even successful students work out proofs.

Being aware of the iterative character of performing a proof, mathematics educators generally argue that the teaching and learning of proof should not be restricted to the presentation of a correct result but should emphasize the procedural aspects of proving. It is well known through a number of reports from mathematicians how this process might work (e.g., van der Waerden, 1954; Wertheimer, 1945; cf. Reiss & Törner, 2007, for an overview). In particular, mathematicians stress that proving is a process in which not only deductive reasoning but also exploration plays a dominant role (Pólya, 1957). The iterative nature of proving may be regarded as basis for an expert model of mathematical proof presented by Boero (1999). In order to differentiate between process and outcome of proving, Boero distinguishes different phases and gives insight into the combination of explorative empirical-inductive and hypothetical-deductive steps during the generation of a mathematical proof.

The first phase described in this model is (1) the production of a conjecture. This includes the exploration of a problem leading to a conjecture as well as the identification of arguments to support its evidence. Boero refers to this stage as "the private side of mathematicians' work." This work will not be publicly shared with the mathematics community but can obviously be based on discussions with other mathematicians. (2) The formulation of the statement according to shared textual conventions defines the second phase. This phase aims at providing a precisely formulated conjecture as a basis for all further activities. It may be revised in the forthcoming processes but this revision would have consequences for most activities performed by the mathematician. The third phase combines (3) the exploration of the (precisely stated) conjecture, the identification of appropriate mathematical arguments for its validation, and the generation of a rough proof idea. This is also part of the "private work" since exploration might, for

example, lead to errors or at least to preliminary formulations within the proof. Only the following three phases are subject to public communication. They include (4) the selection and combination of coherent arguments in a deductive chain, (5) the organization of these arguments according to mathematical standards, and sometimes (6) the proposal of a formal proof (see Reiss & Renkl, 2002, for an example in an educational context). Boero's model describes an expert's proving process, but it might also be adequate as a model for learning to prove. The first four phases of the model are regarded particularly important for learners as they describe the process of finding a solution and seeking evidence that it is correct.

It seems obvious that performing this process in the different phases depends on certain prerequisites concerning the knowledge of mathematical facts and procedures. Students who learn how to prove might lack this knowledge and might need specific help concerning the facts and procedures involved. In addition, it seems plausible to make the students aware of the proof process and its different phases in order to support their learning of proof. Thus, learning environments for mathematical proof should take into account both aspects, teaching the relevant mathematical content in which the proof problems under consideration are situated and encouraging processes of exploration.

3 Learning from self-explaining worked-out examples and heuristic worked-out examples

In recent years, worked-out examples have received increasing attention from psychologists (Zhu & Simon, 1987; Carroll, 1994). Worked-out examples consist of a problem and its detailed solution. In particular, worked-out examples present the algorithmic steps towards the solution of a problem. Research results show that in the beginning of the learning process on a topic, worked-out examples lead to higher learning gains than other forms of instruction, particularly in well-structured domains such as mathematics (for an overview see Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 2002; Sweller, van Merriënboer, & Paas, 1998). The advantage of worked-out examples is usually explained by the Cognitive Load Theory (cf. Renkl & Atkinson, 2003; Sweller et al. 1998). In regular instruction, problems are presented and the students are supposed to solve them at a very early stage of their learning process. Often, the students still lack an understanding of the underlying principles and thereby try to solve the problem by strategies such as means-ends analyses or shallow strategies such as key-word strategies (i.e., guessing what solution procedure could be adequate from surface features of problems). Such strategies may lead to a

solution of the problem at hand but do not deepen the understanding of mathematical principles and their application in problem solving. In the opposite, these strategies "occupy" cognitive resources in working memory. Thus, just few resources are left for the process of understanding and the acquisition of generalizable problem-solving schemata (Sweller, 1988, 1994). Worked-out examples support the students in focusing on a gain of understanding. They therefore foster a more adequate use of cognitive resources. Moreover, many researchers suggest that there might be positive learning effects of worked-out examples because learners prefer examples to other forms of information. Examples are easier to handle and to understand in comparison to a learning content, which is presented as a regular text (VanLehn, 1986; Recker & Pirolli, 1995; LeFevre & Dixon, 1986).

There are some characteristics of worked-out examples, which enable students to learn successfully. Atkinson et al. (2000) argue that the structure of the single worked-out example (intra-example feature) or of the set of worked-out examples (inter-example feature) may influence to what extent learners can profit from this learning environment. Beneficial intra-example features are blanks inserted in the solutions, emphasized intermediate aims, and the presentation of information in an integrated format (in contrast to the split-source format). Important inter-example features are the presentation of several worked-out examples for the same type of problem (multiple examples) and the accentuation of their common structure. Additionally, including self-explanation activities in worked-out examples is another important issue. There are individual differences how deeply worked-out examples are processed that lead to different learning outcomes (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989). Most students do not adequately self-explain the solution steps to themselves (Renkl, 1997). As a consequence, the students' self-explanations have to be fostered in order to fully exploit the potential of example-based learning (e.g., Renkl, 2002).

Thus, learning by self-explaining examples seems to be most promising even within a school context and worked-out examples have been positively evaluated for mathematics learning. However, this type of examples might not be fully adequate for learning to prove as students' own explorations are not particularly encouraged as suggested by the Boero (1999) model. This expert model indicates that the final mathematical proof as solution of a proof task gives only an incomplete representation of activities performed during the proving process. Consequently, a worked-out example consisting of a problem formulation and its (perfect) solution will not reflect the solution process but simply display the product. Worked-out examples, which are supposed to foster the ability to perform mathematical proof should accordingly offer process-oriented

learning opportunities. Thus, they may lead to a deeper understanding of the heuristics used during the solution process. As methodological approach, Reiss and Renkl (2002) introduced the idea of heuristic worked-out examples, which combine results by Schoenfeld (1983) on the teaching of heuristics for problem solving and the concept of worked-out examples. Schoenfeld (1983) investigated experts' thinking processes during problem solving and found out that they used various heuristic methods. Moreover, experts were able to manage these heuristics properly. In contrast to novices who spent much time in uncontrolled exploration, experts spent most of their time analyzing the problem constraints and making sense of the problem. Schoenfeld (1983, 1985) taught students some of the heuristics used by experts and showed them how they ought to be applied in different kinds of mathematical problems. This approach, namely making heuristics explicit, was used by Reiss and Renkl (2002) in order to design heuristic worked-out examples that did not simply provide the final solution steps (like traditional worked-out examples), but heuristic strategies that guided the problem solving process and made the way to the final solution transparent for the student. Heuristic worked-out examples include characteristics of traditional worked-out examples and aspects of the heuristics that are important for the solution process. They provide some scaffolding but try to encourage the student's own activity.

Research suggests that example-based learning can be primarily recommended in the beginning of a skill acquisition process (Renkl & Atkinson, 2003; Kalyuga, Ayres, Chandler, & Sweller, 2003) when the students still lack a basic understanding and are not able to work on problems on their own. Accordingly, we assume that an intervention with heuristic worked-out examples might primarily be helpful for weaker students. The heuristic examples that have to be self-explained provide them with a model how to solve proof problems. More advanced students have already a basic understanding how to deal with proof problems. Therefore self-explaining with the help of an example how to proceed in proving may be redundant, or the provided heuristic may even interfere with the students' own strategies (cf. Renkl & Atkinson, 2003; Kalyuga et al., 2003).

4 Hypotheses

The study aimed at investigating to what extent learning mathematical proof could be fostered by explicitly encouraging students to use and self-explain heuristic worked-out examples in the mathematics classroom. The following research hypotheses were addressed.

H1 Self-explaining heuristic worked-out examples has a positive influence on students' proof competencies.

Students working with self-explaining heuristic worked-out outperform students participating in regular mathematics instruction in solving proof problems.

It is reasonable to expect an improvement of students' argumentation and proof skills after a learning sequence with heuristic worked-out examples. This approach takes into account that higher-order mathematical problem solving requires both the student's activity and sufficient scaffolding so that students do not need to fully reinvent mathematical proofs. Thus, self-explaining heuristic worked-out examples should have a positive effect compared to regular instruction, which can be characterized by a mainly teacher-guided discourse between teacher and students (cf. Baumert et al., 1997).

H2 Low-achieving students should benefit more than high-achieving students from learning with self-explaining heuristic worked-out examples.

Mathematical proving is one of the most complex tasks for students. Performing a proof requires a sound knowledge of facts as well as the ability to combine these facts in a deductive chain in order to generate new knowledge. High-achieving students with respect to these prerequisites are already on a high level of performance and might therefore profit less from a supportive example-based learning environment. They should be well aware how to integrate knowledge of facts into the process of proving. Low-achieving students working with heuristic worked-out examples will experience both facets of knowledge as relatively unfamiliar and will thus improve their knowledge of facts as well as of procedures for performing proofs.

5 Sample and methods

5.1 Sample

The sample of this field study consisted of ten 8th-grade classrooms with a total of 243 students (93 female and 150 male). Six classrooms were assigned to the experimental group (150 students) and four classrooms to the control group (93 students). The classrooms were assigned to these groups according to their results in a prior test on reasoning and proof administered at the end of grade 7. Moreover, the results of a questionnaire on interest and motivation with respect to mathematics (also administered at the end of grade 7) were taken into account for the assignment to one of the groups so that the learning prerequisites were comparable in both respects across the groups.

5.2 Procedure and instruments

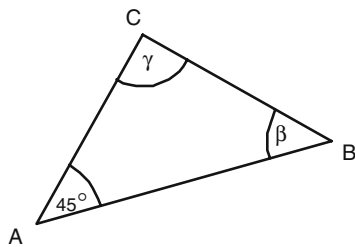All students took part in a regular teaching unit on geometrical reasoning and proof at the beginning of grade 8 (with the same teachers as in grade 7). At the end of this unit, the experimental group worked for five lessons with self-explaining heuristic worked-out examples whereas the students of the control group received instruction on proof according to the mathematics curriculum and in the way their teachers usually taught this topic (e.g., teacher guided work on proof-related exercises from the mathematics textbook, teacher guided classroom discussions on proof methods, teacher guided development of proofs in the classroom setting). Subsequently, all students took part in a posttest on reasoning and proof in geometry (closely related to the topics of the teaching unit). The students of the experimental group were asked to complete a short feedback questionnaire on their perception of those five lessons in which they used self-explaining heuristic worked-out examples.

All tests were administered to the complete classrooms and had been tried-and-tested in former studies (cf. Reiss, Hellmich, & Thomas, 2002). In particular, the mathematics pretest (13 items, 35 min processing time) and the mathematics posttest (11 items, 35 min processing time) had been scaled unidimensionally in one latent dimension by the Rasch model, based on a rating of its items in a dichotomous way as correct or incorrect (Reiss, Hellmich, & Reiss, 2002). Since both tests consisted of only a small number of items we decided to conduct a classical statistical data analysis. For this purpose the students' solutions for each item were categorized by a bottom-up analysis and these categories scored by 0 points (incorrect, no response), 1 point (correct with minor mistake or minor gap) or 2 points (correct). The scoring was comparatively liberal in the sense that we were interested in students' competency in argumentation and proof and not in their ability to write stylistic elegant sentences.

The items of both tests could be assigned to three levels of competency. The first level encompassed items that required the knowledge of facts and simple applications. The second level was characterized by simple proofs using a single deductive conclusion from facts. Level three items were proof tasks that needed more than one deduction for a correct solution (see Fig. 1 for examples of test items at the various levels).
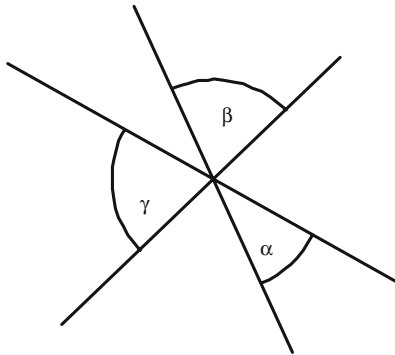
The geometry pretest and posttest comprised different items. In particular, the pretest considered the grade 7 geometry curriculum whereas the posttest took into account aspects of the grade 8 geometry curriculum. There were some items that were used in the pretest as well as in the posttest. Moreover, the posttest was closely related to the teaching unit.

The questionnaire on interest and motivation regarding mathematics consisted of different scales concerning interest in mathematics. It was adapted from a more comprehensive questionnaire (Götz, Pekrun, Perry, & Hladkyi, 2001).

This triangle is isosceles with |AC| = |BC|. Calculate the size of β and γ.
Competency level 1

Show that α + β + γ = 180°.
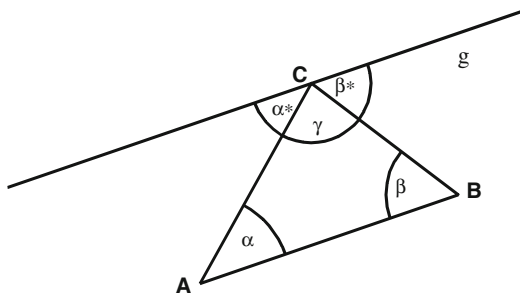
Competency level 2

Why is α = α * and β = β *, if g is parallel to AB?

Show that the angles of a triangle add up to 180°.

Competency level 3

Fig. 1 Test items of the posttest

the use of heuristic worked-out examples in their mathematics classrooms. The heuristic worked-out examples were given to the group, and the main ideas were presented. In particular, the teachers were asked not to intervene into the students' working processes but to take care that all students were aware of the correct proof to be performed at the end of a lesson. Learning with each heuristic worked-out example took about 75 min plus the time for homework. The heuristic worked-out examples for mathematical proof presented to the students were specifically designed for this study. They offered a complex mathematical problem and its solution with respect to the following main aspects (Groß, 2003):

1. Each heuristic worked-out example was structured according to Boero's model of the proving process. It started with an exploration of a problem situation and the identification of empirical arguments to support the evidence of a conjecture, followed by a precise formulation of the hypothesized statement. Students were supposed to explore the conjecture, to identify appropriate mathematical arguments, and to generate a rough idea of the proof. Moreover, they had to select coherent arguments, combine them in a deductive chain, and organize these arguments into a proof.
2. The heuristic worked-out examples were embedded into different stories. In each example two or three (hypothetical) students encountered a problem situation they wished to solve. Hence, the learner could follow the proving activities of the protagonists, which were accompanied and structured by explicit explanations from a meta-perspective.
3. Every heuristic worked-out example provided important geometry knowledge, which might be useful in the specific context. Thus, the students could concentrate on the proving process rather than on the recapitulation of facts.
4. Students were encouraged to perform self-explanation activities by working with integrated exercises and short texts with blanks. The students were asked to make drawings, to measure angles and sides of geometrical figures, to give their own conjectures, to complete statements, and to look back at the end of the proving process.

An important aspect was that students were asked to identify arguments leading to a solution and to combine these arguments in order to get a coherent proof. Moreover, the final proof was presented in detail at the end of each heuristic worked-out example. All examples dealt with topics from the geometry curriculum of the specific grade (e.g., the students were encouraged to prove that connecting the midpoints of the sides of a rectangle will result in a rhombus). The heuristic worked-out examples used in this

Students of the experimental group were asked to work individually on three heuristic worked-out examples during mathematics instruction in the classroom setting. They were told to self-explain the examples. During this work they were allowed to talk about their problems or progress with other students as this behavior was normally accepted in the classrooms of both experimental and control group. After finishing their work, the teacher discussed the proof and the proving process presented in the example with the students. There was a 90 minute teacher training in order to make the teachers of the experimental group familiar with

study were quite elaborate and encompassed more than ten pages (see Reiss & Renkl, 2002, and the appendix for a shortened example). As mentioned before, students in the control group classes got their regular teaching. In these classrooms, the teacher developed geometric proofs in a teacher-students-interaction at the black board in a whole class setting. Additionally, there were short phases of individual work consisting mainly of copying the blackboard notes. This means that the students learnt mathematical proof primarily by a teacher guided process of solving proof problems. This teaching style, guiding students through the development of a procedure by eliciting ideas and procedures from the class, is typical for German mathematics classroom even during proof instruction (cf. Stigler et al., 1999; Heinze & Reiss, 2004).

6 Results

The pretest on reasoning and proof in geometry had an overall mean of $M = 60.9\%$ (SD = 15.8) of the test points. The test scores did not significantly differ between experimental and control group (experimental: $M = 62.3\%$, SD = 15.5, control: $M = 58.5\%$, SD = 16.2; $t(241) = 1.86$, $p = 0.64$). A more detailed analysis of the pretest results according to the levels of proof competencies (cf. Sect. 1) reveals that the students of the experimental group solved 68.9% of the items at competency level II and 33.4% of the items at competency level III. The students of the control group solved 71.1% of the items at competency level II and 31.0% of the items at competency level III. The groups did not differ significantly in their proof competencies with respect to the pretest (level II: $t(241) = -0.52$,; $p = 0.60$; level III: $t(241) = 0.81$; $p = 0.42$).

In comparison to the pretest, the posttest included more items presupposing mathematical reasoning than items demanding a basic knowledge of concepts. Since those items were more difficult to solve, we did not necessarily expect the mean score of the posttest on reasoning and proof to be higher than the pretest mean score. This is supported by the results (posttest mean score: $M = 51.0\%$, SD = 17.9). Comparing the mean posttest scores of the experimental and the control group, there was a significant difference between experimental and control group. The experimental group scored higher in the posttest than the control group (experimental: $M = 54.2\%$, SD = 17.1, control: $M = 45.9\%$, SD = 18.0, $t(241) = 3.59$; $p < 0.001$). The effect size $d = 0.47$ indicates a medium effect. Analyzing the posttest according to the different levels of proof competencies shows that the better overall performance of the control group is due to a better performance at level II and level III items (experimental: 61.8% at level II, 30.8% at level III; control: 54.1% at level

II, 17.6 at level III). These differences between experimental and control group are significant (see Table 1). Moreover, we found the strongest effect for level III items. There was no significant difference with respect to those items requiring only basic competencies in geometry (experimental: 71.8% at level I; control: 68.1% at level I; see Table 1).

The data suggest that learning with self-explaining heuristic worked-out examples has no specific effect on the students' basic knowledge but enhances their proof competencies. In order to identify possible differences with respect to the learning gains of students varying in their pretest achievement, the sample was divided into three groups, namely a low achievement group ($0 \leq$ score ≤ 14; $N = 84$; less than 54% solved correctly), an average achievement group ($14 <$ score ≤ 17; $N = 81$; between 54 and 68% solved correctly), and a high achievement group ($17 <$ score ≤ 26; $N = 78$; more than 68% solved correctly) according to the pretest results (see Table 2).

The posttest data show that the students at the different achievement levels could not equally benefit from the learning environment. The significant gain can be primarily assigned to low-achieving and average-achieving students. For low-achieving students there is a significant difference on level II as well as on level III items (level II items: $t(82) = 2.17$; $p < 0.05$, $d = 0.48$; level III items: $t(82) = 3.27$, $p < 0.01$, $d = 0.74$). Average-achieving students from the experimental group perform significantly better only on level III items ($t(79) = 2.69$, $p < 0.01$, $d = 0.62$). There is no significant difference between high-achieving students from the experimental and the control group.[1]

It is important to note that differences between low-achieving and average-achieving students of the experimental group with respect to the pretest vanished in the posttest. Both groups attained nearly the same scores on level II and level III items (level II: 53.4 vs. 57.8%; level III: 25.5 vs. 26.7%; no significant differences). Moreover, we found no significant differences between the level III score of the high-achieving students of the control group and the low-achieving students of the experimental group (25.5–31.2%, $t(69) = 0.99$, $p = 0.326$).

As classroom differences have been described in other studies (e.g., Reiss, Hellmich, & Thomas, 2002), we tested in a post-hoc analysis whether the classroom also determined the effectiveness of this intervention. For this

---

[1] The fact that the significant differences in the posttest results were obtained only for specific student groups for specific levels indicates that results cannot be explained by "poor" instruction in the control group classes. Otherwise the differences between experimental and control groups in the achievement would be more apparent on all levels. A second point is that there were the same teachers in grade 7 and 8 and the grade 7 results were similar for both groups.

Table 1 Results of the posttest—parameters

	Level of competency I M (SD)	Level of competency II M (SD)	Level of competency III M (SD)
Experimental group $N = 150$	71.9 (22.9)	61.8 (23.1)	30.8 (24.7)
Control group $N = 93$	68.1 (25.2)	54.1 (26.3)	17.6 (19.1)
t value	$t(241) = 1.22$	$t(241) = 2.38$	$t(241) = 4.38$
	$p = 0.223$	$p = 0.018$	$p < 0.001$
Effect size		$d = 0.31$	$d = 0.59$

Table 2 Posttest results for different achievement groups: means and standard deviations (in brackets)

Solution percentages	Lower third		Average third		Upper third	
	Control ($N = 35$)	Experimental ($N = 49$)	Control ($N = 36$)	Experimental ($N = 45$)	Control ($N = 22$)	Experimental ($N = 56$)
Level II pretest	45.7 (37.6)	42.2 (32.6)	82.9 (20.9)	72.2 (26.6)	92.4 (13.3)	89.6 (17.0)
Level III pretest	16.8 (12.5)	14.3 (14.2)	31.9 (14.8)	32.2 (19.7)	52.3 (18.4)	51.1 (19.3)
Level II posttest	41.9* (25.7)	53.4* (22.3)	56.0 (24.9)	57.8 (20.6)	70.5 (19.9)	72.3 (22.1)
Level III posttest	11.4** (15.6)	25.5** (21.8)	15.3** (14.4)	26.7** (21.9)	31.2 (24.3)	38.6 (27.4)

$* p < 0.05, ** p < 0.01$

purpose, we focused on the experimental classrooms and we computed residual gains scores for the proof competency (taking the pretest as predictor of the posttest score). The residual gain scores were taken as dependent variables and the classroom as factor. If our intervention effects depended on the classroom, there should be a significant effect. We did not, however, find significant classroom differences with respect to the residual gain scores ($F(5) = 0.519$, $p = 0.761$). Hence, we could assure that the intervention was successfully implemented in all six classrooms.

7 Discussion

Research suggests that worked-out examples may foster learning processes in a number of different contexts; however, most studies that support learning with worked-out examples are based on relatively well-defined (algorithmic) problems. The idea of this study was to investigate whether this learning environment might be adapted to mathematical argumentation and proof, which can be regarded as a complex mathematical activity. However, there is an important obstacle in simply implementing a traditional worked-out example and this obstacle has its roots in mathematics as a subject. Traditional worked-out examples might leave the learner as a recipient of knowledge who will know that a statement is true but not why it is true. Accordingly, worked-out examples might probably

not enhance the engagement of the students in their learning as research in mathematics education presupposes that mathematics learning is heavily dependent on an active role of the learner who should take part in the process of doing mathematics. Accordingly, if worked-out examples are used in the mathematics classroom, students' cognitive activation should be well considered.

Obviously, the concept of worked-out examples has to be extended in order to better integrate process-oriented features. The idea led in this study to the use of *heuristic* worked-out examples as an instrument for learning proof in the mathematical classroom. Heuristic worked-out examples are based on traditional worked-out examples but make explicit the heuristics of the problem solving respectively proving process. For modeling mathematical proof in a heuristic worked-out example, Boero's model of proving was adapted in order to adequately reflect the problem-solving process (Boero, 1999).

Based on the positive learning effects of traditional worked-out examples in well-structured domains it was appropriate to expect better posttest results for the experimental group than for the control group and, as described above, the students of the experimental classrooms obtained significant better results. A detailed analysis of the data revealed that these positive effects could not be assigned to a gain in concept knowledge, but was due to a higher achievement of the experimental group on items of competency level II and III. Accordingly, these students were able to increase their performance level on items that

required mathematical argumentation. With respect to different achievement groups we identified a major achievement gain for low-achieving and average-achieving students. The low-achieving students improved for competency level II and III items, the average-achieving students for items on competency level III. However, there was no significant effect for high-achieving students.

Moreover, the data suggest, that the positive effect of self-explaining heuristic worked-out examples might be independent of the specific teacher. There was a gain of competency in all classrooms participating in the experiment. Probably the effects of autonomous and self-regulated learning particularly in a well-structured learning environment are still underestimated in the mathematics classrooms. However, it remains as an open question whether other forms of independent learning of the students might cause similar effects.

The results indicate that self-explaining heuristic worked-out examples are a qualified instrument for improving students' achievement on reasoning and proof in the mathematics classroom. Moreover, they suggest that low-achieving and average-achieving students may take particular advantage of this learning environment. There are a number of reasons that might have caused this effect. Evidently, the learning environment has influenced students' abilities to argue in a mathematical setting. Probably the scaffold that the heuristic worked-out examples provided might have enabled the students to better understand what a mathematical proof respectively a deductive argument is. In addition, learning with heuristic worked-out examples may be regarded as activating for every single student and thus foster students' self-determined learning. It is probably this mixture of guidance through a complex process and individual learning opportunities that are appropriate for initiating robust learning processes.

The fact that high-achieving students could not benefit in a similar way from the learning environment might have an explanation in the topics introduced during instruction. The students were assigned to the achievement groups according to their pretest results. However, distinguished pretest results are linked to an appropriate understanding of mathematical argumentation and proof. We assume that heuristic worked-out examples emphasize aspects of the proving process those students were to some extent already familiar with. Possibly, the structured learning environment did not activate high-achieving students appropriately as it provided insight into a process that they already understood. Moreover, high-achieving students might have felt unchallenged and thus did not work with the material as motivated and concentrated as other students. However, the results of feedback questionnaires administered after the treatment gave no evidence for this explanation. There

were no significant differences between the achievement groups with respect to items on motivation within the treatment phase.

Theoretical arguments from both, educational psychology and mathematics education, suggested that heuristic worked-out examples might be helpful for learning mathematical proof. The data revealed that in particular students with low proving competencies were able to benefit from working with examples that emphasized the heuristic nature of proving and encouraged them to explicate the process of proving. These students showed better results in the posttest than their high-achieving counterparts. Accordingly self-explaining heuristic worked-out examples may be apt to foster students' understanding of a quite complex mathematical topic. It remains as an open question whether high-achieving students might profit from some other and probably less controlled forms of heuristic worked-out examples. In particular, it would be important to know if more challenging respectively more difficult problems would have a positive effect on these students. Moreover, it should be investigated whether and to what extent heuristic worked-out examples could be complemented by forms of instruction that provide even more openness in problem solving.

Appendix

A heuristic worked-out example: Opposing sides and angles of parallelograms (slightly shortened version)

I. The problem

Nina and Tom have drawn and measured parallelograms. In doing so, they noticed that opposing sides were always of equal length. Moreover, opposing angles were always of equal size.

Tom: "We measured so many parallelograms: We have drawn all kinds of quadrangles and always we recognized that the opposing sides were of equal length and opposing angles were of equal size. I think, it has to be like this!"

Nina: "I think you are right, but I don't know a reason. Maybe by chance, we have only drawn parallelograms for which the statement is correct? We cannot measure the angles and sides exactly. Perhaps they were only approximately of the same size."

Tom: "So let's try to prove our assumption like mathematicians would do!"

Tom and Nina try to prove the following mathematical proposition:

"In a parallelogram opposing sides are of equal length and opposing angles are commensurate!"

In the following we have a look at how they solved the mathematical problem. Please read their solution, but try to complete all steps on your own.

II. Examination of the Problem

First we want to reproduce the measurement results of Tom and Nina. You will need a set square, paper, and pencils.

(a) Draw a parallelogram ABCD and mark the angles with α, β, γ, δ. Afterwards measure and note the sizes of its sides and angles.

(b) The experiments suggest that opposing sides and angles are of equal size in all parallelograms. You may remember that this characteristic is called congruence. In 7th grade you learned that congruent sides and angles can be transformed on one another by using congruency mappings.

Nina and Tom remember some facts:

Nina: "When did we hear of angles and sides that have the same size?"

Tom: "In the 7th grade."

Nina: "Yes, when we learned about congruency mappings."

What kind of congruency mappings do you remember?
Answer: ——

(c) The pictures on this page display several congruency mappings. Figure out what kind of congruency mappings are displayed and mark the congruent sides of the triangles.

II – c) The pictures on this page display several congruency mappings. Figure out what kind of congruency mappings are displayed and mark the congruent sides of the triangles.
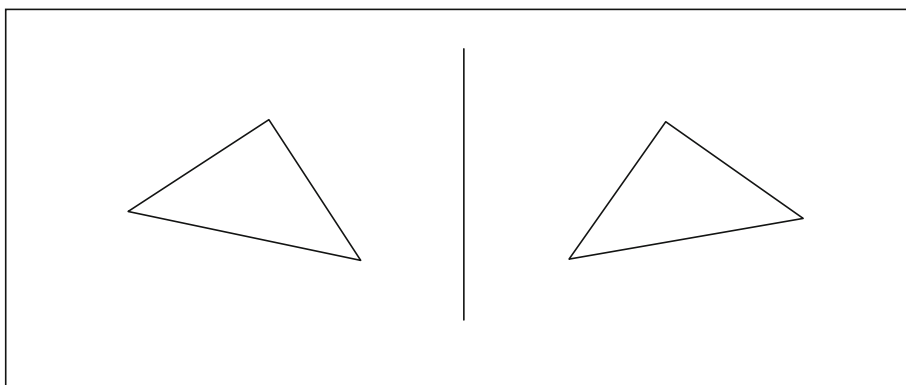
Congruency mapping: _____
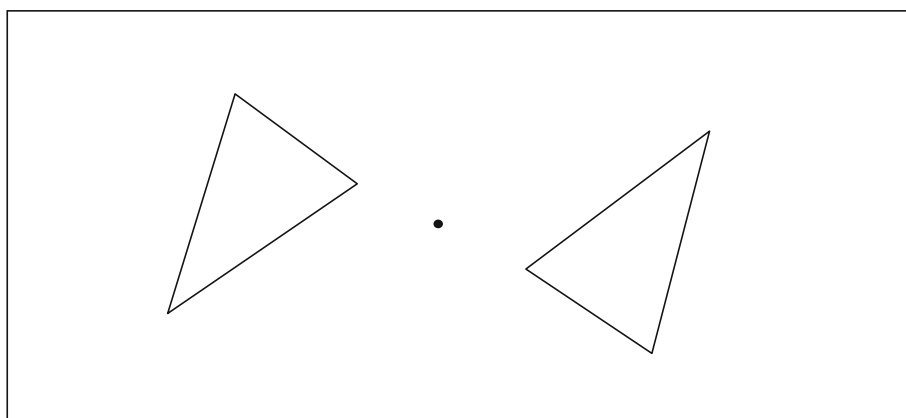
Congruency mapping: _____

Nina and Tom think about using the properties of congruency mappings for parallelograms.

Tom: "So far, by using congruency mappings we got new parallelograms."

Nina: "Now, what can we do with the congruency mappings?"

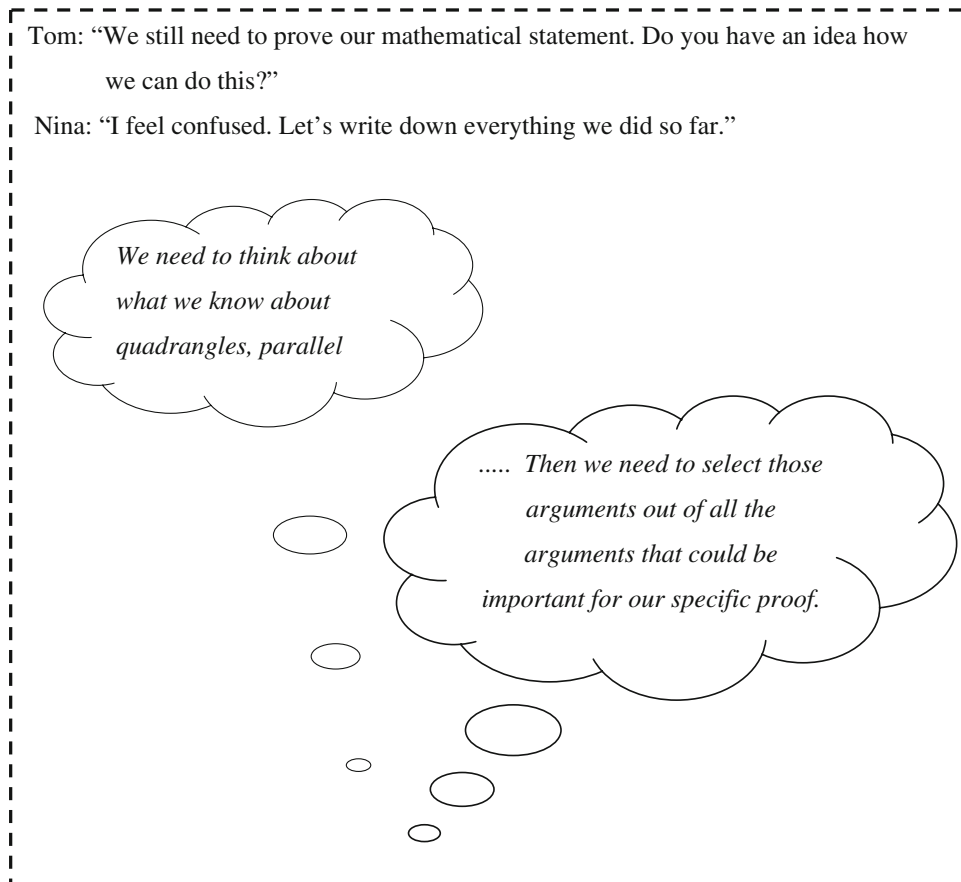Tom: "We could transform all the figures in congruent figures."

We know that a parallelogram is a ——, in which the —— sides are parallel.

We claim:

If A, B, C, D are the—of a parallelogram and α, β, γ, δ are its——, then you can say that:

— = —, — = —, — = —, — = — .

Compare your statement to the statement of Tom and Nina on the first page.

Tom: "We still need to prove our mathematical statement. Do you have an idea how we can do this?"

Nina: "I feel confused. Let's write down everything we did so far."

We need to think about what we know about quadrangles, parallel

..... Then we need to select those arguments out of all the arguments that could be important for our specific proof.

Nina: "But we don't want to construct new parallelograms. Rather we want to demonstrate that opposing sides and angles are of equal sizes. Therefore we need to transform the parallelogram on itself."

Try to detect all the symmetry axes, rotation centers, and the center of the point of reflection of the parallelogram and mark them in a figure.

III. Statement

So far it seems that Tom's and Nina's statement was correct. Try to rephrase this statement in a formula:

Prove what you wrote by filling in the gaps in the following text:

IV. What do you know about quadrangles, parallel straight lines and congruency transformations

There are a lot of arguments that could be important for the proof of the conjecture. In particular, please try to remember the following facts:

- Congruency mappings map
  straight lines on ——,
  circles on ——,
  sections on ——with equal ——,
  angles on —— of equal ——.
- The point of reflection is a rotation of —— degrees.
- The sum of angles in a quadrangle always is —— degrees.

- In a point of reflection every point lies on a ——
  with its image point and the ——.
  The center is exactly ——
  between —— and ——
  ——.
- The point of reflection maps straight lines on——
- With the line of reflection straight lines that are orthographic to the axis of reflection are mapped on the ——
- With the line of reflection straight lines that are orthographic to the axis of reflection are mapped on the ——
- Straight lines that are parallel to the axis of the reflection are mapped on the ————.

V The Proof Idea

The length of a side and the size of an angle remain unchanged when congruency mappings are applied. Thus, if we find a mapping, which maps each side of a parallelogram on the opposing side and each angle on the opposing angle we will know that they are of equal size. Have another look on section (d):

Where would the point of symmetry or the axis of reflection have to be?

Try to extend your proof idea to a proof.

VI. Proving the Conjecture

ABCD is an arbitrary parallelogram with the angles α, β, γ, δ. We draw the diagonal [AC] and call the middle of it M, as in the figure.
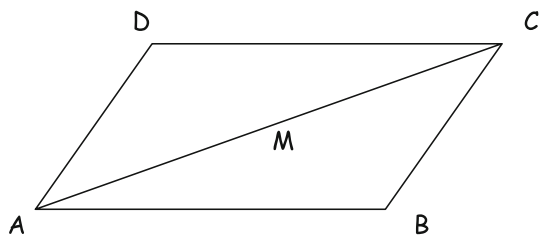
Then A is mapped with a —— on M to A' = —— and C to C' = ——, because M is the—— between A and C. The straight-line AB is mapped to a —— straight-line, which goes trough A' = ____, so to _____. Also the straight-line BC is mapped to a —— straight-line through C' = ——, so to ——.

Now we find B', the image point of B. To do this we use the fact that B is the point of intersection of the straight-lines —— and ——. The image point of B has to be the point of intersection of both image lines, that means the point of intersection of—— and ——. This point of intersection is —, that is why B' = ——.

In the same way we conclude that D' = ——. So the —— on M maps the parallelogram to ——. This means that the section [AB] is mapped to ——, and the section [BC] is mapped to ——. Because the —— maps the sections to ——, it can be concluded that —— = —— and —— = ——.

Furthermore it follows that α' = —— and β' = ——. Because the —— maps angles to —— of —— size, we conclude that —— = —— and —— = ——.

References

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. W. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*, 181–214.

Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller O., & Neubrand, J. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde.* Opladen: Leske + Budrich.

Balacheff, N. (1982). Preuve et demonstration en mathematiques au college. *Recherche de Didactique Math, 3*(3), 261–304.

Balacheff, N. (1999). Is argumentation an obstacle? *International newsletter on the teaching and learning of mathematical proof (5/6)*. Retrieved March 2008 from http://www.lettredelapreuve.it

Boero, P. (1999). Argumentation and mathematical proof: A complex, productive, unavoidable relationship in mathematics and mathematics education. *International newsletter on the teaching and learning of mathematical proof, 7/8.* http://www.lettredelapreuve.it/Newsletter/990708Theme/990708ThemeUK.html

Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology, 86*, 360–367. doi:10.1037/0022-0663.86.3.360.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145–182.

Götz, T., Pekrun, R., Perry, R. P., & Hladkyi, S. (2001). *Academic emotions questionnaire: Codebook for English-language scale versions (Technical report)*. University of Munich: Department of Psychology.

Groß, C. (2003). *Beweisen lernen mit heuristischen Lösungsbeispielen. In W. Henn (Hrsg.) Beiträge zum Mathematikunterricht 2003 (S. 257–260)*. Hildesheim: Franzbecker.

Hanna, G., & de Villiers, M. (2008). ICMI Study 19: Proof and proving in mathematics education. *ZDM-International Journal on Mathematics Education, 40*(2), pp. tba (online first).

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. H. Schoenfeld, J. Kaput & E. Dubinsky (Eds.), *Research in collegiate mathematics education III* (pp. 234–283). Providence, RI: American Mathematical Society.

Healy, L., & Hoyles, C. (1998). *Justifying and proving in school mathematics*. University of London, Institute of Education: Technical report on the nationwide survey. London.

Heinze, A., & Reiss, K. (2004). The teaching of proof at lower secondary level—a video study. *ZDM International Journal on Mathematics Education, 36*(3), 98–104.

Heinze, A., Reiss, K., & Rudoph, F. (2005). Mathematics achievement and interest in mathematics from a differential perspective.

ZDM. Zentralblatt für Didaktik der Mathematik, 37(3), 212–220. doi:10.1007/s11858-005-0011-7

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31. doi:10.1207/S15326985EP3801_4.

Klieme, E., Reiss, K., & Heinze, A. (2003). Geometrical competence and understanding of proof. A study based on TIMSS items. In F. L. Lin & J. Guo (Eds.), *Proceedings of the international conference on science and mathematics learning 2003* (pp. 60–80). Taipei (Taiwan): National Taiwan Normal University.

LeFevre, J. -A., & Dixon, P. (1986). Do written instructions need examples? *Cognition and Instruction, 3*, 1–30. doi:10.1207/s1532690xci0301_1.

Lin, F. L. (2000). An approach for developing well-tested, validated research of mathematics learning and teaching. In T. Nakahara & M. Koyama (Eds.), *Proceedings of the 24th conference of the international group for the psychology of mathematics education* (Vol. 1, pp. 84–88). Hiroshima: Hiroshima University.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Pólya, G. (1957). *(1957). How to solve it*. Princeton, NJ: Princeton University Press.

Recker, M. M., & Pirolli, P. (1995). Modelling individual differences in students' learning strategies. *Journal of the Learning Sciences, 4*, 1–38. doi:10.1207/s15327809jls0401_1.

Reiss, K., Hellmich, F., & Reiss, M. (2002a). Reasoning and proof in geometry: Prerequisites of knowledge acquisition in secondary school students. In A. D. Cockburn & E. Nardi (Eds.), *Proceedings of the 26th conference of the international group for the psychology of mathematics education* (Vol. 4, pp. 113–120). Norwich: University of East Anglia.

Reiss, K., Hellmich, F., & Thomas, J. (2002). Individuelle und schulische Bedingungsfaktoren für Argumentationen und Beweise im Mathematikunterricht. In M. Prenzel, M. & J. Doll J. (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen*. 45. Beiheft der Zeitschrift für Pädagogik (S. 51–64). Weinheim: Beltz.

Reiss, K., Klieme, E., & Heinze, A. (2001). Prerequisites for the understanding of proofs in the geometry classroom. In M. van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25th conference of the international group for the psychology of mathematics education* (Vol. 4, pp. 97–104). Utrecht: Utrecht University.

Reiss, K., & Renkl, A. (2002). Learning to prove: The idea of heuristic examples. *ZDM. Zentralblatt für Didaktik der Mathematik, 34*(1), 29–35. doi:10.1007/BF02655690

Reiss, K., & Törner, G. (2007). Problem solving in the mathematics classroom: The German perspective. *ZDM. The International Journal on Mathematics Education, 39*(5–6), 431–442.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*, 1–29.

Renkl, A. (2002). Learning from worked-out examples: Instructional explanations supplement self-explanations. *Learning and Instruction, 12*, 529–556. doi:10.1016/S0959-4752(01)00030-5.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skills acquisition: A cognitive load perspective. *Educational Psychologist, 38*, 15–22. doi:10.1207/S15326985EP3801_3.

Schoenfeld, A. (1983). Episodes and executive decisions in mathematical problem-solving. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics. Concepts and processes* (pp. 345–395). New York, NY: Academic Press.

Schoenfeld, A. (1985). *Mathematical problem-solving*. Orlando, FL: Academic Press.

Singh, S. (1997). *Fermat's last theorem*. London: Fourth Estate.

Stigler, J., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study. U.S. Department of Education*. National Center for Education Statistics, Washington, DC: U.S. Government Printing Office.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.

Sweller, J. (1994). Cognitive load, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312. doi:10.1016/0959-4752(94)90003-5.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. doi:10.1023/A:1022193728205.

VanLehn, K. (1986). Arithmetic procedures are induced from examples. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 133–180). Hillsdale, NJ: Erlbaum.

van der Waerden, B. L. (1954). *Einfall und Überlegung*. Basel: Birkhäuser.

Wertheimer, M. (1945). *Productive thinking*. New York: Harper Brothers.

Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction, 4*, 137–166. doi:10.1207/s1532690xci0403_1.