



Using heuristic worked examples to promote solving of reality-based tasks in mathematics in lower secondary school

Martin Hänze¹ · Dominik Leiss²

Received: 17 March 2021 / Accepted: 20 February 2022 / Published online: 19 April 2022
© The Author(s) 2022

Abstract

This study examined whether learning with heuristic worked examples can improve students' competency in solving reality-based tasks in mathematics (mathematical modeling competency). We randomly assigned 134 students in Grade 5 and 180 students in Grade 7 to one of three conditions: control condition (students worked on reality-based tasks), worked example condition (students studied worked examples representing a realistic process of problem-solving by fictitious students negotiating solutions to the tasks), and prompted worked example condition (students additionally received self-explanation prompts). In all three conditions, the students worked on the tasks individually and independently for 45 min. Dependent measures were mathematical modeling competency (number of adequate solution steps and strategies) and modeling-specific strategy knowledge. Results showed that although strategy knowledge could be improved through the intervention for fifth and seventh graders, modeling competency was improved only for seventh graders. The prompting of self-explanations had no additional effect for either fifth or seventh graders.

Keywords Heuristic worked examples · Mathematical word problem · Modeling competency · Mathematical problem solving

Introduction

An important goal of mathematics education is to enable learners to develop the ability to solve reality-based tasks. Accordingly, there has been an immense amount of research over the past 50 years on word-problem solving (Verschaffel et al., 2020). Whereas good problem solvers have acquired metacognitive skills that help them navigate through the complex task space (Newell & Simon, 1972) of reality-based tasks, many students fall

✉ Martin Hänze
haenze@uni-kassel.de

¹ Department of Psychology, University of Kassel, Hollaendische Str. 36-38, 34109 Kassel, Germany

² Leuphana University of Lüneburg, Lüneburg, Germany

short of this goal (De Corte et al., 2000). It remains unclear how teachers may best support their students' acquisition of the competency to solve reality-based tasks. In this paper, we advance and evaluate the idea of using a special sort of worked examples in mathematics instruction to improve students' competence in solving reality-based tasks.

The application of mathematics to solving problem situations in the real world is described in the framework of mathematical modeling (Verschaffel et al., 2002). In this paper, we first provide an overview of reality-based tasks and research on promoting mathematical problem-solving skills. Since reality-based tasks require a translation between extra-mathematical context and intra-mathematical content, we present relevant research studies on mathematical modeling. We then introduce the concept of learning using worked examples. Whereas previously the focus of research on worked examples was on well-defined problems, now several approaches attest to the efficacy of worked examples for complex and ill-defined problems, such as mathematical proving (Hilbert et al., 2008) or inquiry-based learning (Mulder et al., 2014). These worked examples were labeled *heuristic* because they combined the teaching of problem-solving heuristics by cognitive modeling (Schoenfeld, 1985) with the worked example approach (Hilbert et al., 2008).

The study presented here examines whether heuristic worked examples can be a helpful instructional strategy for supporting lower secondary school students' acquisition of the competency to solve reality-based tasks.

Reality-based tasks in mathematics

This section provides a brief and selective review of research on reality-based tasks in mathematics education. Our goal is to build on this research and extract specific steps for constructing heuristic worked examples. The competency to solve real-world problems by mathematics is a central goal of mathematical education (e.g., Common Core State Standards Initiative, 2010). Although within the mathematics research community there is no consensus on how to differentiate word problems and real-world problems (Verschaffel et al., 2020), in agreement with Verschaffel and colleagues, we view word problems as valuable “simulacra” of authentic problems one could encounter in real life. Moreover, we concede that there is a wide range of word problems with different complexity and modeling requirements (Leiss et al., 2019). In this study, we focus on word problems located in the middle of the spectrum between authentic problems and “dressed up” word problems; we call them *reality-based tasks*. We focus on the problem-solving process triggered by these tasks (see Verschaffel et al., 2020).

The research literature describes typical difficulties that students have with reality-based tasks. In an early study, Reusser (1988) reported that students often solve word problems correctly without understanding them and that the main reason for difficulties with word problems lies in a fundamental weakness of the students' epistemic control behavior. Many students do not monitor and regulate their solution attempts enough. Daroczy et al. (2015) stressed that word problems belong to the most complex and challenging problem types. They pointed out that word problems can be differentiated by linguistic factors, numerical factors, and the interaction of linguistic and numerical factors. Not only mathematical abilities but also linguistic and domain-general abilities contribute to student performance. It can be shown that these non-mathematical abilities can help students when solving reality-based problems. Though Koedinger and Nathan (2004) found that students could solve mathematical story problems better than equivalent equations, this effect was limited to simple story problems. As Koedinger et al. (2008) demonstrated, this effect was vice versa

for complex story problems. For complex problems, abstract, symbolic problem representations were easier to solve than verbal representations.

In the field of mathematics learning, there are different approaches to promoting the acquisition of competence in solving reality-based tasks. However, according to Verschaffel et al. (1999), studies show that after several years of mathematical education, many students have not yet acquired the skills needed to approach mathematical application problems efficiently and successfully. In addition to shortcomings in the domain-specific knowledge base, many learners suffer from deficits in the metacognitive aspects of mathematical competence. Verschaffel et al. concluded that the vast majority of students' attempts to solve problems show a lack of self-regulating activities such as analyzing the problem, monitoring the solution process, and evaluating its results.

In the following, we summarize studies that address this deficit by explicitly promoting a multi-stage solution process during instruction. Effective elements should then find their way into the formulation of our worked examples for solving reality-based tasks. In Verschaffel et al.'s (1999) experimental study, fifth graders were taught a model with five stages and a set of eight heuristics for solving mathematical application problems. Within the learning environment of the experimental condition, the aim was for learners to become aware of the different phases of the problem-solving process, develop an ability to monitor and evaluate their actions during the different phases of the problem-solving process, and gain mastery of the eight heuristic strategies. The five steps and heuristics of Verschaffel et al.'s "competent problem-solving model" (1999, p. 202) are the following: (step 1) students build a mental representation of the problem using heuristics such as drawing a picture, making a list, a scheme, or a table, distinguishing relevant from irrelevant data, and using real-world knowledge; (step 2) students decide how to solve the problem using one specific heuristic, such as making a flowchart, guessing and checking, looking for a pattern, or simplifying the numbers; (step 3) students execute the necessary calculations; (step 4) students interpret the outcome of step 3 and formulate an answer; (step 5) students evaluate the solution. Verschaffel et al. found that a learning environment based on the competent problem-solving model had a significant positive effect on the development of pupils' mathematical problem-solving skill compared to a control group.

In a similar attempt, Montague et al. (2011) developed a cognitive strategy instructional program (*Solve it!*) for middle school learners. Based on a framework by Mayer (1985), the program focused on the development of cognitive strategies needed for the two phases involved in mathematical word problem solving, according to Mayer: problem representation and problem solution. In Montague et al.'s cognitive strategy instruction, students were introduced to a strategic approach consisting of a sequence of seven cognitive processes (read, paraphrase, visualize, hypothesize, estimate, compute, and check). The instructional setting incorporated teaching strategies such as cueing, modeling, rehearsal, and feedback. Although the program was designed for students with learning disabilities, low-achieving students and average-achieving students were found to benefit to the same extent. They showed a much more positive development of their mathematical problem-solving skills compared to students in the control group, who received typical classroom instruction.

Self-regulated learning with this type of task is also often associated with the term *mathematical modeling*. Mathematical modeling means to express a real-world task in the language of mathematics (mathematical model) in order to solve the given problem with the help of mathematical tools (Blomhøj & Jensen, 2003). It can be defined as a complex process involving several phases (Van Dooren et al., 2006). Modelers do not move through the different phases sequentially but instead run through several modeling cycles as they gradually refine, revise, or reject the original model (Panaoura, 2012). Researchers

on mathematical modeling describe several central stages of the modeling process (for example, see Blum & Leiss, 2007; Verschaffel et al., 2000, 2014): First, a situation model is built. It contains relevant elements, relations, and conditions that are embedded in the given problem (Leiss et al., 2019). Second, a mathematical model of relevant elements, relations, and conditions available in the situation is constructed. Third, the mathematical model yields results, that, fourth, have to be interpreted within the initial problem. Fifth, the results must be evaluated (is the solution appropriate and reasonable for the problem?). Sixth, the solution must be communicated.

Some approaches have used descriptive knowledge on mathematical modeling to support learners' mathematical problem-solving. Panaoura (2012) developed a computerized approach using a cartoon animation that supports the mathematical problem-solving process by using a mathematical model. The main goals of the model were to help students divide the problem-solving procedure into stages, develop strategic problem-solving procedures, and apply the procedures during problem-solving. A focus of Panaoura's intervention program was to support students' self-reflection on their learning behavior when they encounter obstacles. The computerized intervention program supported students in Grade 5 through cartoons, interactive prompts, and questions. The program was found to enable students to recognize their strengths and limitations compared to a control group and enhance their mathematical performance.

In a study by Schukajlow et al. (2015), teachers scaffolded small groups of 15-year-old students working with reality-based tasks using an instrument called the "solution plan." The solution plan comprised four steps: (1) understanding task, (2) searching mathematics, (3) using mathematics, and (4) explaining results. The experimental group, which was scaffolded with this solution plan during the treatment phase, outperformed the control group working on the same problems without being scaffolded.

A four-step account of mathematical modeling with heuristic worked examples

Based on the research presented, we think that four critical steps in modeling emerge. These four steps are strongly linked to the four principles formulated by Polya (1957) in *How to Solve It*. In that seminal work, Polya differentiated the steps *understand the problem*, *make a plan*, *carry out the plan*, and *look back*. Our model takes up Polya's ideas but integrates further aspects from modeling research and formulates the following four steps, which correspond to the central stages of the modeling process: (1) understand the problem (i.e., read the task carefully, check for understanding, make a sketch, and note the central questions of the task), (2) mathematize it (i.e., identify relevant size specifications or estimate missing information, look for mathematical relations), (3) work mathematically (formulate and solve mathematical equations), and (4) explain the result (translate the solution into the real-world context, evaluate whether the result is correct and suited to a given situation).

The studies cited above contain a more or less abstract and explicit instruction of these solution steps. The worked-example research takes a different approach by illustrating abstract principles with concrete examples. This idea corresponds to Verschaffel et al.'s (2014) approaches on instructional support for word problems in mathematics. They assume that specific teaching methods and learner activities are crucial for teaching word problem-solving in mathematics, such as expert modeling of strategic aspects of the problem-solving process and appropriate forms of scaffolding. We suppose that learning with worked examples can be a fruitful way to develop mathematical modeling competency. In

some way, the worked example simulates the expert model and provides a scaffold for the problem-solving process. We, therefore, turn now to research on worked examples.

Worked examples present a problem, describe the problem-solving steps, and provide the correct solution. They have been shown to be superior to unguided problem-solving in terms of learning outcomes and efficiency (for reviews, see Atkinson et al., 2000; Sweller et al., 1998). Whereas most of the older studies on worked examples focused on highly structured cognitive tasks [e.g., in domains such as algebra (Carroll, 1994) or physics (Kal-yuga et al., 2001)], more recent studies have found that worked examples can also be effective for less highly structured cognitive tasks. When problems are not very well defined, such as when constructing a mathematical proof (Hilbert et al., 2008), it is usually not possible to present a procedure that leads directly to a successful solution (Kollar et al., 2014). Instead, in ill-defined domains, worked examples can demonstrate general solution approaches. Heuristic worked examples, specifically, focus on the strategic level and demonstrate heuristics for choosing adequate principles to solve a given problem (Kollar et al., 2014). In a wealth of non-algorithmic content areas, the effectiveness of worked examples has now been proven. In Hilbert et al. (2008), heuristic examples helped learners develop better conceptual knowledge about mathematical proving and proving skills than was developed by learners in a control condition focusing on mathematical contents. Mulder et al. (2014) found that heuristic worked examples improved inquiry-based learning in high school students. In Kollar et al. (2014), heuristic worked examples improved mathematical argumentation skills in university students studying mathematics teaching.

According to Renkl (2014), the effectiveness of worked examples depends on the learners' self-explanation activities. Self-explanation refers to constructive cognitive activities, whereby the learners explain the rationale of example solutions to themselves. Self-explanations can be enacted spontaneously or in response to a prompt. Bisra et al. (2018) showed in their meta-analysis that self-explanation prompts are a potentially powerful intervention across a wide range of instructional interventions. Thus, many studies integrated the prompting of self-explanations into worked examples as well. For example, Hefter et al., (2014, 2015) investigated the effectiveness of training programs that include a video-based worked example. Several prompts stimulated self-explanations. The studies confirm the excellent effectiveness of prompted worked examples in ill-defined domains, namely argumentation (Hefter et al., 2014) and epistemological understanding as a component of argumentation skills (Hefter et al., 2015). Schworm and Renkl (2007) investigated different types of self-explanation prompts in learning with worked examples aiming at the acquisition of argumentation skills. In this study, learners benefited from self-explanation prompts that focused the learner's attention on argumentation. The study of Roelle et al. (2012) showed that the combination of worked examples with self-explanation prompts fostered learning strategy acquisition. Renkl et al. (2009) pointed out the importance of self-explanation prompts for skill acquisition using heuristic worked examples; at the same time, they emphasized that the benefits of self-explanation prompts can be hindered in complex tasks if they induce too many processing demands. In their review, Dunlosky et al. (2013) rate self-explanation as having moderate utility. Moreover, they see further research needed to establish the efficacy in representative educational contexts. A meta-analysis on the relevance of self-explanations in mathematics learning (Rittle-Johnson et al., 2017) showed that scaffolding of high-quality explanations through structuring the responses is beneficial for learning. For example, learners can fill in blanks in partially complete explanations (Rittle-Johnson et al., 2017).

Within the mentioned studies with worked examples in ill-defined domains, self-explanation prompts and worked examples were often combined. Our study tried to disentangle

World's Tallest Mohawk



The Japanese fashion designer Kazuhiro Watanabe has the tallest mohawk hairdo in the world. It took him a long time to achieve his world record, since hair grows only about 0.25 cm per week. The disadvantage of his hairdo is that his arms are too short to style his hair himself. He requires a team of three hairdressers who spend 2 hours creating the hairdo each time using gel and a hair dryer.

About how long did Mr. Watanabe have to grow out his hair to achieve the current height of his mohawk? Show your work/explain your answer.

Fig. 1 Modeling task “Mohawk.”

the effects of worked examples and of explicit prompting of self-explanations. We tried to encourage spontaneous self-explanations through the design of the worked examples; moreover, we had an additional condition with explicit self-explanation prompts.

We adapted the heuristic worked example framework to reality-based tasks in mathematics. Figure 1 presents an example of an ill-defined problem used in our study. For the worked examples, we created dialogues between two fictitious students, who discuss their ideas throughout the four phases of the solving process and make realistic solution attempts (see Fig. 2). They go a little wrong sometimes, discuss their detours and mistakes, but ultimately demonstrate adequate strategies. This kind of worked example demonstrates the cognitive and metacognitive problem-solving approaches by making them obvious. This specific feature of the worked examples, wherein the worked example goes wrong sometimes, should support the active processing of the tasks and sensitize students for possible pitfalls. This innovative design element also connects to studies showing benefits from learning with erroneous worked examples. For example, Barbieri and Booth (2020) demonstrated that exposure to errors improves algebraic equation solving. As the study of Große and Renkl (2007) showed an advantage of learning with erroneous worked examples only for learners with a high level of prior knowledge, the two fictitious students in our worked examples ultimately solve the problem adequately.

Developmental aspects

In the debate on mathematical literacy, the handling of reality-based tasks and, therefore, promoting modeling competency at the lower secondary level has become more and more important. Reality-based tasks are demanding in terms of metacognitive skills: Students must not only apply cognitive strategies correctly but also learn to plan and regulate their strategy use. Schneider (2008) reviewed research on the development of metacognitive knowledge in children and adolescents and concluded that meta-cognitive knowledge develops from early primary school age on and does not peak before young adulthood. Schneider et al. (2017) differentiate between declarative, procedural, and conditional strategy

② Find the important information

As always, the two students go through the text of the word problem and check every number to see if it is needed in order to solve the problem:

Paul: "We definitely need the 0.25 cm per week. The three hairdressers are not important, and neither are the 2 hours for blow-drying. It's funny; this time there is only one important number given."

Paul highlights this piece of information in the text:

World's Tallest Mohawk



The Japanese fashion designer Kazuhiro Watanabe has the tallest mohawk hairdo in the world. It took him a long time to achieve his world record, since hair grows only about **0.25 cm per week**. The disadvantage of his hairdo is that his arms are too short to style his hair himself. He requires a team of three hairdressers who spend 2 hours creating the hairdo each time using gel and a hair dryer.

About how long did Mr. Watanabe have to grow out his hair to achieve the current height of his mohawk? Show your work.

Sara: "But we can only solve the problem if we know how long his hair is. Can we make just any assumption on that?"

Paul: "How tall is a head, then, approximately? If we knew that, we could look at the photo to see how much taller the hairdo is than his head."

Sara: "I'll just measure your head! And you could measure in the photo how much taller the hairdo is than his head."

The two students take out their rulers and measure.

Sara: "Okay; your head is about 25 cm tall."

Paul: "And in the photo his hairdo is almost exactly 5 times taller than his head. That means we get 5 times 25, or 125 cm."

They write down their estimation: **Assumption: The Mohawk hairdo is about 125 cm tall.**

Fig. 2 Second section of the heuristic worked example "Mohawk."

knowledge. They reported a substantial gain in declarative strategy knowledge during the course of lower secondary school; the growth processes assessed between Grades 5 and 9 were found to be decelerated (p. 298): During the first years of lower secondary school, students showed the most remarkable improvement in metacognitive knowledge. Declarative strategy knowledge, as well as procedural and conditional strategy knowledge, are

necessary prerequisites in order to apply strategies effectively. Based on these findings, we think that instructional support of metacognitive competencies could be particularly fruitful during the first years of lower secondary school, that is, for fifth to seventh graders.

The present study

This study aims at promoting the mathematical modeling competency of lower secondary school students through learning with heuristic worked examples. In our understanding, mathematical modeling competency is the ability to solve reality-based tasks using appropriate strategies. We developed a learning environment in which students work on reality-based tasks for one school period (45 min duration). In the control condition the students worked without instructional support; in the experimental conditions the students worked on the same tasks using the heuristic worked examples. The study design included two different worked example conditions: one with and one without additional self-explanation prompts. Preliminary studies showed that most students are able to study at least 3–4 worked examples in a 45-min period. Based on other studies on heuristic worked examples (Hilbert et al., 2008; Mulder et al., 2014) that also came up with a similar intervention duration, this amount of time and the number of problems should have the potential to produce at least initial learning effects. Our main research questions were: (1) Does students' processing of heuristic worked examples with reality-based tasks improve their mathematical modeling competency and their explicit modeling-specific strategy knowledge? and (2) Can this effect of heuristic worked examples be strengthened through self-explanation prompts?

The two research questions were separately checked for fifth and seventh graders using material adapted for their age and curriculum. That way, we could test our hypotheses in two different samples and curricular contexts, increasing the explanatory power of our findings. The aim was not to bring developmental psychological aspects to the fore – for that, the materials and tasks would have to be kept constant across the age groups. Moreover, although we primarily expected short-time effects (for a relatively immediate post-test), we also examined possible effects measured in a follow-up test several weeks after the treatment.

Method

Power analysis, sample, and design

Based on previous studies on heuristic worked examples (Hilbert et al., 2008), we expected strong effect sizes; however, we tried to reach a sample size large enough to detect medium size effects because our intervention was short. Moreover, we strengthened statistical precision by adding relevant covariates. Statistical power analysis for a three-group analysis of covariance yielded a total $N=162$ to detect an effect size of $\eta^2=.05$ (corresponds to $f=0.25$) with a reasonable power of $1 - \beta=.80$ (Shieh, 2020). The actually analyzable study sample comprised 134 students in Grade 5 ($M_{\text{age}}=11.3$ years, $SD=0.50$, 49% girls) and 180 students in Grade 7 ($M_{\text{age}}=13.4$ years, $SD=0.46$, 54% girls) at different types of lower secondary level schools in Germany.

The students were randomly assigned to one of three experimental conditions (one control group and two treatment groups) within each school class. In the *control condition*, students had to work individually on various reality-based tasks. In the *worked example condition*, students had to process heuristic worked examples that were based on the reality-based tasks used in the control group. In the *prompted worked example condition*, students learned using the heuristic worked examples along with prompts for self-explanations. In all three conditions, the time-on-task was 45 min, so although the number of processed tasks or examples could differ, all students worked on the instructional material for the same amount of time.

Instructional materials

The design was implemented in parallel for fifth and seventh graders, whereby the instructional material and the reality-based tasks in pretest and posttest were selected for fifth-grade topics (basic numeracy, plane geometry) and seventh-grade topics (linear functions, plane and spatial geometry), respectively. It was not aligned with the learners' classroom content; the study was detached from the usual teaching.

During the treatment, each student had to work on a sequence of materials. For students in the control condition, this was a series of reality-based tasks comparable to the one presented in Fig. 1. There were two treatment conditions. A series of worked examples was constructed for students in the worked examples condition based on the control group's reality-based tasks. The examples were constructed following the principles presented in Sect. *A four-step account of mathematical modeling with heuristic worked examples* above. First, each example was segmented into four sections relating to the four specific parts of the modeling process: (1) understand the problem, (2) mathematize it, (3) work mathematically, (4) explain the result. Second, the examples had varying surface characteristics but focused on a standard structure that was determined by a strategic scheme for processing modeling problems. In particular, we constructed heuristic worked examples that depicted a realistic solution process of two fictitious students cooperatively negotiating solutions to the given problems. For an example, see the second section of the heuristic worked example 'Mohawk' in Fig. 2. In the prompted worked example condition, students additionally received self-explanation prompts; here, for each solution step, students were additionally prompted to write down central solution steps along with justifications. The sequence of the reality-based tasks was held constant over the three treatment conditions.

Dependent measures (pretest)

The participants' modeling competence, basic cognitive skills, and reading ability were assessed in a pretest. It can be assumed that basic cognitive skills and reading ability are related to modeling competence since they represent essential prerequisites for processing reality-based tasks. By adding these covariates, statistical precision could be strengthened.

Modeling competency (short version). Fifth and seventh grade students worked on different sets of three reality-based mathematical tasks similar to the intervention tasks (Fig. 1). For one of the tasks, they had to answer questions in a multiple-choice format. For

example, they were asked what information provided in the task was important and needed for solving the problem. For two of the tasks, they had to give and explain their answer. We used a coding scheme with several specific items to indicate the students' use of adequate solution steps and strategies (e.g., marking numerical data, making assumptions or estimations, giving an adequate answer; for interrater agreement see posttest). Due to low item-total correlations, four items were excluded from analysis for the fifth graders; the subset of seven remaining items yielded an internal consistency of Cronbach's $\alpha = .71$. For the seventh graders all items could be used; Cronbach's α was $.73$ for 11 items.

Cognitive abilities: Students worked on the figural reasoning task from the *Berlin Test of Fluid and Crystallized Intelligence* (BEFKI 5–7; Schroeders et al., 2020). Students had to detect regularities in a sequence of geometric figures that changed in certain aspects and choose two missing figures in that sequence. Internal consistency was Cronbach's $\alpha = .78$ for the fifth graders and $.83$ for the seventh graders.

Basic reading abilities: We used a standardized test measuring reading speed (Auer et al., 2005). A list of simple sentences, adapted to the students' knowledge, had to be read as quickly as possible; for each sentence, students had to decide if it was true or false. Auer et al. (2005) reported parallel test reliability of $r_{tt} = .89$.

Dependent measures (posttest)

Modeling competency (long version): Fifth and seventh grade students worked on different sets of eight reality-based mathematical tasks; compared with the pretest, the tasks were much more demanding. For two of the tasks, they had to answer questions in a multiple-choice format (see pretest). For six of the tasks, they had to give and explain their answer in an open format. To rate the answers, we used the same coding scheme as in the pretest, with overall 45 evaluative items for all eight tasks, which indicated the quality of the answers. Interrater agreement (Cohen's κ) was above $.71$ for all single ratings indicating at least substantial strength of agreement according to the benchmarks of Landis and Koch (1977). Internal consistency was Cronbach's $\alpha = .76$ for the fifth graders and $.82$ for the seventh graders.

Modeling-specific strategy knowledge: Students in Grade 5 and Grade 7 were given the same real-world task but were not asked to solve it. Instead, they were asked to help a fictitious classmate by answering four specific questions about the solution. For example, they were asked: "Jana is overwhelmed because the word problem text contains so many numbers. What would you advise her to do? Come up with several hints." To obtain a measure of the students' modeling-specific strategy knowledge, expert raters counted how many adequate and helpful hints the students produced. Interrater agreement (Cohen's κ) was above $.71$ for all single ratings indicating at least substantial strength of agreement again according to the benchmarks of Landis and Koch (1977). Internal consistency over the four task-related questions was Cronbach's $\alpha = .68$ for the fifth graders and $.62$ for the seventh graders.

Procedure

First, the teachers, parents, and students were informed about the study. They received information about the content and the procedure of the study as well as their anonymity and data protection. It was pointed out that the students could develop their skills in solving

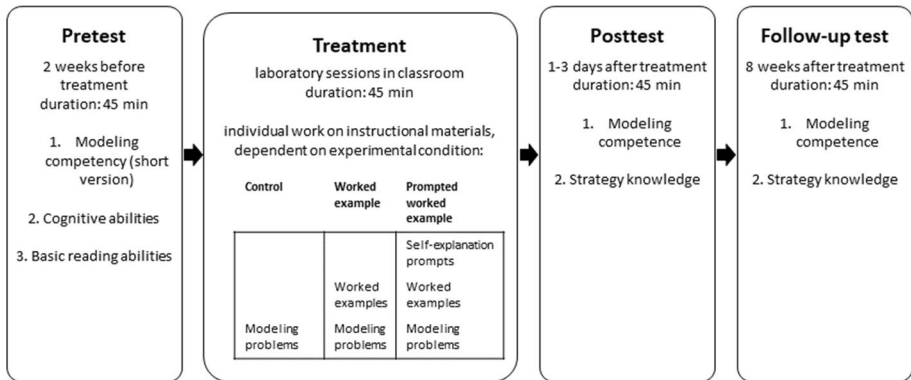


Fig. 3 Overview of the study design

application-oriented mathematics problems. We further asserted the participants that their results would not be passed on to the teachers or even be graded. Participation was part of the lessons but voluntary. All tasks were done with paper and pencil. The actual study (see Fig. 3) began with the pretest, which lasted 45 min. Approximately two weeks later, the lesson with the treatment took place, in which the students were randomly assigned to one of the three conditions. They worked individually on the corresponding tasks for about 45 min, the sequence of the reality-based tasks was held constant over the three treatment conditions. One to four days later, the posttest was conducted with a duration of 45 min as well. Students performed the follow-up test (again 45 min) approximately eight weeks later.

Results

Table 1 shows the mean scores and standard deviations for the three experimental conditions for fifth and seventh graders on all study measures. Table 2 reports correlations between all variables for fifth and seventh graders, respectively. First, we checked if the three treatment groups differed in the pretest variables. For all three pretest variables (modeling competence, cognitive abilities, basic reading abilities), there were no significant differences between the experimental conditions for the fifth graders as well as for the seventh graders (all $F < 1$, all p 's > 0.50); all values are in an expectable medium range. During the treatment, all students had 45 min; the number of completed tasks therefore varied. Students in the control condition worked on more tasks than students in the worked example and prompted worked example conditions $F(2,127) = 82.6$, $p < .01$, part. $\eta^2 = .57$ for Grade 5 and $F(2,175) = 185.4$, $p < .00$, part. $\eta^2 = .68$ for Grade 7, respectively.

In the next and central step, we checked the posttest and follow-up measures modeling competency and strategy knowledge for the treatment effects. Due to different tasks and test materials, this was done in separate analyses for fifth and seventh graders. We performed analyses of covariance for the posttest and follow-up measures modeling competency and strategy knowledge.

Table 1 Means and standard deviations for study variables

	5th graders			7th graders		
	Treatment group		Control	Treatment group		Control
	Worked examples	Prompted worked examples		Worked examples	Prompted worked examples	
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Pretest						
Modeling competency (short version) ^{a,b}	.56 (.26)	.62 (.28)	.54 (.29)	.50 (.20)	.50 (.23)	.48 (.21)
Cognitive abilities	8.62 (3.6)	8.76 (3.21)	8.8 (2.8)	9.61 (3.16)	9.73 (3.15)	9.20 (3.48)
Basic reading abilities	93.5 (16.9)	90.1 (10.4)	94.8 (17.6)	101.6 (17.0)	98.9 (15.6)	99.7 (14.6)
During treatment						
Number of completed tasks	8.9 (2.7)	5.3 (1.0)	4.0 (1.32)	9.3 (2.2)	5.5 (0.78)	4.2 (1.08)
Posttest						
Modeling competency ^b	.33 (.10)	.34 (.11)	.33 (.13)	.28 (.11)	.33 (.12)	.29 (.11)
Strategy knowledge ^c	2.15 (2.01)	3.17 (1.96)	3.08 (2.39)	3.21 (2.14)	3.95 (2.36)	4.32 (2.29)
Follow-up						
Modeling competency ^b	.31 (.11)	.33 (.11)	.31 (.12)	.28 (.12)	.33 (.13)	.28 (.11)
Strategy knowledge ^c	2.41 (2.22)	3.04 (2.37)	2.59 (2.14)	3.39 (2.42)	3.83 (2.12)	4.27 (2.35)

The sample sizes for the control, worked example, and prompted worked example groups were $n=48$, 46, and 40, respectively, for the fifth graders and $n=62$, 62, and 56 for the seventh graders.

^aThe short version (pretest) of the modeling competency test was shorter and less demanding than the posttest and follow-up versions.

^bProportion of the maximal possible number of points. The values are not comparable between 5 and 7th-grade students because we used different versions of the tests.

^cNumber of adequate and helpful hints the students produced.

Table 2 Correlations for all variables, separately for fifth and seventh graders

	Time	Variable	Grade	1	2	3	4	5	6	7	8
1	Pretest	Modeling competency	5								
			7								
2	Pretest	Cognitive abilities	5	.21*							
			7	.26**							
3	Pretest	Basic reading abilities	5	-.01	.13						
			7	.23**	.05						
4	During treatment	Number of completed tasks	5	.07	-.05	-.01					
			7	-.01	.00	.17*					
5	Posttest	Modeling competency	5	.45**	.46**	.37**	.01				
			7	.63**	.33**	.26**	-.09				
6	Posttest	Strategy knowledge	5	.10	.24**	.31**	-.17	.38**			
			7	.30**	.18*	.20**	-.16*	.40**			
7	Follow-up	Modeling competency	5	.31**	.38**	.27**	-.02	.68**	.35**		
			7	.54**	.37**	.16*	-.08	.75**	.36**		
8	Follow-up	Strategy knowledge	5	.12	.23**	.25**	-.03	.40**	.65**	.41**	
			7	.25**	.21**	.20**	-.12	.42**	.68**	.41**	

* $p < .05$; ** $p < .01$

Table 3 The proportion of variances (partial η^2) of modeling competency and strategy knowledge in posttest and follow-up measurement explained by the pretest variables (covariates)

Pretest variables	Grade	Modeling competency		Strategy knowledge	
		Posttest	Follow-up	Posttest	Follow-up
Modeling competency (short version)	5	.20**	.07**	.00	.01
	7	.33**	.24**	.06**	.03*
Cognitive abilities	5	.16**	.10**	.04*	.03*
	7	.05**	.08**	.02	.03*
Basic reading abilities	5	.17**	.07**	.10**	.06**
	7	.03*	.01	.02**	.03*

* $p < .05$; ** $p < .01$

In addition to the main effect of the treatment, we specifically tested each of the two research questions using contrast coding. The two research questions were covered by contrast coding with the contrast variables $C_{\text{worked-examples}}$ ($-1[\text{control}], 1/2[\text{worked example}], 1/2[\text{prompted worked example}]$) and C_{prompts} ($0, -1, 1$). The contrast $C_{\text{worked-examples}}$ tested research question 1, comparing the two worked example conditions with the control condition; The contrast C_{prompts} tested research question 2, comparing prompted worked examples with worked examples. Pretest scores (modeling competency, cognitive abilities, basic reading abilities) were included as covariates. We performed analyses of covariance separately for the two dependent measures (modeling competency and strategy knowledge) for posttests and follow-up measurements. For the proportion of variances explained by the covariates for all analyses, see Table 3. The condition of homogenous regression slopes was met for all covariates in all analyses (all p 's $> .05$).

For fifth graders, modeling competency was not affected by the treatment; posttest: $F(2, 128) < 1$, part. $\eta^2 < .01$; both contrasts $C_{\text{worked-examples}}$ and C_{prompts} with $p > .20$; follow-up: $F(2, 121) < 1$, part. $\eta^2 \leq .01$; both contrasts $p > .20$. Strategy knowledge was affected by the treatment at posttest measurement, $F(2, 128) = 4.18$, $p < .05$, part. $\eta^2 = .06$. Contrast $C_{\text{worked-examples}}$ showed that students in the worked examples conditions had better strategy knowledge than students in the control condition ($p < .01$) at posttest. At follow-up measurement the treatment effect was no longer significant, $F(1, 121) = 1.40$, $p > .20$, part. $\eta^2 = .02$). There was no effect of C_{prompts} neither for posttest nor for follow up measurement.

A different picture emerges for the seventh graders. The treatment affected modeling competency ($F(2, 174) = 5.10$, $p < .01$, part. $\eta^2 = .06$) and strategy knowledge ($F(2, 174) = 5.26$, $p < .01$, part. $\eta^2 = .06$) at posttest. Here, the $C_{\text{worked-examples}}$ contrasts were significant for modeling competency ($p < .05$) as well as for strategy knowledge ($p < .01$) at posttest measurement, confirming hypothesis 1. The C_{prompts} contrasts were not significant (modeling competency: $p = .07$; strategy knowledge: $p > .20$). The treatment effects faded out somewhat by the follow-up measurement (modeling competency: $F(2, 163) = 3.86$, $p < .05$, part. $\eta^2 = .05$; strategy knowledge: $F(2, 163) = 2.83$, $p = .06$, part. $\eta^2 = .03$). Regarding contrast $C_{\text{worked-examples}}$, the results indicated no effect for modeling competency ($p > .10$) but still greater use of adequate strategies after worked examples compared to control condition ($p < .05$). Regarding C_{prompts} , the results indicated contrary to expectations an

advantage of worked examples over prompted worked examples ($p < 0.05$) regarding modeling competency, but not use of adequate strategies.

Discussion

The present study showed that students' competency in solving reality-based task in mathematics, the so-called mathematical modeling competency, can be improved through worked examples showing fictitious students discussing their problem-solving approaches. In a field-experimental study, we investigated the potential of heuristic worked examples to support the development of the competency to solve reality-based tasks. For fifth graders, learning with worked examples improved their strategic knowledge. However, the improved strategy knowledge was not reflected in actual behavior: There were no experimental effects on modeling competency. Self-explanation prompts did not show any additional effects: the prompted worked examples group did not differ from the worked example condition. For the seventh graders, in contrast, there were clear effects of the worked examples. For both modeling competency and strategy knowledge, the experimental condition triggered effects of considerable magnitude. Modeling competency and strategy knowledge were improved by working with worked examples compared to solving comparable tasks without worked examples. Again, prompting self-explanations did not increase the effects for the seventh graders. Instead, modeling competency seemed to be impeded by the prompting of self-explanations.

The answers to our first research question are different for the Grade 5 and Grade 7 samples: Mathematical modeling competency is improved through worked examples, but only for seventh graders; modeling-specific strategy knowledge is improved through worked examples for fifth and seventh graders. A good part of the effects even lasts until the follow-up measurement. Regarding the second research question, we find that the overall pattern of results does not confirm differential effectiveness of the worked example conditions with vs. without self-explanation prompts. None of the comparisons between these conditions yielded significant results in the postulated way. Moreover, in one case (modeling competency for the 7th graders), we even find detrimental effects of prompting self-explanations.

Furthering the knowledge of heuristic worked examples

First, we would like to look at the importance of our results for the field of worked example research. Our study adds to the research on worked examples in non-algorithmic content areas (e.g., Hefter et al., 2014, 2015; Hilbert et al., 2008; Kollar et al., 2014; Mulder et al., 2014; Roelle et al., 2012; Schworm & Renkl, 2007). Unlike most of the studies mentioned, we tried to disentangle the effects of worked examples from those resulting from the explicit prompting of self-explanation. Beyond that, we developed and tested an innovative element. Our operationalization of the worked examples supported the active processing of the tasks. It is a dialogue between two fictitious students who discuss their ideas, go a little wrong sometimes, discuss their detours and mistakes, and ultimately demonstrate adequate strategies. Our results point to the excellent effectiveness of this dialogical approach.

Instructional support for modeling competency

Although very extensive research literature on mathematical modeling exists that discusses different modeling processes and cycles, types of modeling problems, and factors influencing modeling competency, there is relatively little evidence-based research available on instructional methods for developing and improving this competency (e.g., Stillman et al., 2017). Valuable studies are available, for example, on qualitative analyses of solution paths (Schukajlow et al., 2015), but experimental classroom studies focusing on effects on output variables are rare. Our study closes this gap and points to first steps to improve modeling competency by innovative instructional means. Working through heuristic solution examples that focus on general solutions can increase a student's general ability to solve reality-based mathematical tasks, at least in the seventh grade. The central element of the worked examples that we used was modeling in four steps: (1) understand the problem, (2) mathematize it, (3) work mathematically, and (4) explain the result. The heuristic worked examples improved students' strategy knowledge. We were thus able to show that worked examples can be used to promote not only algorithmic processing of solution paths but also metacognitive knowledge for self-regulated solving of mathematical problems.

Shortness and permanence of intervention

One might well ask how a short intervention of 45 min' duration can at all improve such a complex competency. We consider this study to be a first and encouraging approach to the development of possible instructional designs for the training of modeling competency. To create the worked examples, we used a simple four-step approach; moreover, the worked examples were from selected content domains adapted to the grade curriculum. Thus, the transfer to similar problems was facilitated. Our results reveal short-term effects that are still present about 1 to 3 days after the treatment. We certainly do not assume that the effect of the treatment is the building of comprehensive let alone sustainable modeling competency. Nevertheless, the data show that the students benefit from working through several structure-like examples for their work on similar tasks—at least in the short run. And this result adds a noteworthy and innovative element to modeling research.

Prompting self-explanation

In our study, learning with worked examples could not be improved further by the addition of self-explanation prompts. As Renkl et al. (2009) found, self-explanation prompts can only develop beneficial effects if the learners are not cognitively overwhelmed by the additional instructions. Perhaps the appealing solution examples in the form of dialogues between fictitious students provided a good potential for attentive thinking and *spontaneous* self-explanations. We suspect that the seventh graders were prone to spontaneous self-explanations, and therefore, further instructions did not show clear-cut effects. Moreover, the self-explanations prompts might have induced too many processing demands due to the complexity of the tasks (cf. Renkl et al., 2009). For example, Berthold et al. (2011) showed that under specific conditions, prompts focusing the learner's attention might have some costs for learning content that is not in the focus of the prompts. Our focus is on

conceptualizing, explaining, and on metacognitive aspects. Perhaps this results in a cognitive overload for students in the prompted worked-example condition.

Hiller et al. (2020) differentiate two perspectives on the potential effects of self-explanation prompts. Following the perspective of generative learning theory (Fiorella & Mayer, 2016), prompts activate generative learning activities. From the point of view of retrieval-based learning theory, self-explanations prompts rather support retrieval processes and stimulate episodic or contextual associations with the learning content. Like Roelle et al. (2017), learners in our study were working with the instructional material (here, worked examples) while generating self-explanations; thus, no retrieval was required. This might be an explanation for the explicit prompting of self-explanations not yielding effects. At the same time, our results do not indicate any generative learning activities stimulated by the prompts.

When linking our results to the meta-analysis of Bisra et al. (2018), we would categorize our self-explanation prompts as follows: The inducement timing was *concurrent*, the content specificity was *specific*, and the inducement format was *interrogative* and *imperative*. The prompts elicited *conceptualizations*, *explanations*, and *metacognitions*. Whereas the inducements of conceptualizations and explanations rendered medium to large effects, the prompting of metacognitions showed no significant effects. As we also prompted metacognitive processes, our findings correspond to this study.

Grade level

Due to different instructional material and different pretests and posttests, our study was not designed to check for grade or age effects in a comprehensive model. We investigated our research questions with two samples, students in Grade 5 and students in Grade 7. It is interesting to note that for the younger group of fifth graders, the heuristic worked examples did not have the desired effect on modeling competency. Although the fifth graders acquired the corresponding strategic knowledge, they were not yet able to translate it into competent and planned solutions to the problems. The task of actually applying that knowledge might have been too complex for the younger students. Studies on the development of metacognitive competencies (Schneider, 2008) point out that these abilities remain expandable up to young adulthood. Accordingly, our results give first hints that the promotion of metacognitive strategies during mathematical modeling with reality-based tasks should not be started too early.

Limitations and perspectives

Some limitations ought to be mentioned. First, due to the confounding of task material with grade level, this study is not suitable for drawing conclusions regarding the role of age development. Second, the study lacks a measure of quantity or quality of learners' self-explanation activities. Therefore, statements on the non-effects of the self-explanation prompts remain speculative. Third, although the results show short-term effects and even some long-term effects, we would need repeating and longer-lasting training or instructional interventions to find longer-lasting effects. Fourth, especially for the implementation of heuristic worked examples in the classroom, we need research on the appropriate embedding of this instructional method into teaching choreographies. Fifth, since the operationalization was carried out on the basis of concrete content areas,

the results must always be interpreted against this background. This does not rule out similar findings in other mathematical areas, but they cannot be postulated lightly. It is precisely because of the general importance of metacognitive strategies for mathematics teaching as a whole that further research is needed in this area.

Our study shows that when working on reality-based mathematical tasks, even complex skills like metacognitive strategies can be supported by a guided example-based instructional approach. Further research should focus on the question as to how mathematics instruction can integrate the student's autonomous learning with worked examples. Competency in dealing with reality-based mathematical tasks effectively can only be built up sustainably using longer-term approaches.

Funding Open Access funding enabled and organized by Projekt DEAL. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and material The data are available at <https://osf.io/w6945>.

Code availability Not applicable.

Declarations

Conflict of interest The authors have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*, 181–214. <https://doi.org/10.3102/00346543070002181>
- Auer, M., Gruber, G., Wimmer, H., & Mayringer, H. (2005). *Salzburger Lese-Screening für die Klassenstufen 5–8 [Salzburg Reading Screening Test for Grades 5–8]*. Huber.
- Barbieri, C. A., & Booth, J. L. (2020). Mistakes on display: Incorrect examples refine equation solving and algebraic feature knowledge. *Applied Cognitive Psychology, 34*, 862–878. <https://doi.org/10.1002/acp.3663>
- Berthold, K., Röder, H., Knörzer, D., Kessler, W., & Renkl, A. (2011). The double-edged effects of explanation prompts. *Computers in Human Behavior, 27*(1), 69–75. <https://doi.org/10.1016/j.chb.2010.05.025>
- Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review, 30*, 703–725. <https://doi.org/10.1007/s10648-018-9434-x>
- Blomhøj, M., & Jensen, T. H. (2003). Developing mathematical modelling competence: Conceptual clarification and educational planning. *Teaching Mathematics and Its Applications, 22*(3), 123–139. <https://doi.org/10.1093/teamat/22.3.123>
- Blum, W., & Leiss, D. (2007). How do students and teachers deal with modelling problems? In C. Haines (Ed.), *Mathematical modelling (ICTMA 12). Education, engineering and economics: Proceedings from the Twelfth International Conference on the Teaching of Mathematical Modelling and Applications* (pp. 222–231). Horwood.

- Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology*, 86(3), 360–367. <https://doi.org/10.1037/0022-0663.86.3.360>
- Common Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics*. National Governors Association Center for Best Practices and Council of Chief State School Officers. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf
- Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. -C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, 6, Article 348. <https://doi.org/10.3389/fpsyg.2015.00348>
- De Corte, E., Verschaffel, L., & Op 'T Eynde, P. (2000). Self-regulation: A characteristic and a goal of mathematics education. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 687–726). Academic Press.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, 28(4), 717–741. <https://doi.org/10.1007/s10648-015-9348-9>
- Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, 17(6), 612–634. <https://doi.org/10.1016/j.learninstruc.2007.09.008>
- Hefter, M. H., Berthold, K., Renkl, A., Riess, W., Schmid, S., & Fries, S. (2014). Effects of a training intervention to foster argumentation skills while processing conflicting scientific positions. *Instructional Science*, 42, 929–947. <https://doi.org/10.1007/s11251-014-9320-y>
- Hefter, M. H., Renkl, A., Riess, W., Schmid, S., Fries, S., & Berthold, K. (2015). Effects of a training intervention to foster precursors of evaluativist epistemological understanding and intellectual values. *Learning and Instruction*, 39, 11–22. <https://doi.org/10.1016/j.learninstruc.2015.05.002>
- Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction*, 18(1), 54–65. <https://doi.org/10.1016/j.learninstruc.2006.10.008>
- Hiller, S., Rumann, S., Berthold, K., & Roelle, J. (2020). Example-based learning: Should learners receive closed-book or open-book self-explanation prompts? *Instructional Science*, 48, 623–649. <https://doi.org/10.1007/s11251-020-09523-4>
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, 93(3), 579–588. <https://doi.org/10.1037/0022-0663.93.3.579>
- Koedinger, K. R., Alibali, M. W., & Nathan, M. J. (2008). Trade-offs between grounded and abstract representations: Evidence from algebra problem solving. *Cognitive Science*, 32, 366–397. <https://doi.org/10.1080/03640210701863933>
- Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representation on quantitative reasoning. *Journal of the Learning Sciences*, 13, 129–164. https://doi.org/10.1207/s15327809jls1302_1
- Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction*, 32, 22–36. <https://doi.org/10.1016/j.learninstruc.2014.01.003>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Leiss, D., Plath, J., & Schwippert, K. (2019). Language and mathematics: Key factors influencing the comprehension process in reality-based tasks. *Mathematical Thinking and Learning*, 21(2), 131–153. <https://doi.org/10.1080/10986065.2019.1570835>
- Mayer, R. E. (1985). Mathematical ability. In R. J. Sternberg (Ed.), *Human abilities: Information processing approach* (pp. 127–150). Freeman.
- Montague, M., Enders, C., & Dietz, S. (2011). Effects of cognitive strategy instruction on math problem solving of middle school students with learning disabilities. *Learning Disability Quarterly*, 34(4), 262–272. <https://doi.org/10.1177/0731948711421762>
- Mulder, Y. G., Lazonder, A. W., & de Jong, T. (2014). Using heuristic worked examples to promote inquiry-based learning. *Learning and Instruction*, 29, 56–64. <https://doi.org/10.1016/j.learninstruc.2013.08.001>
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall Inc.

- Panaoura, A. (2012). Improving problem solving ability in mathematics by using a mathematical model: A computerized approach. *Computers in Human Behavior*, 28(6), 291–2297. <https://doi.org/10.1016/j.chb.2012.06.036>
- Polya, G. (1957). *How to solve it*. Princeton University Press.
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38, 1–37. <https://doi.org/10.1111/cogs.12086>
- Renkl, A., Hillbert, T., & Schworm, S. (2009). Example-based learning in heuristic domains: A cognitive load theory account. *Educational Psychology Review*, 21(1), 67–78. <https://doi.org/10.1007/s10648-008-9093-4>
- Reusser, K. (1988). Problem solving beyond the logic of things: Contextual effects on understanding and solving word problems. *Instructional Science*, 17, 309–338. <https://doi.org/10.1007/BF00056219>
- Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM Mathematics Education*, 49, 599–611. <https://doi.org/10.1007/s11858-017-0834-z>
- Roelle, J., Hiller, S., Berthold, K., & Rumann, S. (2017). Example-based learning: The benefits of prompting organization before providing examples. *Learning and Instruction*, 49, 1–12. <https://doi.org/10.1016/j.learninstruc.2016.11.012>
- Roelle, J., Krüger, S., Jansen, C., & Berthold, K. (2012). The use of solved example problems for fostering strategies of self-regulated learning in journal writing. *Education Research International*, Article ID 751625. <https://doi.org/10.1155/2012/751625>
- Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education*, 2(3), 114–121. <https://doi.org/10.1111/j.1751-228X.2008.00041.x>
- Schneider, W., Lingel, K., Artelt, C., & Neuenhaus, N. (2017). Metacognitive knowledge in secondary school students: Assessment, structure, and developmental change. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Competence assessment in education: Methodology of educational measurement and assessment* (pp. 225–250). Springer.
- Schoenfeld, A. H. (1985). *Mathematical problem-solving*. Academic.
- Schroeders, U., Schipolowski, S., & Wilhelm, O. (2020). *BEFKI 5–7. Berliner Test zur Erfassung fluider und kristalliner Intelligenz (Testform 5–7)* [Berlin Test of Fluid and Crystallized Intelligence for Grades 5–7]. Hogrefe.
- Schukajlow, S., Kolter, J., & Blum, W. (2015). Scaffolding mathematical modelling with a solution plan. *ZDM Mathematics Education*, 47(7), 1241–1254. <https://doi.org/10.1007/s11858-015-0707-2>
- Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, 99(2), 285–296. <https://doi.org/10.1037/0022-0663.99.2.285>
- Shieh, G. (2020). Power analysis and sample size planning in ANCOVA designs. *Psychometrika*, 85, 101–120. <https://doi.org/10.1007/s11336-019-09692-3>
- Stillman, G. A., Blum, W., & Kaiser, G. (Eds.). (2017). *Mathematical modelling and applications: Crossing and researching boundaries in mathematics education*. Springer. <https://doi.org/10.1007/978-3-319-62968-1>
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–295. <https://doi.org/10.1023/A:1022193728205>
- Van Dooren, W., Verschaffel, L., Greer, B., & De Bock, D. (2006). Modelling for life: Developing adaptive expertise in mathematical modelling from an early age. In L. Verschaffel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds.), *Instructional psychology: Past, present, and future trends. Sixteen essays in honour of Erik de Corte* (pp. 91–109). Elsevier.
- Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H., & Ratnicks, E. (1999). Learning to solve mathematical application problems: A design experiment with fifth graders. *Mathematical Thinking and Learning*, 1(3), 195–229. https://doi.org/10.1207/s15327833mtl0103_2
- Verschaffel, L., Depaepe, F., & Van Dooren, W. (2014). Word problems in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 641–645). Springer.
- Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Swets & Zeitlinger.
- Verschaffel, L., Greer, B., & De Corte, E. (2002). Everyday knowledge and mathematical modeling of school word problems. In K. Gravemeijer, R. Lehrer, B. van Oers, & L. Verschaffel (Eds.), *Symbolizing, modeling and tool use in mathematics education* (pp. 257–276). Kluwer Academic. https://doi.org/10.1007/978-94-017-3194-2_16
- Verschaffel, L., Schukajlow, S., Star, J., & van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM Mathematics Education*, 52, 1–16. <https://doi.org/10.1007/s11858-020-01130-4>

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