



Effects of problem–example and example–problem pairs on gifted and nongifted primary school students' learning

Leonora C. Coppens¹ · Vincent Hoogerheide¹ · Elleke M. Snippe¹ · Barbara Flunger¹ · Tamara van Gog¹

Received: 27 August 2018 / Accepted: 18 March 2019 / Published online: 1 April 2019
© The Author(s) 2019

Abstract

Example-based learning (i.e., studying examples to learn a problem–solution procedure, often alternated with solving practice problems) leads to better learning outcomes than solving practice problems only, and video examples are increasingly being used in online and blended learning environments. Recent findings show that the presentation order of examples and problems affects learning: Example–problem pairs have been found to be more effective than problem–example pairs. We investigated a motivational explanation for this difference, which states that starting with a practice problem might be too difficult, causing learners to lose confidence and motivation to study. We investigated this by presenting gifted ($n=61$) and nongifted primary school students ($n=65$) with two problem–example or example–problem pairs. We hypothesized that gifted students, who generally report higher perceived competence and autonomy and higher need for cognition, would be less affected by the difficulty of starting with a problem. As expected, gifted students indeed reported higher motivation and confidence than nongifted students, and gifted students were more efficient learners. In contrast to our expectations, however, there was no difference between gifted and nongifted students in the effect of the different task sequences on test performance. Studying example–problem pairs was more efficient than studying problem–example pairs, both for gifted and nongifted students.

Keywords Example-based learning · Gifted children · Instructional design · Motivation · Primary school students

Introduction

Decades of research have shown that for novices, example-based learning leads to better learning outcomes, often achieved with less time and/or mental effort investment, than only solving practice problems (Atkinson et al. 2000; Renkl 2014; Sweller et al. 2011; Van Gog and Rumel 2010; Van Gog et al. 2019). Example-based learning entails studying examples in which the solution procedure is demonstrated, and example study is often (though not necessarily)

✉ Leonora C. Coppens
l.c.coppens@gmail.com

¹ Department of Education, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands

alternated with solving isomorphic practice problems. Many online and blended learning environments contain a mix of video examples and problem solving exercises (see e.g., www.khanacademy.org). It has been suggested, however, that the benefits of example-based learning can depend on the order in which examples and practice problems are alternated. In a study by Van Gog et al. (2011) students who learned how to troubleshoot faulty electrical circuits by means of written example–problem pairs invested less effort during the learning phase and showed better performance on a subsequent problem-solving test than students who received problem–example pairs (this replicated findings by Reisslein et al. 2006, for novice learners, and the finding that starting with example study was more effective than starting with problem solving was later replicated in other studies: Kant et al. 2017; Leppink et al. 2014).

It is still an open question *why* example–problem pairs would result in better learning outcomes with less perceived effort investment during the learning phase than problem–example pairs. After all, both approaches provide learners with a similar amount of instruction; only the order in which the instruction is provided differs. Van Gog et al. (2011) speculated on potential cognitive or motivational explanations.

Cognitive and motivational explanations for effects of example–problem sequences

First, starting with an example to study provides learners with the opportunity to acquire knowledge of how the problem should be solved, and they can then apply this knowledge during subsequent practice problem solving. As such, learners might be more successful at (and thereby reap more benefit from) attempting to solve the practice problems than learners who start with a practice problem. Second, learners might be more motivated to engage with a practice problem once they have a clue on how to solve it (cf. Sweller and Cooper 1985). Starting with a practice problem without knowing how to handle it is very difficult (forcing learners to engage in inefficient yet effortful search processes such as trial-and-error or means–ends analysis; see Sweller 1988; Sweller and Levine 1982), which may be frustrating for them. Consequently, learners might lose the confidence in their own abilities and the motivation needed for studying the example and other tasks that follow (Van Gog et al. 2011).

Tentative support that this second, ‘motivational explanation’ may be more likely than the first, ‘cognitive explanation’, comes from a study by Van Gog (2011), in which no differences between an example–problem pairs and problem–example pairs condition were found. Crucially, the task in this study consisted of learning how to solve a puzzle problem. Engaging in puzzle tasks is presumably more intrinsically motivating than engaging in troubleshooting faulty electrical circuits such as in the study of Van Gog et al. (2011). Starting with (and failing at) solving a puzzle problem is unlikely to lower learners’ confidence in their own ability (i.e., it does not threaten a learner’s academic self-concept) and instead of frustration it may even spark curiosity about the correct solution. As such, for students learning how to solve a puzzle task, starting with a practice problem might not have a negative effect on their confidence and motivation. By contrast, the traditional ‘cognitive’ explanation of the worked example effect would predict better performance in the example–problem than the problem–example condition: starting with an example allows participants to learn about the problem before attempting to solve it, which should be more effective, regardless of whether the task is a motivating puzzle task. However, the motivational explanation proposed by Van Gog et al. (2011) has not yet been tested.

Self-regulated learning research suggests that students prefer to start with a problem instead of an example, however. Foster et al. (2018) showed that when given the choice, the majority

of learners selected a problem as their first task. However, prior to their decision to study a problem instead of an example, participants had received an instructional text, which would have allowed them to acquire some knowledge of the tasks and may have given them the confidence to select a problem instead of an example. Moreover, the selection of learning materials by a learner does not necessarily reflect confidence and motivation. Furthermore, in research on productive failure (e.g., Kapur 2008), failure to solve an initial problem is under certain circumstances found to be helpful for learning (although this applies mainly to the acquisition of conceptual knowledge, whereas the main goal in research on example-based learning is typically the acquisition of procedural knowledge; see the review by Loibl et al. 2017).

Moreover, it is likely that individual differences might play a role in the effects that the order of examples and problems has on learning outcomes. For instance, need for cognition (NFC), which refers to “the tendency for an individual to engage in and enjoy thinking” (Cacioppo and Petty 1982, p. 116), might affect how students respond to the challenge posed by starting with a practice problem to solve. There is, however, very little research that has investigated the role of individual differences (other than prior knowledge) in example-based learning (for an exception, see Schwaighofer et al. 2016).

The present study

The main purpose of the present study was to examine whether presenting novice learners with example–problem pairs or problem–example pairs has differential effects on their performance, effort investment, as well as motivational outcomes (Van Gog et al. 2011). We investigate this question both in gifted and nongifted primary school students, because how children deal with a failed problem-solving attempt may also depend on their cognitive abilities. Compared to students with typical cognitive ability (henceforth denoted as nongifted students), students with high cognitive ability (henceforth denoted as gifted students) have been found to show a higher NFC (Meier et al. 2014). Because individuals with a high NFC tend to enjoy engaging in effortful cognitive activities more than students with lower NFC (Cacioppo et al. 1996), it may be that it is not giftedness or the task sequence that affects performance and other (motivational) outcomes, but that the effects depend on students’ NFC. Therefore, we explored the comparison of gifted/nongifted students concerning the effects of problem–example and example–problem pairs while controlling for students’ NFC.

Presenting students with problem–example versus example–problem pairs in problem solving tasks might also affect motivational outcomes. Specifically, gifted students have been shown to report higher competence beliefs, such as self-efficacy, than nongifted students in many contexts (Pajares 1996; Pajares and Graham 1999), also when faced with complex and difficult tasks. Thus, when letting gifted students solve tasks with problem–example pairs they should feel greater competence than non-gifted students presented with the same task sequence. Moreover, gifted students are known to be particularly motivated for tasks that allow them to have autonomy (i.e., the feeling of having the option to choose, e.g., Assor 2012) over the problem–solution approach (Clinkenbeard 2012). Presenting a problem first (instead of the worked-out solution steps) could give gifted students a stronger feeling of autonomy over the approach to take, and could enhance their enjoyment when performing the task. Thus, whereas problem–example pairs compared to example–problem pairs might impose higher cognitive load, and decrease confidence, motivation, and learning outcomes for nongifted students, problem–example pairs might not have this negative effect for gifted students, who could perceive even higher autonomy and competence when first working with a problem and then with an example than vice versa.

The main effects of giftedness and task sequence were tested in the present study via a 2 (gifted/nongifted) \times 2 (problem–example/example–problem pairs) ANOVA. Gifted and nongifted primary education students learned how to solve a math problem by means of two example–problem pairs or two problem–example pairs and were subsequently presented with a posttest consisting of isomorphic problems (i.e., different numbers but same underlying problem-solving procedure as the tasks used in the learning phase) and near transfer problems (i.e., different numbers and a slightly different underlying problem-solving procedure). We chose a complex math task to ensure that both gifted and nongifted students would not have the level of prior knowledge needed to be able to solve the first practice problem (without studying an example first). Note that we chose a sample of primary education students because the effect of example–problem sequences has not yet been tested in this age group and because questions concerning (differential) instructional strategies for gifted students are important in primary education (Morisano and Shore 2010). We measured perceived competence, perceived autonomy, and learning enjoyment as indicators of motivational aspects of learning, in order to verify the expected motivational differences between gifted and nongifted students. Furthermore, we measured NFC, performance, and self-reported mental effort investment (i.e., an indicator of experienced cognitive load; Paas 1992) in the learning phase. The combination of performance and effort investment ratings provides more insight than performance scores alone, as it can be seen as an indicator of cognitive efficiency: the same performance achieved with less effort can be considered more efficient (Van Gog and Paas 2008). To get more insight into learning progress, we measured and analyzed performance and effort on the two practice problems separately as well as together.

The traditional, ‘cognitive’ explanation of worked example effects would not predict a difference in the effect of the order of worked examples and problems between gifted and nongifted students (i.e., also gifted students’ practice problem performance would be expected to benefit from the opportunity to build a schema through worked example study before attempting to solve a practice problem). However, if the motivational explanation proposed by Van Gog et al. (2011) would hold true, then we would expect a different pattern of results for gifted students than for nongifted students. Specifically, we would expect to replicate their finding that receiving example–problem pairs would result in higher test performance (Hypothesis 1a) and require less investment of mental effort (Hypothesis 1b) than receiving problem–example pairs when learning a difficult math task in *nongifted* students, but not in *gifted* students.

In addition, we expected gifted students to report higher autonomy (Hypothesis 2a), competence (Hypothesis 2b), and enjoyment (Hypothesis 2c) on the posttest than nongifted students, especially after starting with a practice problem. Finally, we expected that gifted students would show a higher NFC than nongifted students. Because we assumed that NFC would buffer against the complexity of problem–example pairs, we also explored whether the effect of giftedness and task sequence on performance and perceived mental effort could be found when controlling for students’ NFC with an ANCOVA.

Method

Participants and design

Assuming a medium effect size of $f = .25$, a total of 128 participants were needed to reach a power of .80. Participants were 134 Dutch primary school students: 66 nongifted students

from a regular Dutch primary school and 68 gifted students from a primary school for gifted students. A requirement for admission to the school for gifted students was an IQ of 130 or higher. The regular school had no IQ requirement, so the average IQ of their students was presumably around the population average of 100. Six students (three gifted and three nongifted) did not complete the experiment and were removed from the sample. In addition, two nongifted students for whom pre-test data were unavailable due to experimenter error were removed from the sample. This resulted in a final sample of 126 students, consisting of 61 nongifted students (34 boys; $M_{\text{age}} = 10.46$, $SD = 0.98$) and 65 gifted students (35 boys; $M_{\text{age}} = 10.65$, $SD = 0.74$). Nongifted and gifted students were quasi-randomly assigned to either the example–problem pairs condition (EPEP) or problem–example pairs condition (PEPE); that is, they were matched for gender and scores on a standardized math test that is part of the regular curriculum (e.g., for each gifted girl assigned to the EPEP condition, another gifted girl with a very similar if not identical score on the math test was assigned to the PEPE condition).

Materials

Pretest

Because using the same task for the pretest and the main experiment would provide all participants with an additional problem (i.e., make the EPEP condition a PEPEP condition), the pretest consisted of a different task that had a similar problem-solving procedure to the procedure needed to solve the water jug problems in the main experiment: a paper-based Tower of Hanoi problem (see [Appendix 1](#)). In this task we asked students to ‘move’ three disks between three pegs of a different size from the initial state (all disks on the first peg, the largest disk on the bottom, and the smallest one on top) to the goal state (the same stack of disks, but on the third peg), under instructions to only move one disk at a time and to never place a larger disk on top of a smaller disk. The Tower of Hanoi task could be completed in seven steps.

Learning tasks: water jug problems and examples

The learning tasks were presented through Qualtrics (i.e., a web-based survey program; www.qualtrics.com) and consisted of water jug problems (Luchins 1942). Snapshots of the first example are shown in [Fig. 1](#). In the water jug problems, there were four water jugs (A, B, C, and the goal jug) of which the sizes were given. The task was to fill the goal jug with the required amount of water, using the other jugs, in as few steps as possible. Each jug could at any point be filled to the top or emptied completely. Pouring water from jugs A, B, or C into another jug was also allowed, but only until the jug that the water was poured from was empty or the receiving jug was full. The rules of the task were explained using a rules sheet at the beginning of the learning phase (see [Appendix 2](#)). The problems in the learning phase could be solved in six steps.

The learning phase consisted of two video modeling examples and two practice problems (with presentation order depending on assigned condition). Two video modeling examples were created (consisting of PowerPoint slides with a female voice-over) in which it was explained and demonstrated how to solve water jug task problems. First the starting situation was presented, then the problem was solved in six steps, and at the end of the video modeling example an overview of all steps was shown. Both video examples were

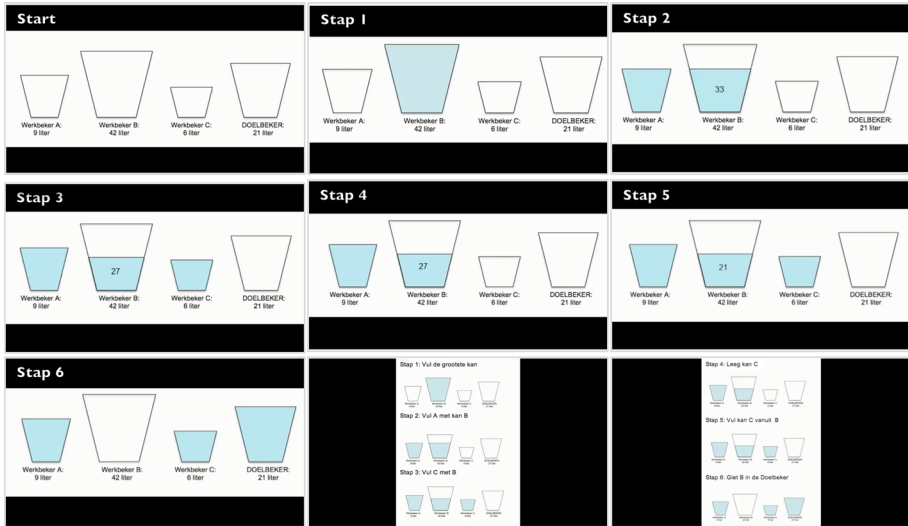


Fig. 1 Snapshots of the start-state, six problem-solving steps, and the end-state of the first video modeling example

4 min (240 s) long. The problems in the two video modeling examples and the two practice problems were isomorphic (i.e., the underlying problem structure was the same, but the values differed).

Test tasks

The posttest consisted of three problems that were isomorphic to the learning phase problems (i.e., could be solved in six steps, using the procedure demonstrated in the examples), and three near transfer problems, which required seven or eight steps to solve. Cronbach's α for the isomorphic posttest tasks was $\alpha = .81$ and $.62$ for the near transfer tasks.

Mental effort

To measure experienced cognitive load, students were asked to rate how much effort they had invested in studying each video modeling example, solving each practice problem, and solving each posttest problem, on a scale of 1 (very, very low effort) to 9 (very, very high effort) (Paas 1992).

Perceived autonomy and perceived competence

Perceived autonomy and perceived competence questionnaires were administered before and after the learning phase using a translated version of a questionnaire developed by Flunger et al. (2013). We adapted the wording of the first part of the items to measure autonomy and competence during math tasks for the pretest and to measure autonomy and

competence during the water jug tasks for the posttest (see Table 1). Both questionnaires consisted of four items rated on a scale from 1 (does not describe me at all) to 4 (describes me perfectly). The items were averaged per questionnaire to obtain the overall perceived competence and perceived autonomy scores. Cronbach's α for the perceived autonomy questionnaire was $\alpha = .64$ on the pretest and $\alpha = .68$ on the posttest. For the perceived competence questionnaire, Cronbach's α was $\alpha = .80$ on the pretest and $\alpha = .76$ on the posttest.

Learning enjoyment

After the learning phase, students rated how much they enjoyed learning from the video modeling examples and problems on a scale from 1 (not at all enjoyable) to 9 (very much enjoyable) (Hoogerheide et al. 2014).

Need for cognition

Students completed a Dutch translation of the shortened version of the NFC questionnaire (Cacioppo et al. 1984), consisting of 18 items they rated on a scale from 1 (completely disagree) to 6 (completely agree). Cronbach's α for the NFC questionnaire was $\alpha = .84$.

Procedure

Potential participants and their parents were informed of the experiment through the school newsletter which was distributed 4 weeks before the first session. In the letter, the study was described and parents could contact the researcher for any questions or if they did not want their child to participate. At the start of the experiment, participants were again informed of the nature of the experiment and signed a consent form. The study consisted of two sessions. The first session lasted around 20 min and took place in the participants' regular classroom. The experimenter first provided some general information about the nature of the study, after which students provided informed consent. Afterwards, participants were first provided with the pretest, for which they received 4 min, and then completed the (paper-based) perceived autonomy, perceived competence, and NFC questionnaires.

The second session took place 1 week later and lasted approximately 50 min. Before students arrived, an experimenter created 25 individual 'work stations' in a regular classroom, each consisting of a table, a laptop, a headset, a calculator, a pen, a piece of scrap paper, and a piece of paper containing the name of a student and a link to the Qualtrics questionnaire. Upon arrival, participants were instructed to find their work station and to log into the Qualtrics environment. The Qualtrics questionnaire provided participants with three 'blocks' of questions. The first block presented a short demographic data questionnaire, asking participants for their name, age, and educational year. At the start of the second block, participants were provided with the rules sheet, which they were instructed to study for 2 min. Afterwards, participants were presented with the two video modeling examples and the two practice problems, either in an example–problem–example–problem sequence (EPEP condition) or a problem–example–problem–example sequence (PEPE condition), depending on their assigned condition. They were instructed to study each example once. Note that participants in both conditions were provided with the exact same examples and problems, but in a different order. Participants received four min for each task, were allowed to use the rules sheet and a calculator while working on the practice problems, and completed the mental effort rating scale after each task. At the end of the second block, participants completed the learning enjoyment, perceived autonomy, and

Table 1 Items of the perceived autonomy and perceived competence questionnaire

Original item (Flunger et al. 2013)	Pretest item (translated from Dutch)	Posttest item (translated from Dutch)
Perceived autonomy		
During physics lessons, do you in general have the feeling that you can follow your own procedures of solution?	While working on math tasks, I generally have the feeling that I can follow my own procedures of solution	While working on the water jug tasks, I generally have the feeling that I can follow my own procedures of solution
During physics lessons, do you in general have the feeling that you can do the work your own way?	While working on math tasks, I generally have the feeling that I can do the work my own way	While working on the water jug tasks, I generally have the feeling that I can do the work my own way
During physics lessons, do you in general have the feeling that you can decide on your own how to work on a task?	While working on math tasks, I generally have the feeling that I can decide on my own how to work on a task	While working on the water jug tasks, I generally have the feeling that I can decide on my own how to work on a task
During physics lessons, do you in general have the feeling that you can work on new contents your own way?	While working on math tasks, I generally have the feeling that I can work on new contents my own way	While working on the water jug tasks, I generally have the feeling that I can work on new contents my own way
Perceived competence		
During physics lessons, do you in general have the feeling that you can solve the difficult tasks as well?	While working on math tasks, I generally have the feeling that I can solve the difficult tasks as well	While working on the water jug tasks, I generally have the feeling that I can solve the difficult tasks as well
During physics lessons, do you in general have the feeling that you are up to the requirements of the lessons?	While working on math tasks, I generally have the feeling that I am up to the requirements of the lessons	While working on the water jug tasks, I generally have the feeling that I am up to the requirements of the lessons
During physics lessons, do you in general have the feeling that you can successfully solve the tasks?	While working on math tasks, I generally have the feeling that I can successfully solve the tasks	While working on the water jug tasks, I generally have the feeling that I can successfully solve the tasks
During physics lessons, do you in general have the feeling that you can understand difficult learning materials as well?	While working on math tasks, I generally have the feeling that I can understand difficult learning materials as well	While working on the water jug tasks, I generally have the feeling that I can understand difficult learning materials as well

perceived competence questionnaires. Lastly, the third block contained the posttest, in which participants received 4 min per problem to solve six problems. Participants could use the rules sheet and a calculator while working on the test tasks, and rated how much mental effort they invested after each task.

Scoring of measures and inter-rater reliability

Performance on the Tower of Hanoi pretest was scored based on the model (i.e., ideal) solution, which consisted of seven steps. Two raters (first and third author) independently scored whether all steps had been performed correctly, and a high degree of inter-rater reliability was found. The average measure intra-class correlation (ICC) was .98 and Cohen's Kappa was .97. The scores of the first rater (third author) were used for the analyses.

Performance on the water jug tasks was also scored based on model (i.e., ideal) solutions. A correct solution using the fastest method was awarded 2 points. A correct solution using a less efficient method or a correct solution in which the final step was missing received 1 point. An incorrect or missing solution received 0 points. To measure the reliability of the ratings, two raters (first and third author) independently scored 38% of the water jug tasks. Because the inter-rater reliability was high (the average measure ICC was .99 and Cohen's Kappa was .95), the first rater (third author) continued to score all data and we used her scores in the analyses. The maximum performance score was 4 points on the learning tasks, 6 points on the isomorphic posttest problems, and 6 points on the near transfer posttest problems. Average perceived mental effort scores were computed for the learning tasks, the isomorphic posttest problems, and the near transfer posttest problems separately. Negatively worded items of the scales for perceived autonomy, perceived competence, and NFC questionnaires were reversed and scores were then averaged per questionnaire.

Results

To test our main research question, 2 (giftedness: nongifted vs. gifted) \times 2 (task sequence: EPEP vs. PEPE) ANOVAs were conducted on the performance, effort, and motivation outcome variables. Any significant interaction effects were followed up by independent-samples t-tests. Descriptive statistics are presented in Table 2.

Before addressing our hypotheses, we first checked whether students' pretest performance was low and did not differ among the task sequence conditions. Students indeed performed poorly on the Tower of Hanoi task: 23.5% of students completed the task successfully. Of the gifted students, 27.5% completed the task successfully, versus 19.6% of the nongifted students. A χ^2 test showed no difference in scores on the Tower of Hanoi task between gifted and nongifted students: $\chi^2(1)=0.54$, $p=.462$. Of the students in the EPEP condition, 16.7% were successful at completing the Tower of Hanoi task, versus 31.3% in the PEPE condition. This difference was also not significant: $\chi^2(1)=1.85$, $p=.173$. We explored whether performance on the pretest would be a useful covariate. Performance on the pretest was not associated with performance during the learning phase or the posttest, nor did it explain any additional variance above and beyond giftedness and task sequence. Therefore we decided not to include pretest performance as a covariate.

We checked whether there were differences between the groups in NFC. In line with our expectations, gifted students reported a higher NFC ($M=4.12$, $SD=0.65$) than nongifted students ($M=3.35$, $SD=0.60$), $F(1, 122)=48.87$, $p<.001$, $\eta_p^2=.29$. None of the

Table 2 Means (SD) of the dependent variables per condition

	Nongifted students		Gifted students	
	EPEP (<i>n</i> =30)	PEPE (<i>n</i> =31)	EPEP (<i>n</i> =33)	PEPE (<i>n</i> =32)
Learning phase performance problem I (range 0–2)	0.40 (0.77)	0.00 (0.00)	1.36 (0.86)	0.06 (0.35)
Learning phase performance problem II (range 0–2)	0.57 (0.90)	0.68 (0.95)	0.97 (1.02)	1.13 (0.94)
Posttest performance isomorphic (range 0–6)	2.33 (2.52)	3.19 (2.55)	5.06 (1.41)	5.41 (1.04)
Posttest performance near transfer (range 0–6)	0.40 (0.97)	0.84 (1.32)	2.76 (2.00)	2.28 (1.80)
Learning phase effort (range 1–9)	3.98 (1.46)	4.22 (1.52)	3.05 (1.03)	3.86 (0.77)
Posttest effort isomorphic (range 1–9)	4.10 (2.27)	3.18 (2.41)	2.67 (1.29)	2.60 (1.39)
Posttest effort near transfer (range 1–9)	6.64 (2.16)	5.62 (2.20)	5.31 (1.65)	4.90 (1.78)
Pretest perceived autonomy (range 1–4)	2.67 (0.52)	2.59 (0.49)	2.77 (0.55)	2.78 (0.45)
Pretest perceived competence (range 1–4)	2.69 (0.64)	2.83 (0.54)	2.99 (0.62)	2.89 (0.59)
Posttest perceived autonomy (range 1–4)	2.66 (0.63)	2.47 (0.54)	2.79 (0.45)	2.79 (0.48)
Posttest perceived competence (range 1–4)	2.80 (0.64)	2.62 (0.57)	3.13 (0.49)	2.78 (0.40)
Learning enjoyment (range 1–9)	6.70 (2.22)	6.97 (1.76)	7.42 (1.46)	6.56 (2.08)
Need for cognition (range 1–6)	3.25 (0.56)	3.44 (0.64)	4.24 (0.61)	4.00 (0.66)

other effects were significant [task sequence: $F(1, 122)=0.05$, $p=.818$, $\eta_p^2 < .01$; Giftedness * Task Sequence: $F(1, 122)=3.90$, $p=.050$, $\eta_p^2 = .03$; note that the last result may seem significant at first glance, in which case according to the data in Table 2 it would seem to suggest that gifted students in the EPEP condition reported a higher NFC than gifted students in the PEPE condition, whereas nongifted students in the EPEP condition reported a lower NFC than nongifted students in the PEPE condition. However, the effect is strictly speaking not significant: $p=.050435 > .05$, so one has to take care not to over-interpret these data].

Do giftedness and task sequence affect performance (Hypothesis 1a)?

For completeness, we added NFC as a covariate to the 2×2 ANOVAs on performance on the learning tasks and the isomorphic and transfer test to explore whether NFC could predict any variance of students' test performance above and beyond the variance explained by giftedness and task sequence.¹

Performance during the learning phase

Levene's test indicated that equal variances in the population could not be assumed for the total performance during the learning phase, and for the first practice problem [total: $F(3, 122)=3.03$, $p=.032$; first practice problem: $F(3, 122)=38.27$, $p < .001$; second practice

¹ We originally analyzed the data with ANOVAs first, in order to clearly separate the exploratory analysis from the confirmatory analyses. However, because the ANCOVAs are more complete and did not change the results, we followed a reviewer's suggestion and report only the ANCOVAs.

problem: $F(3, 122)=2.24, p=.087$]. However, the statistical tests were not affected by this heterogeneity of variance, because the groups were approximately equal in size (Field 2013). Therefore, the initially planned 2×2 ANCOVAs were carried out.

There was a main effect of giftedness on total performance during the learning phase, $F(1, 121)=10.77, p=.001, \eta_p^2=.08$, and a main effect of task sequence, $F(1, 121)=10.96, p=.001, \eta_p^2=.08$. NFC was not a significant predictor, $F(1, 121)=0.54, p=.464, \eta_p^2<.01$. As expected, gifted students performed significantly better ($M=1.77, SD=1.31$) than nongifted students ($M=0.82, SD=1.23$) and students in the EPEP conditions performed significantly better ($M=1.68, SD=1.57$) than students in the PEPE conditions ($M=0.94, SD=0.97$). Note that these means and standard deviations reflect the sum of students' performance on both practice problems. There was no significant interaction between giftedness and task sequence, $F(1, 121)=3.33, p=.070, \eta_p^2=.03$.

There was a main effect of giftedness on performance on the first practice problem, $F(1, 121)=13.59, p<.001, \eta_p^2=.10$, and a main effect of task sequence, $F(1, 121)=61.55, p<.001, \eta_p^2=.34$, but no effect of NFC, $F(1, 121)=0.36, p=.550, \eta_p^2<.01$. As expected, gifted students performed significantly better on the first practice problem than nongifted students and students in the EPEP conditions performed significantly better than students in the PEPE conditions. There was a significant interaction between giftedness and task sequence, $F(1, 121)=15.94, p<.001, \eta_p^2=.12$. Follow-up independent-samples *t*-tests with a Bonferroni-correction (α set to 0.25) showed that performance on the first problem was better for those in the EPEP condition than those in the PEPE condition, both for gifted students, $t(63)=7.94, p<.001, d=2.07$, and for nongifted students, $t(59)=2.89, p=.005, d=1.06$.

On the second practice problem, there was no main effect of giftedness, $F(1, 121)=3.29, p=.072, \eta_p^2=.03$, task sequence, $F(1, 121)=0.63, p=.43, \eta_p^2<.01$, or NFC, $F(1, 121)=0.30, p=.583, \eta_p^2<.01$, and no interaction effect between giftedness and task sequence, $F(1, 121)=0.05, p=.822, \eta_p^2<.01$.

Performance on the final test

On both measures, there was no NFC * Task Sequence interaction [isomorphic: $F(1, 118)=0.34, p=.563, \eta_p^2<.01$; transfer: $F(1, 118)=1.14, p=.289, \eta_p^2=.01$] and no NFC * Giftedness interaction [isomorphic: $F(1, 118)=1.37, p=.245, \eta_p^2=.01$; transfer: $F(1, 118)=1.73, p=.191, \eta_p^2=.01$], which indicates that the assumption of homogeneity of regression slopes was met. NFC was not a significant predictor for either isomorphic test performance [giftedness: $F(1, 121)=28.05, p<.001, \eta_p^2=.19$; task sequence: $F(1, 121)=3.02, p=.085$, Giftedness * Task Sequence: $F(1, 121)=0.26, p=.610, \eta_p^2<.01$, NFC: $F(1, 121)=1.40, p=.239, \eta_p^2=.01$] or transfer test performance [giftedness: $F(1, 121)=25.84, p<.001, \eta_p^2=.18$; task sequence: $F(1, 121)<0.01, p=.965, \eta_p^2<.01$; Giftedness * Task Sequence: $F(1, 121)=1.95, p=.165, \eta_p^2=.02$, NFC: $F(1, 121)=1.22, p=.272, \eta_p^2=.01$].

Role of perceived competence and perceived autonomy

We explored whether perceived competence and perceived autonomy concerning the water jug tasks had an effect on (1) performance on the final test and (2) the effect of task sequence, by conducting ANCOVAs with task sequence (EPEP vs. PEPE) as the independent variable and perceived autonomy and perceived competence as covariates. On performance on

the isomorphic posttest problems, there was a main effect of perceived competence [$F(1, 120)=5.70, p=.018, \eta_p^2=.05$], indicating that participants who rated their competence higher performed better on the isomorphic posttest problems. There were no other main or interaction effects [condition: $F(1, 120)=0.66, p=.419, \eta_p^2<.01$; perceived autonomy: $F(1, 120)=2.82, p=.095, \eta_p^2=.02$; Condition * Perceived Autonomy: $F(1, 120)=0.08, p=.781, \eta_p^2<.01$; Condition * Perceived Competence: $F(1, 120)<0.01, p=.929, \eta_p^2<.01$].

On performance on the transfer posttest problems, there was a main effect of perceived autonomy [$F(1, 120)=5.58, p=.020, \eta_p^2=.04$], indicating that participants who rated their autonomy higher performed better on the transfer posttest problems. There were no other main or interaction effects [condition: $F(1, 120)=0.41, p=.525, \eta_p^2<.01$; perceived competence: $F(1, 120)=0.81, p=.371, \eta_p^2<.01$; Condition * Perceived Autonomy: $F(1, 120)=0.27, p=.607, \eta_p^2<.01$; Condition * Perceived Competence: $F(1, 120)=0.02, p=.904, \eta_p^2<.01$].

Do giftedness and task sequence affect mental effort (Hypothesis 1b)?

We first checked for differences among conditions in the average mental effort investment reported during the learning phase. Results showed that gifted students ($M=3.45, SD=1.00$) reported significantly lower effort investment than nongifted students ($M=4.10, SD=1.48$), as indicated by a main effect of giftedness, $F(1, 122)=8.79, p=.004, \eta_p^2=.07$. There was also a main effect of task sequence, $F(1, 122)=5.93, p=.016, \eta_p^2=.05$, with students in the EPEP conditions ($M=3.49, SD=1.33$) reporting lower levels of effort investment than students in the PEPE conditions ($M=4.04, SD=1.20$). There was no significant interaction, $F(1, 122)=1.66, p=.200, \eta_p^2=.01$.

Regarding the isomorphic problems, there was a significant main effect of giftedness, $F(1, 122)=5.07, p=.026, \eta_p^2=.04$. Gifted students reported having invested less effort ($M=2.64, SD=1.33$) than nongifted students ($M=3.63, SD=2.37$). None of the other effects were significant [task sequence: $F(1, 122)=2.11, p=.149, \eta_p^2=.02$; Giftedness * Task Sequence: $F(1, 122)=1.60, p=.208, \eta_p^2=.01$].

Concerning the near transfer posttest problems, results showed a main effect of giftedness, $F(1, 122)=8.75, p=.004, \eta_p^2=.07$. Gifted students reported significantly lower levels of effort investment ($M=5.11, SD=1.71$) than nongifted students ($M=6.13, SD=2.22$). The main effect of task sequence was also significant, $F(1, 122)=4.27, p=.041, \eta_p^2=.03$, as students in the EPEP conditions ($M=5.95, SD=2.01$) reported having invested significantly more effort in solving the near transfer posttest problems than students in the PEPE conditions ($M=5.25, SD=2.01$). There was no interaction between giftedness and task sequence, $F(1, 122)=.75, p=.388, \eta_p^2<.01$.

Do giftedness and task sequence affect aspects of student motivation (Hypothesis 2)?

We firstly checked whether there were pre-existing differences among conditions in terms of perceived autonomy and perceived competence regarding math tasks in general. There were no significant differences among conditions on perceived autonomy [giftedness: $F(1, 122)=2.62, p=.108, \eta_p^2=.02$; task sequence: $F(1, 122)=0.12, p=.732, \eta_p^2<.01$; Giftedness * Task Sequence: $F(1, 122)=0.27, p=.602, \eta_p^2<.01$] or perceived competence

[giftedness: $F(1, 122)=2.79, p=.097, \eta_p^2=.02$; task sequence: $F(1, 122)=0.03, p=.860, \eta_p^2<.01$; Giftedness * Task Sequence: $F(1, 122)=1.39, p=.241, \eta_p^2=.01$].

Regarding students' perceived autonomy after the learning phase (Hypothesis 2a), results showed a main effect of giftedness, $F(1, 122)=5.75, p=.018, \eta_p^2=.05$. Gifted students reported higher levels of perceived autonomy ($M=2.79, SD=0.46$) than nongifted students ($M=2.56, SD=0.59$). None of the other effects were significant [task sequence: $F(1, 122)=1.01, p=.316, \eta_p^2<.01$; Giftedness * Task Sequence: $F(1, 122)=1.04, p=.310, \eta_p^2=.01$].

As for students' perceived competence after the learning phase (Hypothesis 2b), results showed a main effect of giftedness, $F(1, 122)=6.65, p=.011, \eta_p^2=.05$. Consistent with students' performance on the posttest, gifted students indicated higher levels of perceived competence ($M=2.96, SD=0.48$) than nongifted students ($M=2.71, SD=0.61$). There was also a main effect of task sequence, $F(1, 122)=7.71, p=.006, \eta_p^2=.06$. Students in the EPEP condition ($M=2.97, SD=0.59$) reported greater levels of perceived competence than students in the PEPE condition ($M=2.70, SD=0.49$). There was no interaction effect between giftedness and task sequence, $F(1, 122)=0.79, p=.376, \eta_p^2<.01$.

There were no differences among conditions on learning enjoyment (Hypothesis 2c; giftedness: $F(1, 122)=0.22, p=.637, \eta_p^2<.01$; task sequence: $F(1, 122)=0.77, p=.381, \eta_p^2<.01$; Giftedness * Task Sequence: $F(1, 122)=2.80, p=.097, \eta_p^2=.02$].

Discussion

The main purpose of this experiment was to yield more evidence on why in previous research, novice students were often found to benefit more from example–problem pairs than problem–example pairs (i.e., attain better learning outcomes with less perceived effort investment during the learning phase; Kant et al. 2017; Reisslein et al. 2006; Van Gog et al. 2011). We examined the motivational hypothesis proposed by Van Gog et al. (2011) with a sample of nongifted and gifted primary school students, who learned a difficult math task from example–problem or problem–example pairs. The motivational hypothesis would predict that nongifted students benefit more from example–problem pairs than from problem–example pairs, because a failed problem-solving attempt would be demotivating for them, resulting for instance in a loss of confidence in own abilities and motivation needed to work on the remaining tasks. Gifted students should not be negatively affected by a failed problem-solving attempt, because the task complexity and open task structure should stimulate their perceptions of autonomy (e.g., in terms of trying out distinct problem solving approaches) and the challenge should directly boost their perceived competence. Therefore, the motivational hypothesis would predict that gifted students would particularly benefit from problem–example pairs.

Our gifted and nongifted student samples indeed showed individual differences in terms of NFC, confidence, and motivation (Hypotheses 2a and 2b). That is, gifted students reported higher levels of NFC prior to the experiment, as well as higher levels of perceived competence and perceived autonomy than nongifted students after the learning phase. As expected, gifted students also attained better test performance with less effort investment during both the learning and test phase. We did not find the expected difference in learning enjoyment between gifted and nongifted students, however (Hypothesis 2c). In contrast to Hypothesis 1a, the effects of example–problem and problem–example pairs did not depend on giftedness (i.e., there were no significant interaction effects between task sequence and

giftedness on any of the outcome variables in the test phase). There was no difference in the effect of the different task sequences between gifted and nongifted students on test performance. Task sequence did have a different effect on gifted students than on nongifted students in the learning phase: both gifted and nongifted students performed better on the first problem in the learning phase when they started with a problem, but for nongifted students the effect of the order of problems and examples on first problem performance was smaller.

Furthermore, although we did find that it was more efficient to study example–problem pairs, because the example–problem conditions attained that similar level of test performance with less effort investment during the learning phase than the problem–example conditions (for an elaboration on the concept of cognitive efficiency, see Van Gog and Paas 2008), this higher efficiency was—contrary to Hypothesis 1b—present for both gifted and nongifted students. Example–problem pairs also led to higher levels of perceived competence than problem–example pairs, but there were no differences in perceived autonomy or learning enjoyment between example–problem and problem–example pairs.

A possible explanation for our findings might be that both gifted *and* nongifted students scored above a certain motivational threshold and that these relatively high levels of motivation shielded them from becoming demotivated after starting the learning phase with a failed practice problem-solving attempt. Primary school students are known to have relatively high levels of (intrinsic) math motivation compared to, for instance, secondary education students (Gottfried et al. 2001; Jacobs et al. 2002). Furthermore, students might have perceived the water-jug task, although it is a complex mathematical task, as a puzzle problem. Hence, starting with a failed practice problem-solving attempt might have sparked curiosity and interest rather than a feeling of lack of control over the task and demotivation. Van Gog (2011) made a similar argument after finding no differences between an example–problem pairs and problem–example pairs condition for university students learning how to solve a puzzle task. Evidence for this explanation comes from the finding that despite the fact that being provided with example–problem pairs resulted in higher perceived competence than problem–example pairs, learning enjoyment levels were similar across the task sequence conditions. Moreover, both gifted and nongifted students reported remarkably high levels of learning enjoyment, much higher levels than reported in prior video modeling example research in which students learned how to troubleshoot electrical circuits (e.g., Hoogerheide et al. 2014: 39% enjoyment; Hoogerheide et al. 2018: 31% enjoyment).

Strengths of our study were that we extended research on learning from examples to younger students and gifted students. Most research has focused on nongifted adult or adolescent students. As such, not much is known about the conditions under which example study is most effective for children, and whether gifted students would benefit from example study at all. Our findings suggest that example study can also be an effective instructional strategy for gifted students.

In the current study, we compared gifted and nongifted students. Based on the literature, we expected these two groups of students to differ in a number of aspects. We measured several motivational aspects before the experiment, and the gifted students in our sample reported higher levels of NFC, perceived competence, and perceived autonomy than the nongifted students. However, a limitation is that our sample of gifted and nongifted students likely differed on other aspects than just the motivational aspects of learning that we measured, such as their use of learning strategies and/or metacognitive abilities (Greene et al. 2008; Snyder et al. 2011). One could expect these additional differences between our gifted and nongifted participants to have an effect on how the students study and deal

with examples and practice problems in general, even though it did not differentially affect learning from the sequences studied here.

Another limitation of the current experiment is that we only collected motivational measures before and after the learning phase. The second practice problem (on which we found no performance difference between the two task sequence conditions) may have had a larger impact on motivation than the first practice problem, on which participants scored higher when they started with an example. We decided against measuring students' confidence and motivation during the learning phase in order to limit the length of the experiment. However, asking students for their confidence and motivation after each task in the learning phase would have allowed for a more fine-grained test of the motivational hypothesis. Nevertheless, had there been an effect on motivation, our data suggests that this most likely would have been temporary as the overall measures showed no effect of sequence (and this would be in line with the analysis of the performance on the practice problems during the learning phase, which showed that students recovered quickly from starting with a practice problem). This explanation could be further examined in a replication with single problem–example pairs, which would allow measuring the motivation after the first practice problem and would limit the length of the experiment.

An interesting avenue for future research would be to compare a task like the water-jug task that students seem to find enjoyable, with a task that they would find less enjoyable, to be able to tease apart the motivational effects of the type of task and the order in which examples and problems are presented. Another interesting suggestion for future research would be to further examine how gifted students, who reported higher autonomy than nongifted students, would respond to other sequences of example study and/or problem solving, such as example study only versus example–problem pairs versus problem–example pairs versus problem solving only (cf. Van Gog et al. 2011). Recent findings suggest that example study only is as effective as and perhaps even more efficient than example–problem pairs, even on delayed posttests (Leahy et al. 2015; Van Gog and Kester 2012; Van Gog et al. 2015). This effect might be different for gifted students, who typically prefer to have more autonomy over their learning.

Our findings also have relevance for the design of online learning environments. It seems solid advice to start the learning phase with an example and not a practice problem. Although example–problem and problem–example pairs might not always have differential effects on test performance, example–problem pairs seem to be more efficient and more beneficial for perceived competence for gifted and nongifted students.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix 1

Goal of the puzzle: make sure the tower is built in the same way (from broad to narrow), but on the right peg. You are only allowed to move one disc at a time. The puzzle can be solved in seven steps, but you can use all the pegs below to draw, if you want. Draw your final solution on the back of this sheet. You have 4 min to solve the puzzle. Good luck!



STEP 1:



STEP 2:



STEP 3:



STEP 4:



STEP 5:



STEP 6:

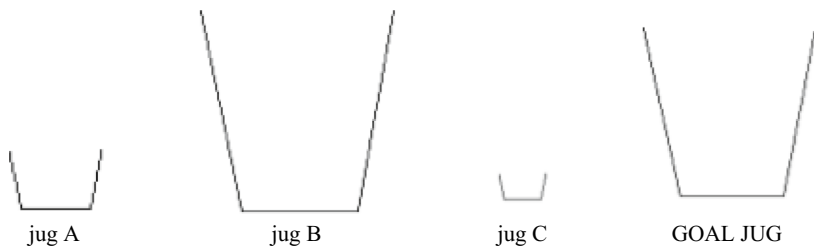


STEP 7:



Appendix 2

The water jug task



You will get a task, in which you will see four different water jugs. These are the work jugs and one goal jug. Under each jug will be a number, indicating how many liters of water fit into the jug. Using the work jugs, you have to fill the goal jug in a clever way. The size of the goal jug is also indicated under the jug, and you have to fill the goal jug with exactly enough water, using the work jugs. You have to fill the work jugs, pour the water into other jugs and try to fill the goal jug with the right amount of water in as little steps as possible. Of course, there are rules you will have to follow. Study the rules below:

Rules:

1. You can fill the work jugs you want to use with water.
2. A work jug is always filled completely, until it contains the maximum amount of liters.
3. You can pour water from a work jug into other work jugs and into the goal jug.
4. You can also empty the work jugs and use them again.
5. The goal jug can only be filled using the work jugs.
6. If you pour water from a work jug into the goal jug, you have to completely empty the work jug.
7. You cannot pour water from the goal jug back into the work jugs.

References

- Assor, A. (2012). Allowing choice and nurturing an inner compass: Educational practices supporting students' need for autonomy. In *Handbook of research on student engagement* (pp. 421–439). Boston: Springer. https://doi.org/10.1007/978-1-4614-2018-7_20.
- 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/10.2307/1170661>.
- Cacioppo, J. T., & Petty, R. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131. <https://doi.org/10.1037//0022-3514.42.1.116>.
- Cacioppo, J. T., Petty, R. E., Feinstein, A. J., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*, 197–253. <https://doi.org/10.1037/0033-2909.119.2.197>.
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*, 306–307. https://doi.org/10.1207/s15327752jpa4803_13.

- Clinkenbeard, P. (2012). Motivation and gifted students: Implications of theory and research. *Psychology in the Schools*. <https://doi.org/10.1002/pits.21628>.
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). Los Angeles: Sage.
- Flunger, B., Pretsch, J., Schmitt, M., & Ludwig, P. (2013). The role of explicit need strength for emotions during learning. *Learning and Individual Differences*, 23, 241–248. <https://doi.org/10.1016/j.lindif.2012.10.001>.
- Foster, N. L., Rawson, K. A., & Dunlosky, J. (2018). Self-regulated learning of principle-based concepts: Do students prefer worked examples, faded examples, or problem solving? *Learning and Instruction*, 55, 124–138.
- Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology*, 93(1), 3–13. <https://doi.org/10.1037/0022-0663.93.1.3>.
- Greene, J. A., Moos, D. C., Azevedo, R., & Winters, F. I. (2008). Exploring differences between gifted and grade-level students' use of self-regulatory learning processes with hypermedia. *Computers and Education*, 50, 1069–1083. <https://doi.org/10.1016/j.compedu.2006.10.004>.
- Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior*, 41, 80–91. <https://doi.org/10.1016/j.chb.2014.09.013>.
- Hoogerheide, V., Van Wermeskerken, M., Van Nassau, H., & Van Gog, T. (2018). Model-observer similarity and task-appropriateness in learning from video modeling examples: Do model and student gender affect test performance, self-efficacy, and perceived competence? *Computers in Human Behavior*, 89, 457–464. <https://doi.org/10.1016/j.chb.2017.11.012>.
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, 73, 509–527. <https://doi.org/10.1111/1467-8624.00421>.
- Kant, J., Scheiter, K., & Oschatz, K. (2017). How to sequence video modeling examples and inquiry tasks to foster scientific reasoning. *Learning and Instruction*, 52, 46–58. <https://doi.org/10.1016/j.learninstruc.2017.04.005>.
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26, 379–424. <https://doi.org/10.1080/07370000802212669>.
- Leahy, W., Hanham, J., & Sweller, J. (2015). High element interactivity information during problem solving may lead to failure to obtain the testing effect. *Educational Psychology Review*, 27, 291–304.
- Leppink, J., Paas, F. G. W. C., Van Gog, T., Van der Vleuten, C. P. M., & Van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction*, 30, 32–42. <https://doi.org/10.1016/j.learninstruc.2013.12.001>.
- Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review*, 29, 693–715. <https://doi.org/10.1007/s10648-016-9379-x>.
- Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, 54, 1–95. <https://doi.org/10.1037/h0093502>.
- Meier, E., Vogl, K., & Preckel, F. (2014). Motivational characteristics of students in gifted classes: The pivotal role of need for cognition. *Learning and Individual Differences*, 33(Supplement C), 39–46. <https://doi.org/10.1016/j.lindif.2014.04.006>.
- Morisano, D., & Shore, B. M. (2010). Can personal goal setting tap the potential of the gifted underachiever? *Roepers Review*, 32, 249–258. <https://doi.org/10.1080/02783193.2010.508156>.
- Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics—A cognitive-load approach. *Journal of Educational Psychology*, 84, 429–434. <https://doi.org/10.1037/0022-0663.84.4.429>.
- Pajares, F. (1996). Self-efficacy beliefs and mathematical problem-solving of gifted students. *Contemporary Educational Psychology*, 21, 325–344. <https://doi.org/10.1006/ceps.1996.0025>.
- Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, 24, 124–139. <https://doi.org/10.1006/ceps.1998.0991>.
- Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and Instruction*, 16, 92–103. <https://doi.org/10.1016/j.learninstruc.2006.02.008>.
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38, 1–37. <https://doi.org/10.1111/cogs.12086>.

- Schwaighofer, M., Bühner, M., & Fischer, F. (2016). Executive functions as moderators of the worked example effect: When shifting is more important than working memory capacity. *Journal of Educational Psychology, 108*, 982–1000. <https://doi.org/10.1037/edu0000115>.
- Snyder, K. E., Nietfeld, J. L., & Linnenbrink-Garcia, L. (2011). Giftedness and metacognition: A short-term longitudinal investigation of metacognitive monitoring in the classroom. *Gifted Child Quarterly, 55*, 181–193. <https://doi.org/10.1177/0016986211412769>.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.
- Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*, 59–89.
- Sweller, J., & Levine, M. (1982). Effects of goal specificity on means–ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 463.
- Van Gog, T. (2011). Effects of identical example–problem and problem–example pairs on learning. *Computers and Education, 57*, 1775–1779.
- Van Gog, T., & Kester, L. (2012). A test of the testing effect: Acquiring problem-solving skills from worked examples. *Cognitive Science, 36*, 1532–1541. <https://doi.org/10.1111/cogs.12002>.
- Van Gog, T., Kester, L., Dirx, K., Hoogerheide, V., Boerboom, J., & Verhoeijen, P. P. J. L. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educational Psychology Review, 27*, 265–289. <https://doi.org/10.1007/s10648-015-9297-3>.
- Van Gog, T., Kester, L., & Paas, F. G. W. C. (2011). Effects of worked examples, example–problem, and problem–example pairs on novices' learning. *Contemporary Educational Psychology, 36*, 212–218. <https://doi.org/10.1016/j.cedpsych.2010.10.004>.
- Van Gog, T., & Paas, F. G. W. C. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*, 16–26.
- Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social–cognitive research perspectives. *Educational Psychology Review, 22*, 155–174. <https://doi.org/10.1007/s10648-010-9134-7>.
- Van Gog, T., Rummel, N., & Renkl, A. (2019). Learning how to solve problems by studying examples. In J. Dunlosky & K. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 183–208). New York: Cambridge University Press.

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