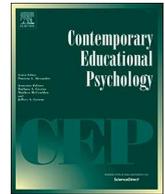




ELSEVIER

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

Effects of different sequences of examples and problems on motivation and learning

Milou van Harsel^{a,b,*}, Vincent Hoogerheide^b, Peter Verkoeijen^{a,c}, Tamara van Gog^b^a Learning and Innovation Centre, Avans University of Applied Sciences, the Netherlands^b Department of Education, Utrecht University, the Netherlands^c Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, the Netherlands

ARTICLE INFO

Keywords:

Example-based learning
Worked examples
Problem solving
Motivation
Mental effort
Instructional design

ABSTRACT

Recent research has shown that example study only (EE) and example-problem pairs (EP) were more effective (i.e., higher test performance) and efficient (i.e., attained with less effort invested in learning and/or test tasks) than problem-example pairs (PE) and problem solving only (PP). We conducted two experiments to investigate how different example and problem-solving sequences would affect motivational (i.e., self-efficacy, perceived competence, and topic interest) and cognitive (i.e., effectiveness and efficiency) aspects of learning. In Experiment 1, 124 technical students learned a mathematical task with the help of EEEE, EPEP, PEPE, or PPPP and then completed a posttest. Students in the EEEE Condition showed higher posttest performance, self-efficacy, and perceived competence, attained with less effort investment, than students in the EPEP and PPPP Condition. Surprisingly, there were no differences between the EPEP and PEPE Condition on any of the outcome measures. We hypothesized that, because the tasks were relevant for technical students, starting with a problem might not have negatively affected their motivation. Therefore, we replicated the experiment with a different sample of 81 teacher training students. Experiment 2 showed an efficiency benefit of EEEE over EPEP, PEPE, and PPPP. However, only EEEE resulted in greater posttest performance, self-efficacy, and perceived competence than PPPP. We again did not find any differences between the EPEP and PEPE Condition. These results suggest that, at least when short training phases are used, studying examples (only) is more preferable than problem solving only for learning. Moreover, this study showed that example study (only) also enhances motivational aspects of learning whereas problem solving only does not positively affect students' motivation at all.

1. Introduction

Example-based learning is an effective and efficient instructional strategy for novices to acquire new problem solving skills. Research has repeatedly shown that instruction that relies more heavily on example study, yields better learning outcomes than engaging in practice problem solving only (for reviews, see Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 2014; Sweller, Ayres, & Kalyuga, 2011; Sweller, Van Merriënboer, & Paas, 1998; Van Gog & Rummel, 2010). This is known as the worked example effect. Notwithstanding several decades of research, an important open question in research on example-based learning is whether and when example study should be alternated with practice problem solving to be effective and efficient for learning.

Historically, most studies on the worked example effect have used *example-problem pairs* (e.g., Carroll, 1994; Cooper & Sweller, 1987; Kalyuga, Chandler, Tuovinen, & Sweller, 2001; McLaren, Lim, &

Koedinger, 2008; Mwangi & Sweller, 1998; Rourke & Sweller, 2009; Sweller & Cooper, 1985). Others used *example study only* (e.g., Van Gerven, Paas, Van Merriënboer, & Schmidt, 2002; Van Gog, Paas, & Van Merriënboer, 2006). Both approaches were found to be more effective and efficient for learning and transfer than *problem solving only*. Another means of implementing examples and problems is to use *problem-example pairs* (e.g., Hausmann, Van de Sande, & VanLehn, 2008; Reisslein, Atkinson, Seeling, & Reisslein, 2006; Stark, Gruber, Renkl, & Mandl, 2000). In a direct comparison of all four approaches of Van Gog, Kester, and Paas (2011), students were randomly assigned to learn how to troubleshoot electrical circuits (in four training tasks) by means of example study only (EEEE), example-problem pairs (EPEP), problem-example pairs (PEPE), or practice problem solving only (PPPP). Time-on-task was kept constant. Results showed no differences in test performance or mental effort investment in the training phase between the EEEE condition and EPEP condition and between the PEPE condition

* Corresponding author at: Avans University of Applied Sciences, Learning and Innovation Centre, PO Box 90116, 4800 RA Breda, the Netherlands.

E-mail address: m.vanharsel@avans.nl (M. van Harsel).

<https://doi.org/10.1016/j.cedpsych.2019.03.005>

Available online 13 March 2019

0361-476X/© 2019 Elsevier Inc. All rights reserved.

and PPPP condition. The EEEE condition and EPEP condition were, however, more effective (i.e., attained significantly higher test performance; medium to large effect) and more efficient (i.e., attained significantly higher test performance with less invested mental effort in the training phase; medium to large effect) compared to the PEPE condition and PPPP condition (for a discussion of efficiency, see Hoffman & Schraw, 2010; Van Gog & Paas, 2008).

Several studies have by now replicated the finding that learning outcomes after EEEE and EPEP do not differ significantly (Van der Meij, Rensink, & Van der Meij, 2018), even on a delayed posttest (Leahy, Hanham, & Sweller, 2015; Van Gog & Kester, 2012; Van Gog et al., 2015). Note though, that these studies did not include motivational variables (which have largely been ignored in worked example research; Renkl, 2014; Sweller et al., 2011; Van Gog & Rummel, 2010). It has been suggested –though not yet tested– that solving a (similar) problem immediately after studying an example may be more motivating for students than passively studying examples only, because practice problem solving requires learners to actively apply what they have learned (e.g., Sweller & Cooper, 1985; Trafton & Reiser, 1993). If EPEP was found to be more motivating than EEEE whilst yielding comparable levels of learning outcomes, it would be highly relevant for educational practice. Outside a laboratory research setting, motivational variables might affect learning outcomes via persistence. That is, in (online) learning environments, students can decide for themselves whether they continue to working on a task (sequence) or not, so how motivating a task sequence is becomes important.

Another noteworthy finding in the Van Gog et al. (2011) study in which motivational aspects of learning might have played a role, was that EPEP was more effective and efficient for learning than PEPE, even though both received the same number of examples to study. Moreover, students in the PEPE condition –despite receiving two examples– did not outperform students in the PPPP condition. This finding, which has since been replicated in two other studies (e.g., Kant, Scheiter, & Oschatz, 2017; Leppink, Paas, Van Gog, Van der Vleuten, & Van Merriënboer, 2014¹), suggests that the order in which example study and practice problem solving is alternated, matters: if novice learners start with a practice problem, example study loses its effectiveness (see also Reisslein et al., 2006). Van Gog et al. (2011) suggested –but did not test– that motivational aspects of learning might explain this finding: “students may not be motivated to study the example because of the negative experience of a failed problem solving attempt” (p. 217). That is, when novices have to learn how to solve a complex task that requires domain-specific knowledge and is not particularly intrinsically rewarding (such as the physics task in the study by Van Gog et al., 2011), then starting the training phase with a practice problem (i.e., PEPE) might lead to a decrease in student motivation. When the practice problem is being experienced as so difficult that students lose confidence in their own abilities or lose interest to learn the task, they may not be motivated to study the subsequent example (and possibly also the tasks that follow). Starting with an example (EPEP) gives students a basis for how to approach the subsequent practice problem, and may therefore prevent students from becoming demotivated.

Three aspects of motivation that may be affected by example-problem sequences are self-efficacy, perceived competence, and topic interest. Self-efficacy is a key construct in Bandura (1986) social learning theory and can be defined as a person’s belief in their own capacity to organize or accomplish a specific task or challenge (see also Bandura, 1997; Schunk, 1987). Self-efficacy has been shown to have a positive effect on factors such as academic motivation, study behavior, and learning outcomes (e.g., Bandura, 1997; Bong & Skaalvik, 2003; Schunk, 2001). Perceived competence plays a central role in Deci and

Ryan (2002) self-determination theory of motivation and has also shown to have significant influence on academic motivation and learning outcomes (e.g., Bong & Skaalvik, 2003). This construct is related to self-efficacy, but covers more general knowledge and perceptions (Bong & Skaalvik, 2003; Hughes, Galbraith, & White, 2011; Klassen & Usher, 2010). Topic interest is a motivational construct that can be described as the level of interest generated by a specific topic (Ainley, Hidi, & Berndorff, 2002; Renninger, 2000; Schiefele & Krapp, 1996) and seems to have a positive effect on (deeper) learning and engagement (e.g., Benton, Corkill, Sharp, Downey, & Khramtsova, 1995; Schiefele & Krapp, 1996; Tobias, 1996). Although research has shown that example study only can foster students’ self-efficacy and perceived competence (e.g., Bandura, 1997; Crippen, Biesinger, Muis, & Orgill, 2009; Hoogerheide, Loyens, & Van Gog, 2014; Hoogerheide, Van Wermeskerken, van Nassau, & Van Gog, 2018), it is an open question how different sequences would affect motivational aspects of learning.

In sum, it is both theoretically and practically relevant to address whether different sequences of examples and practice problems would differentially affect not only cognitive (i.e., effectiveness and efficiency), but also motivational aspects of learning (i.e., self-efficacy, perceived competence, and topic interest). The present study set out to do so, by performing a conceptual replication of Van Gog et al. (2011), extended with motivational measures.

1.1. The present study

The purpose of Experiment 1 was to investigate the effects of different example and problem solving sequences on motivational and cognitive aspects of learning. We conducted a conceptual replication of the Van Gog et al. (2011) study, with the same task sequences (i.e., EEEE, EPEP, PEPE, and PPPP), but a different population (i.e., higher education students rather than secondary education students), different training tasks (i.e., mathematics tasks rather than physics tasks), and a different example format (i.e., video modeling examples consisting of screen recordings with voice-over, rather than worked examples; cf. McLaren et al., 2008). In addition to performance on posttest tasks and reported effort investment in the training phase, we measured time-on-task in the training phase, as well as mental effort and time-on-task in the posttest phase as (explorative) indicators of efficiency of the learning process and learning outcomes. We added a procedural and conceptual transfer task and a delayed posttest one week later to investigate any effects on transfer and whether effects would remain stable over time (cf. Van Gog et al., 2015). The most important novelty of Experiment 1 was that we measured the following motivational aspects of learning before and after the training phase: self-efficacy, perceived competence, and topic interest.

The main aim was to investigate how the different example and problem solving sequences (i.e., EEEE, EPEP, PEPE, and PPPP) would affect (a) motivational aspects of learning (i.e., self-efficacy, perceived competence, and topic interest), and (b) cognitive aspects of learning (i.e., effectiveness and efficiency). We expected that an EPEP sequence would result in higher levels of self-efficacy, perceived competence, and topic interest than an EEEE sequence (cf. the suggestion by Sweller & Cooper, 1985 and Trafton & Reiser, 1993; Hypothesis 1a). Based on the motivational explanation for the effectiveness and efficiency of EPEP over PEPE proposed by Van Gog et al. (2011), we also expected EPEP to be more beneficial for self-efficacy, perceived competence, and topic interest than PEPE (Hypothesis 1b). We had no hypotheses for the other condition comparisons, so we examined them in an exploratory manner (Question 1c).

Regarding cognitive aspects of learning, we expected to replicate the findings by Van Gog et al. (2011) regarding both isomorphic problem-solving performance (i.e., EEEE = EPEP > PEPE = PPPP; Hypothesis 2) and mental effort invested in the training phase (i.e., EEEE = EPEP > PEPE = PPPP; Hypothesis 3). That is, we expected the EEEE and EPEP condition to attain greater posttest performance with

¹ Note though, that EPEP > PEPE was not found when the examples and problems remained fully identical throughout the sequences (Van Gog et al., 2011).

less effort investment in the training phase than the PEPE and PPPP condition, and no differences to arise on these variables between the EEEE and EPEP condition and between the PEPE and PPPP condition. Because example-based learning has been found to be effective not only for learning to solve similar problems, but also for solving transfer problems (e.g., Cooper & Sweller, 1987; Paas, 1992; Paas & Van Merriënboer, 1994), we expected the same pattern of results for performance on the procedural transfer task (i.e., EEEE = EPEP > PEPE = PPPP; Hypothesis 4) and conceptual transfer task (i.e., EEEE = EPEP > PEPE = PPPP; Hypothesis 5).

2. Experiment 1

2.1. Method

2.1.1. Participants and design

An a-priori-power analysis was conducted to determine how many participants we would need to be able to reliably detect the effect sizes reported by Van Gog et al. (2011). Inserting $\eta_p^2 = 0.23$ (i.e., effect size for test performance found in the study by Van Gog et al., 2011) into G*Power and performing an a-priori-power analysis for a one-way ANOVA with four groups, with an alpha of 0.05, and a power of 0.95, yielded a total sample of 64. Participants were 124 first-year students from a Dutch University of Applied Sciences, enrolled in an electrical and electronic or mechanical engineering program ($M^{age} = 19.25$, $SD = 1.90$; 117 male, 7 female). At the time of the experiment, students were novices to the task being taught in this study (i.e., approximating the definite integral using the trapezoidal rule) as this topic had not yet been taught in their curriculum. They received study credits for their participation. The experiment consisted of 3 phases, namely: the pretest, training, and immediate posttest phase. Participants were randomly assigned to one of four conditions: (1) examples only ($n = 34$; EEEE), (2) example-problem pairs ($n = 25$; EPEP), (3) problem-example pairs ($n = 30$; PEPE), or (4) practice problems only ($n = 35$; PPPP).

2.1.2. Materials

All materials were presented in a web-based online learning environment.

Training tasks. The training phase consisted of four math tasks that were developed in collaboration with three mathematics teachers of a Dutch University of Applied Sciences. The tasks required the use of the trapezoidal rule (i.e., a numerical integration method which divides a specific region under the graph of a function into trapezoids and calculates its area) to approximate the region under the graph of a function.

Each task had a different cover story (i.e., task 1: fitness, task 2: energy measurement, task 3: soapsuds, and task 4: running). These cover stories were randomly distributed over the four tasks that were used in the training phase. The four tasks were divided in two pairs (i.e., pair 1: fitness and energy measurement, pair 2: soapsuds and running), based on their complexity level. In the first pair of tasks (complexity level 1), only positive numbers were used in constructing the graph of a function, whereas in the second pair (complexity level 2), negative numbers were used in constructing the graph of a function. Requiring students to calculate with negative numbers made the second pair of tasks slightly more complex than the first pair of tasks. Within each pair, the two tasks were isomorphic (i.e., a similar problem-solving procedure was required, but surface features such as the cover stories and numbers used in functions were slightly different).

Two versions of each task were created, a video modeling example and a practice problem. The video modeling example, a video screen capture, showed a digital recording of a female model's computer screen demonstrating step-by-step how to solve a problem using the trapezoidal rule. The visual demonstration was supported by verbal explanations and handwritten notes. The screen capture started with a

brief introduction of the purpose of the trapezoidal rule, followed by an explanation of the problem state. For example, the problem state in the example format of 'Energy measurement' read as follows: "Jalil has bought a solar cell and wants to know how much energy the solar cell supplies during a certain amount of time. Jalil has used an energy meter to examine how much energy the solar cell produces during a specific amount of time. Jalil has measured the energy at different time points and plotted the results in a graph. The time (in minutes) is plotted on the horizontal axis and the power the solar cell supplies (Joule per minute) is plotted on the vertical axis of the graph. By calculating the area under the graph, Jalil can determine how much energy the solar cell has produced during a certain amount of time." Subsequently, the remainder of the example showed and explained how to interpret the corresponding graph of a function with information that was given (i.e., the left border and right border of the area, the number of intervals, the trapezoidal rule), and showed and explained how to solve the problem by using four steps: (1) 'compute the step size of each subinterval', (2) 'calculate the x-values', (3) 'calculate the function values for all x-values', and (4) 'enter the function values into the formula and calculate the area' (for an example of a video modeling example, see Appendix A).

In the problem format, participants first received a short introduction describing the problem state, along with the graph of a function, the left border and right border of the area to be calculated, the number of intervals, and the trapezoidal rule formula. This information was exactly the same as in the video modeling example. It was, however, not explained how to use this information to solve the problem. Participants had to solve the problem themselves by completing four steps: (1) 'compute the step size of each subinterval', (2) 'calculate the x-values', (3) 'calculate the function values for all x-values', and (4) 'enter the function values into the formula and calculate the area'. In addition, they were asked to write down their solution steps (for an example of a practice problem, see Appendix B).

The order of the four tasks was kept constant across conditions (i.e., the training phase always started with 'fitness' and ended with 'running'), only the format of each task varied among conditions (i.e., EEEE and EPEP started with 'fitness' as an example, whereas PEPE and PPPP started with 'fitness' as a practice problem).

Test tasks. The pretest consisted of two tasks that were isomorphic (i.e., same difficulty, different cover stories) to the training tasks (Cronbach's alpha = 0.63). The complexity level of the first and second pretest tasks was identical to the first and second pair of training tasks, respectively. Students were asked to approach the region under the graph using the information that was given and to write down their solution steps. In these problems, the intermediate (four) steps were not explicitly displayed such as in the training tasks.

The immediate posttest consisted of four tasks. The first two tasks were isomorphic to the pretest and training tasks and were used to measure 'learning' (Cronbach's alpha = 0.66). Students needed to apply the exact problem-solving procedure that they learned during the training phase, but tasks were used that differed in terms of surface features such as cover story and numbers used in the function. The third task was a procedural transfer problem in which participants were asked to use the Simpson rule instead of the trapezoidal rule to approximate the definite integral under a graph. The problem-solving procedure of the Simpson rule is similar to that of the problem-solving procedure of the trapezoidal rule. However, Simpson's rule uses a different formula to calculate the area under a graph and approximates the curve with a sequence of quadratic parabolic segments instead of straight lines (such as the trapezoidal rule). The final conceptual transfer task consisted of five questions that aimed to measure students' conceptual understanding of the underlying principles of the trapezoidal rule as a technique to approximate the area under a graph (Cronbach's alpha = 0.54). Each question consisted of a multiple choice part (from which students had to choose the right answer) and an 'explanation' part (where students had to substantiate their chosen

answer). An example of an isomorphic posttest task, procedural transfer task, and a question concerning conceptual transfer can be found in [Appendix C](#).

Mental effort. Participants were asked to rate how much mental effort they had invested after each task on the pretest, the training phase, and the immediate posttest, using the 9-point mental effort rating scale developed by [Paas \(1992\)](#), with answer options ranging from (1) “very, very low mental effort” to (9) “very, very high mental effort”. Research has shown that this measure is an indicator of experienced cognitive load that is sensitive to variations in task complexity ([Paas, Tuovinen, Tabbers, & Van Gerven, 2003](#)).

Self-efficacy, perceived competence, and topic interest. Self-efficacy was measured just before and directly after the training phase by asking participants to rate to what extent they were confident that they could approximate the definite integral of a graph using the trapezoidal rule on a 9-point rating scale, ranging from (1) “very, very unconfident” to (9) “very, very confident”. This was an adapted version of the item used by [Hoogerheide, Van Wermeskerken, Loyens, and Van Gog \(2016\)](#).

Perceived competence was measured using an adapted version of the *Perceived Competence Scale for Learning* ([Williams & Deci, 1996; Williams, Freedman, & Deci, 1988](#)). The original scale consists of 4 items, namely: “I feel confident in my ability to learn this material”, “I am capable of learning the material in this course”, “I am able to achieve my goals in this course”, and “I feel able to meet the challenge of performing well in this course”. We adapted the scale by removing the third question on the topic of personal goals because this question was not relevant for the present study. For the remaining questions, we rephrased the word ‘course’ to focus on approximating the definite integral of a graph using the trapezoidal rule. Participants could rate from (1) “not at all true” to (7) “very true” to what degree the items applied to them. The adapted scale had a good reliability in our sample (Cronbach’s alpha = 0.96). It should be noted, however, that Cronbach’s alpha has a limited degree of precision due to the large sampling error. Nevertheless, a high alpha measure (i.e., above 0.80) has been demonstrated in previous studies (e.g., [Williams & Deci, 1996; Williams et al., 1998](#)).

To measure students’ topic interest, we developed a topic interest scale. Our scale consisted of 7 items (Cronbach’s alpha = 0.80) adapted from the topic interest scale by [Mason, Gava, and Boldrin \(2008\)](#) and from the Perceived Interest Scale developed by [Schraw, Bruning, and Svoboda \(1995\)](#). We selected the items from both scales that focused on feelings and emotions towards a specific topic and adjusted the items to the context of using or practicing the trapezoidal rule. Each item asked participants to rate on a 5-point scale, ranging from 1 (not at all) to 5 (very true), to what degree each of the items applied to them. The items are shown in [Appendix D](#).

2.1.3. Procedure

The experiment was run in sixteen sessions (i.e., eight first sessions and eight second sessions) in a computer classroom with 5 to 25 participants present per session. In the first session, which lasted 100 min on average, the experimenter first provided participants with a general introduction in which she explained the aim and procedure of the experiment. Participants were told that they would be able to work on the tasks at their own pace with a max. of 130 min, and that they would be provided with a headset, a pen, and scrap paper on which they could write down their calculations. Participants were instructed to do their best but could write down an ‘X’ if they really did not know the answer. Then, the experimenter provided students a form with a link to the learning environment so students could enter the environment as soon as the instruction was finished.

The learning environment presented each task and questionnaire on a separate page, which ensured that participants could not go back to previous tasks or questionnaires, nor look ahead until the current task was completed. Time-on-task was logged. The learning environment first randomly assigned participants to one of the four conditions, and

then presented participants with a short demographic questionnaire (e.g., age and gender) and the pretest. Both pretest tasks were followed by the mental effort rating scale, and the pretest was followed by the self-efficacy, perceived competence, and topic interest questionnaires. During the training phase, participants were provided with a combination of examples and/or practice problems, depending on their assigned condition. Each task was followed by the mental effort rating scale, and the training phase was followed by the self-efficacy, perceived competence, and topic interest questionnaires. Finally, participants were provided with the immediate posttest, which consisted of two tasks that were isomorphic to the tasks in first and second pair of the training phase, a procedural transfer task (i.e., Simpson’s rule task) and five open-ended questions to measure conceptual knowledge. Again, after each immediate posttest task participants were provided with the mental effort rating scale. Before starting with the immediate posttest, students were asked to put away the scrap paper they used in the training phase and received a new scrap paper to make notes. After completing the immediate posttest, participants gave the scrap paper containing their calculations to the experimenter.

2.2. Data analysis

For each training task, a maximum of 8 points could be earned: two points for correctly computing the step size of each subinterval (step 1), two points for correctly calculating all the x-values (step 2), two points for correctly calculating the function values for all x-values (step 3), and 2 points for correctly calculating the area by using the correct formula and giving the right answer (step 4). Students received 1 point when half or more of the solution steps were correct in step two, three, and four. If fewer than half of the solution steps were correct, 0 points were granted. These scoring standards were also used to score the pretest (max. 16 points). The same procedure was used for the procedural transfer problem, so a maximum of 8 points could be earned for this task. A maximum of 9 points could be earned on the 5 open-ended questions in the conceptual transfer problem: one point for the first open-ended question (1 point for the correct answer, 0 points for an incorrect answer) and 2 points for the other open-ended questions. The maximum score of 2 points was only granted when participants got the answer right and provided correct reasoning. Only 1 point was granted if the answer was correct but not substantiated by reasoning and 0 points were granted when the answer was completely incorrect.

The data was scored by the experimenter (i.e., first author) based on a standard developed by the authors in collaboration with the mathematics teachers. To measure the reliability of the ratings, two raters independently scored 15% of the tests. The intra-class correlation coefficient was high, with respectively scores of 0.91 on the pretest tasks, 0.94 on the training phase tasks, and 0.98 on the posttest tasks.

Average mental effort was computed separately for the pretest tasks, training tasks, isomorph tasks, and transfer tasks on the immediate posttest. Average scores on self-efficacy, perceived competence, and topic interest were computed separately for the measurement that took place before the training phase and the measurement directly after the training phase.

3. Results

Because several variables were not normally distributed, we analyzed the data with nonparametric tests (cf. [Field, 2009](#)). We tested the main effects of Test Moment with the Wilcoxon signed-rank test and the main effects of Instruction Condition with the Kruskal-Wallis test. For post-hoc tests, we used Mann-Whitney U tests, with a Bonferroni corrected significance level of $p < .013$ (i.e., $0.05/4$) for the Wilcoxon signed-rank tests and a Bonferroni corrected alpha level of $p < .008$ (i.e., $0.05/6$) for the Kruskal-Wallis test. For the post-hoc tests, the effect size of Pearson r correlation is reported (i.e., Z/\sqrt{N}) with values of 0.10, 0.30, and 0.50 representing a small, medium, and large effect size,

Table 1
Mean (M), Standard Deviation (SD) and Median (Med) of Self-Efficacy (range 1–9) Perceived Competence (range 1–7), and Topic Interest (range 1–7) per Condition in Experiment 1.

	EEEE Condition			EPEP Condition			PEPE Condition			PPPP Condition		
	M	SD	Med									
Pretest												
Self-efficacy	4.27	1.91	5.00	4.26	1.83	5.00	4.68	2.17	5.00	3.89	2.05	4.00
Perceived Competence	3.84	1.40	4.00	3.62	1.51	3.67	3.88	1.79	4.00	3.42	1.63	3.33
Topic Interest	4.15	0.89	4.29	3.92	1.12	4.29	3.52	1.01	3.29	4.21	1.04	4.00
Posttest												
Self-efficacy	7.07	0.87	7.00	5.71	2.01	6.00	6.48	2.14	7.00	3.80	1.98	4.00
Perceived Competence	5.70	0.66	5.83	4.76	1.52	5.00	5.45	1.50	5.67	3.32	1.61	3.00
Topic Interest	4.05	0.92	4.14	3.59	1.03	4.71	3.49	1.12	3.57	3.92	0.93	3.86

Table 2
Mean (M), Standard Deviation (SD), and Median (Med) of Pretest (range 0–16), Training Performance (range 0–24), Isomorphic Tasks Performance (range 0–16), Procedural Transfer (range 0–8), Conceptual Transfer (range 0–9), Mental Effort (range 1–9), and Time-on-Task per Condition in Experiment 1.

	EEEE Condition			EPEP Condition			PEPE Condition			PPPP Condition		
	M	SD	Med	M	SD	Med	M	SD	Med	M	SD	Med
Pretest												
Performance	7.17	5.27	8.00	5.26	4.57	4.00	6.92	4.93	7.00	5.57	5.00	5.00
Training												
Performance				4.06	2.68	3.75	4.82	2.01	5.00	3.64	2.14	3.50
Mental Effort	2.51	1.32	2.13	4.02	1.33	4.25	3.75	1.21	4.00	6.01	1.84	6.25
Time-on-Task	3.40	1.17	3.25	9.56	2.95	9.88	7.07	2.65	6.75	9.32	3.60	7.50
Posttest												
Isomorphic Tasks	10.43	2.40	10.00	7.24	4.88	7.00	9.08	4.69	10.00	5.69	4.13	6.00
Procedural Transfer	3.77	2.78	2.00	3.12	3.24	2.00	3.00	2.97	2.00	2.49	2.89	2.00
Conceptual Transfer	4.60	2.22	5.00	4.74	2.81	5.00	5.28	2.48	5.00	3.74	2.31	3.00
Mental Effort												
Isomorphic Tasks	4.53	1.71	4.50	5.13	1.70	5.50	4.60	1.51	5.00	6.39	1.96	6.50
Procedural Transfer	4.00	1.97	3.00	4.44	2.08	5.00	3.72	1.97	3.00	6.03	2.65	6.00
Conceptual Transfer	4.40	1.83	5.00	3.88	1.87	3.00	3.64	1.29	3.00	5.03	2.16	5.00
Time-on-Task												
Isomorphic Tasks	16.25	5.45	15.00	10.94	5.17	10.25	12.16	5.24	11.50	7.74	4.05	7.50
Procedural Transfer	8.63	3.60	8.00	6.97	2.90	7.00	6.72	3.94	6.00	5.14	4.19	5.00
Conceptual Transfer	8.90	4.40	8.00	7.06	3.66	6.00	8.04	3.36	7.00	6.69	2.55	7.00

respectively (Cohen, 1988). Relevant descriptive statistics of self-efficacy, perceived competence, and topic interest scores are presented in Table 1, and performance scores, mental effort scores, and time-on-task scores are presented in Table 2.

Unfortunately, the delayed posttest data had to be excluded from the analyses. We had made a mistake in designing the delayed posttest, as the delayed posttest was not entirely isomorphic to the immediate posttest. The complexity level of the tasks used in the delayed posttest did not correspond to the complexity level of the tasks used in the immediate posttest (i.e., the tasks used in the delayed posttest were less complex than the tasks used in the immediate posttest because students did not have to calculate with fractions or negative numbers). Please see Appendix E for the raw data of the delayed posttest (i.e., means, standard deviations, and medians per condition).

We first checked for prior knowledge differences among conditions. A Kruskal-Wallis test showed no significant differences among conditions on pretest performance, $H(3) = 3.54, p = .315$, and on pretest scores of self-efficacy, $H(3) = 2.36, p = .501$, perceived competence, $H(3) = 2.47, p = .480$, and topic interest, $H(3) = 6.68, p = .083$.

3.1. Does the sequencing of examples and problems affect self-efficacy, perceived competence, and topic interest?

Firstly, we analyzed whether students' self-efficacy increased from before to after the training phase. We found a main effect of Test

Moment, $Z = 5.79, p < .001, r = 0.520$. Follow-up tests showed that the self-efficacy medians of the EEEE ($Z = 4.57, p < .001, r = 0.834$), EPEP ($Z = 3.00, p = .003, r = 0.514$), and PEPE Condition ($Z = 2.91, p = .004, r = 0.582$) significantly increased over time, whereas the medians of the PPPP Condition did not significantly increase over time ($p = .821, r = 0.038$). Regarding the main question of whether there would be differences among instructional conditions on reported self-efficacy measured after the training phase, we found a main effect of Instruction Condition, $H(3) = 43.46, p < .001$. Our findings were not in line with Hypothesis 1a (EPEP > EEEE) and Hypothesis 1b (EPEP > PEPE). Post-hoc tests showed that self-efficacy ratings did not differ between the EPEP and PEPE Condition ($p = .094, r = 0.218$) and that self-efficacy was even significantly higher in the EEEE Condition than in the EPEP Condition ($U = 293.50, p = .003, r = 0.375$). Further explorations showed that self-efficacy was significantly higher in the EEEE ($U = 66, p < .001, r = 0.759$), EPEP ($U = 287, p < .001, r = 0.450$), and PEPE Condition ($U = 144, p < .001, r = 0.573$) than in the PPPP Condition.

Perceived competence showed the same pattern of results. We found a main effect of Test Moment, $Z = 6.03, p < .001, r = 0.542$. Perceived competence increased in the EEEE ($Z = 4.48, p < .001, r = 0.818$), EPEP ($Z = 3.23, p = .001, r = 0.554$), and PEPE Condition ($Z = 3.23, p = .001, r = 0.646$), but not in the PPPP Condition ($p = .455, r = 0.133$). We also found a main effect of Instruction Condition on perceived competence measured after the training phase,

$H(3) = 38.76, p < .001$. In contrast to our expectations (i.e., Hypothesis 1a: EPEP > EEEE; Hypothesis 1b: EPEP > PEPE), perceived competence was significantly higher in the EEEE Condition compared to the EPEP Condition ($U = 315.50, p = .008, r = 0.331$) and there was no significant difference between the EPEP and PEPE Condition ($p = .042, r = 0.264$). Further explorations showed that perceived competence scores were significantly higher in the EEEE ($U = 100.50, p < .001, r = 0.697$), EPEP ($U = 299.50, p = .001, r = 0.428$), and PEPE Condition ($U = 142.50, p = .001, r = 0.573$) compared to the PPPP Condition.

As for topic interest measured after the training phase, we found a main effect of Test Moment, $Z = -3.62, p < .001, r = 0.325$. Students' topic interest significantly decreased over time in the EPEP ($Z = -3.23, p = .001, r = 0.554$) and PPPP Condition ($Z = -2.69, p = .007, r = 0.455$). There was no main effect of Instruction Condition on topic interest measured after the training phase ($p = .143$), indicating that there were no differences among conditions on topic interest. Hence, the topic interest results contrasted Hypothesis 1a (EPEP > EEEE) and Hypothesis 1b (EPEP > PEPE).

3.2. Does the sequencing of examples and problems affect learning and transfer?

When analyzing whether performance on the test tasks isomorphic to the training phase improved significantly from pretest to posttest, we found a main effect of Test Moment, $Z = 3.86, p < .001, r = 0.311$. Numerically, performance increased over time in all example conditions (see Table 1), but follow-up tests showed that only the EEEE Condition performed significantly better on the posttest than on the pretest ($Z = 3.02, p = .003, r = 0.551$). The other conditions did not show a significant increase (EPEP: $p = .061, r = 0.321$; PEPE: $p = .047, r = 0.397$; PPPP: $p = .029$).

To answer our second main question of whether there would be differences among instructional conditions on learning, we analyzed whether there were any differences among conditions regarding isomorphic posttest performance. We found a main effect of Instruction Condition, $H(3) = 20.63, p < .001$. In line with our expectations (i.e., Hypothesis 2: EEEE = EPEP > PEPE = PPPP), we found that participants in the EEEE Condition scored significantly higher than those in the PPPP Condition ($U = 189, p < .001, r = 0.551$). The results of other post-hoc comparisons were not in line with our expectations, however, because we found no significant differences between the EEEE and PEPE Condition ($p = .469, r = 0.097$), between the EPEP and PPPP Condition ($p = .218, r = 0.148$), and between the EPEP and PEPE Condition ($p = .131, r = 0.197$). Performance was even higher in the EEEE Condition compared to EPEP Condition ($U = 293.50, p = .003, r = 0.366$), and the PEPE Condition performed better than the PPPP Condition ($U = 254.50, p = .006, r = 0.356$).

To test our hypotheses regarding the effects on transfer, we analyzed the differences among conditions on the procedural transfer task (i.e., Hypothesis 4: EEEE = EPEP > PEPE = PPPP) and the conceptual transfer task (i.e., Hypothesis 5: EEEE = EPEP > PEPE = PPPP). Contrary to our hypotheses, we found no significant performance differences among conditions on the procedural transfer task ($p = .276$) and the conceptual transfer task ($p = .104$).

3.3. Does the sequencing of examples and problems affect invested mental effort and time-on-task in the training phase?

With regard to learning efficiency, we analyzed our results on self-reported effort invested in the training tasks and found a main effect of Instruction Condition, $H(3) = 51.48, p < .001$. In line with Hypothesis 3 (EEEE = EPEP < PEPE = PPPP), we found that students in the EEEE ($U = 971.50, p < .001, r = 0.730$) and EPEP Condition ($U = 964.50, p < .001, r = 0.535$) reported significantly lower effort investment than students in the PPPP Condition. The average reported effort

investment was also lower in the EEEE Condition than the PEPE Condition ($U = 571.50, p = .001, r = 0.449$), but – in contrast to our expectations – no significant difference was found between the EPEP and PEPE Condition ($p = .419, r = 0.105$). In addition, we found that students in the EEEE Condition reported significantly lower effort investment than those the EPEP Condition ($U = 807, p < .001, r = 0.501$), and students in the PEPE Condition invested significantly less effort than students in the PPPP Condition ($U = 741, p < .001, r = 0.588$).²

When exploring time-on-task in the training phase, we found a main effect of Instruction Condition, $H(3) = 59.70, p < .001$. The average time invested in the training phase was significantly shorter in the EEEE Condition than in the PEPE ($U = 691.50, p < .001, r = 0.722$), and PPPP Condition ($U = 976, p < .001, r = 0.737$). Surprisingly, students in the EEEE Condition ($U = 990, p < .001, r = 0.808$) and PEPE Condition ($U = 214, p = .001, r = 0.422$) invested significantly less time in the training tasks than students in the EPEP Condition. Other post-hoc comparisons were not significant ($ps > 0.012, rs < 0.322$).

3.4. Does the sequencing of examples and problems affect mental effort and time-on-task in the posttest phase?

As for the exploration of self-reported effort invested in solving the isomorphic posttest tasks, we found a significant main effect of Instruction Condition, $H(3) = 21.88, p < .001$. In line with our findings regarding effort invested in the training phase, reported effort investment while solving the isomorphic posttest tasks was significantly lower in the EEEE ($U = 817, p < .001, r = 0.478$), EPEP ($U = 845, p = .003, r = 0.363$), and PEPE Condition ($U = 690.50, p < .001, r = 0.492$) compared to the PPPP Condition. No other post-hoc comparisons were significant ($ps > 0.141, rs < 0.184$). We found the same pattern of results on students' effort invested in solving the procedural transfer task: A significant main effect of Instruction Condition, $H(3) = 15.86, p = .001$, and perceived effort investment was lower in the EEEE ($U = 759, p = .002, r = 0.388$), EPEP ($U = 812.50, p = .008, r = 0.319$), and PEPE Condition ($U = 659.50, p = .001, r = 0.436$) than in the PPPP Condition. None of the other comparisons were significant ($ps > 0.193, rs < 0.170$). Lastly, results showed a main effect of Instruction Condition regarding students' effort invested in the conceptual posttest task, $H(3) = 10.02, p = .018$. Invested effort was significantly lower in the PEPE Condition than in the PPPP Condition ($U = 621, p = .005, r = 0.365$). Again, no other post-hoc comparisons were significant ($ps > 0.020, rs < 0.281$).

As for the invested time-on-task during the posttest phase, a main effect of Instruction Condition was found for the isomorphic posttest tasks, $H(3) = 39.34, p < .001$. The average time-on-task was significantly longer in the EEEE Condition than in the EPEP ($U = 218.50, p < .001, r = 0.491$), PEPE ($U = 199, p = .002, r = 0.427$), and PPPP Condition ($U = 86, p < .001, r = 0.717$). In addition, students in the PEPE Condition invested significantly less time than students in the PPPP Condition ($U = 218.50, p = .001, r = 0.424$). The other post-hoc comparisons were not significant ($ps > 0.011, rs < 0.305$). Concerning the transfer tasks, we found a main effect of Instruction Condition for the procedural transfer task, $H(3) = 15.47, p < .001$.

² Upon a reviewer's request, we explored whether the mental effort invested in and performance on the two practice problems in the training phase differed between the EPEP and PEPE Condition. On the first practice problem (i.e., EPEP vs. PEPE), we found no performance difference ($p = .257, r = 0.148$), but the EPEP Condition reported significantly lower effort investment ($p = .024, r = 0.294$). On the second practice problem (i.e., EPEP vs. PEPE), the PEPE Condition attained greater performance ($p = .001, r = 0.428$) with less effort investment than the EPEP Condition ($p = .001, r = 0.436$). These results are not in line with the motivational hypothesis, as the initial disadvantage of starting with a practice problem disappeared (and even reversed) on the second practice problem.

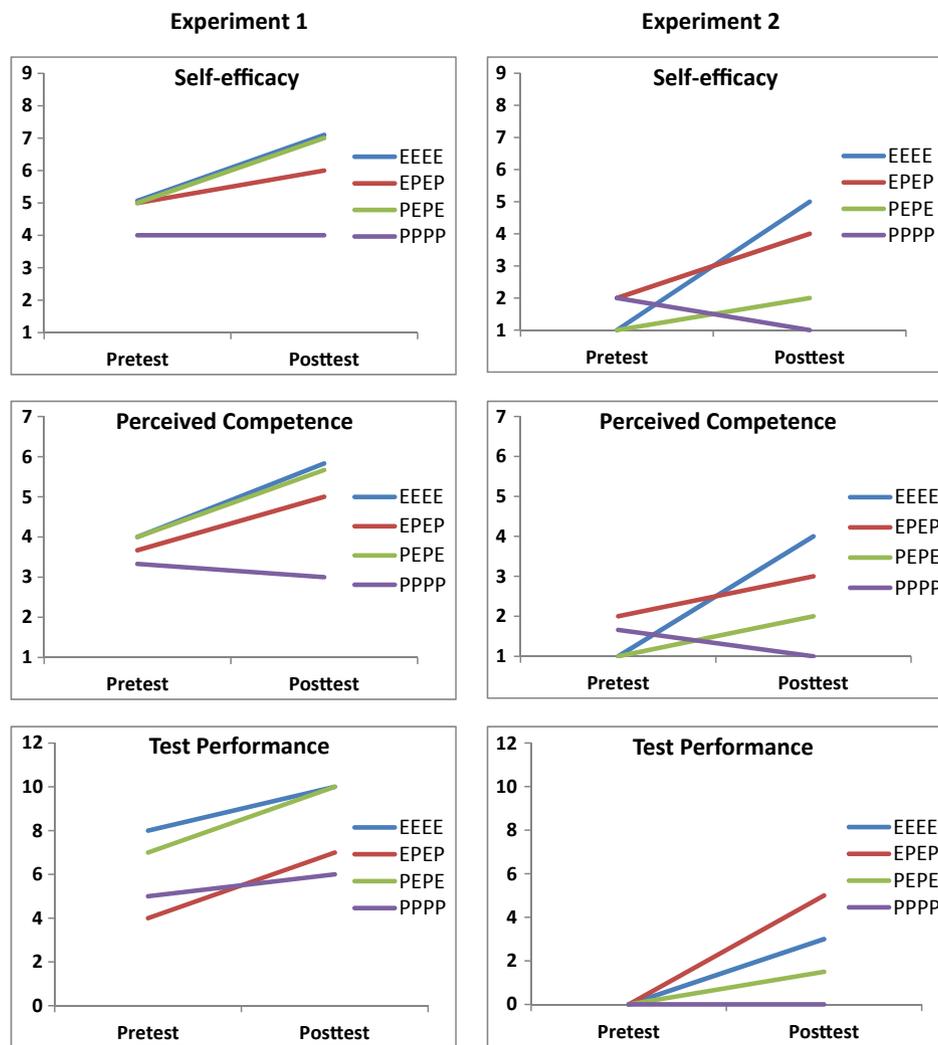


Fig. 1. Median scores on self-efficacy (top row; range 1–9), perceived competence (middle row; range 1–7), and performance on the isomorphic test tasks (range 0–16) on the pretest and immediate posttest in Experiment 1 (left) and 2 (right).

The average time-on-task was significantly longer in the EEEE Condition compared to the PPPP Condition ($U = 250.50, p < .001, r = 0.450$). No other post-hoc comparisons were significant ($ps > 0.015, rs < 0.292$). We found no main effect of Instruction Condition for the conceptual transfer task ($p = .057$).

4. Discussion

Our main aim was to investigate how different example and practice problem sequences (i.e., EEEE, EPEP, PEPE, and PPPP) would affect motivational (self-efficacy, perceived competence, and topic interest) and cognitive (effectiveness and efficiency) aspects of learning. The results were largely inconsistent with our hypotheses (see also Fig. 1 for a graphical overview of the median scores on self-efficacy, perceived competence, and performance on isomorphic test tasks). We predicted that the EPEP condition would show higher levels of self-efficacy, perceived competence, and topic interest than the EEEE Condition (Hypothesis 1a) and the PEPE condition (Hypothesis 1b), but that was not the case. Instead, we found no differences among conditions on topic interest. Moreover, all three example conditions showed significantly higher self-efficacy and perceived competence than the PPPP Condition, which is interesting and extends prior research showing that example study only can foster self-efficacy and perceived competence (e.g., Bandura, 1997; Crippen et al., 2009; Hoogerheide et al., 2014,

2018). Unexpectedly, given that problem solving after example study was implemented in early research on the worked example effect because it was considered to be more motivating (Sweller & Cooper, 1985; Trafton & Reiser, 1993), students in the EEEE Condition showed significantly higher self-efficacy than those in the EPEP Condition.

As for cognitive aspects of learning, we did not find the pattern of results that we expected based on the findings of Van Gog et al. (2011) either. In contrast to Hypothesis 2 (i.e., isomorphic test performance; EEEE/EPEP > PEPE/PPPP) and Hypothesis 3 (mental effort invested in the training phase; EEEE/EPEP < PEPE/PPPP), we found no significant differences between the EPEP and PEPE conditions on both variables. We did find, however, that studying EEEE was more effective (i.e., higher isomorphic posttest performance) and efficient (i.e., with lower effort investment in the training phase) for learning than studying EPEP and PPPP. Furthermore, studying PEPE was more effective and efficient than studying PPPP. Also, while the differences in posttest performance (i.e., mean performance of EPEP seemed higher than PPPP) were not significant, studying EPEP was more efficient than PPPP.

With regards to our exploration of mental effort invested in the isomorphic posttest tasks, our results suggest that all example conditions were more efficient than the PPPP condition. Our exploration of time-on-task in the training phase showed that the EEEE condition spent significantly less time on the learning phase compared to all the

Table 3
Mean (M), Standard Deviation (SD) and Median (Med) of Self-Efficacy (range 1–9), Perceived Competence (range 1–7), and Topic Interest (range 1–7) per Condition in Experiment 2.

	EEEE Condition			EPEP Condition			PEPE Condition			PPPP Condition		
	<i>M</i>	<i>SD</i>	<i>Med</i>									
Pretest												
Self-Efficacy	1.67	1.27	1.00	2.84	1.86	2.00	2.18	2.08	1.00	2.31	1.66	2.00
Perceived Competence	1.54	1.03	1.00	2.19	1.10	2.00	1.98	1.62	1.00	2.06	1.39	1.66
Topic Interest	3.53	0.88	3.43	4.20	1.26	4.29	3.49	1.35	3.43	3.46	1.13	3.29
Posttest												
Self-Efficacy	4.88	1.83	5.00	3.84	2.61	4.00	2.68	2.01	2.00	2.19	2.04	1.00
Perceived Competence	3.72	1.40	4.00	3.47	2.11	3.00	2.35	1.55	2.00	1.94	1.53	1.00
Topic Interest	3.69	0.98	3.93	3.72	1.42	3.57	3.25	1.14	3.36	3.39	1.29	3.57

other conditions. Although this seems to be an efficiency benefit, we must note that time-on-task was not experimenter-paced and watching video modeling examples probably took less time than solving the practice problems. In contrast, we found that the EEEE condition spent most time in the posttest phase compared to all other conditions. On the one hand, this might indicate that students in the EEEE condition may have needed more time to solve the procedure on the posttest because they did not have the possibility to practice during the training phase. On the other hand, it might be possible that students in the EEEE condition mastered the procedure so well and therefore spent more time on solving the posttest tasks. If you cannot figure out how to solve such problems during the training phase, you might drop out and spend less time on the remaining practice and posttest tasks. In sum, our findings show that all example conditions were more efficient than practice problem solving only, as equal or higher performance on isomorphic posttest problems was attained with less investment of effort in the training phase. Example study only was most efficient, requiring less effort (and time) investment in the training phase than all other conditions but attaining the highest scores on the isomorphic posttest tasks.

Our expectations regarding procedural transfer (i.e., Hypothesis 4; EEEE/EPEP > PEPE/PPPP) and conceptual transfer (i.e., Hypothesis 5; EEEE/EPEP > PEPE/PPPP) were not confirmed either. Our results showed no significant differences among conditions on procedural transfer and conceptual transfer.

A possible explanation for finding that EPEP was not more effective and efficient nor more motivating than PEPE, might lie in our participant sample. Although students were novices to this mathematical task, the fact that they were enrolled in a higher technical education program makes it likely that they had experience with learning similar types of mathematical problems that require complex mathematical calculations. This might have shielded those who had started with a practice problem from motivational issues, alleviating the negative effects of starting with a practice problem on motivation (i.e., self-efficacy and perceived competence). The fact that self-efficacy and perceived competence increased in all example conditions and did not differ between the EPEP condition and the PEPE condition indeed suggests that the confidence of those who were provided with problem-example pairs either was unaffected by starting with a practice problem or recovered quickly once provided with the opportunity of studying examples.

If this explanation is correct, the results might be different (i.e., EPEP > PEPE) with a sample of students who are less experienced with these types of mathematical tasks and who would generally be less confident about their mathematical abilities. Therefore, we reran the experiment with a sample of primary education teacher training students, who are normally much less experienced with mathematician problems such as learning how to approximate the definite integral by using the trapezoidal rule. Our hypotheses were identical to those in Experiment 1 (see Section 1.1).

5. Experiment 2

5.1. Method

5.1.1. Participants and design

Participants were 81 first year students from two Primary Education Teacher Training programs from two Dutch Universities of Applied Sciences ($M_{age} = 18.98$, $SD = 1.64$; 17 male, 65 female). Students could earn a monetary reward (eleven 20 euro bills were raffled among participants). Following the design of Experiment 1, students were randomly assigned to one of four conditions: (1) examples only ($n = 24$; EEEE), (2) example-problem pairs ($n = 19$; EPEP), (3) problem-example pairs ($n = 22$; PEPE), or (4) practice problems only ($n = 16$; PPPP). In this experiment, only a pretest, training phase, and (immediate) posttest were used.

5.1.2. Materials, procedure, and data analysis

The materials, procedure, and data analysis were the same as in Experiment 1. The reliability of the test tasks was measured again and showed the following Cronbach's alpha values: 0.81 for pretest tasks, 0.33 for isomorphic posttest tasks, and 0.39 for the conceptual transfer task. The only difference with Experiment 1 was that Experiment 2 was run in seven group sessions in a computer classroom with 5–16 participants per session instead of eight group sessions in a computer classroom with 5–25 participants present per session in Experiment 1.

6. Results

Again, the majority of the variables were not normally distributed, so we analyzed the data with nonparametric tests (cf. Experiment 1). Relevant descriptive statistics of self-efficacy, perceived competence, and topic interest scores are presented in Table 3, and performance scores, mental effort scores and time-on-task scores are presented in Table 4.

We checked for prior knowledge differences among conditions. Kruskal-Wallis test showed that there were no significant differences among conditions in terms of (pretest) performance, $H(3) = 3.35$, $p = .341$, perceived competence, $H(3) = 5.44$, $p = .142$, and topic interest, $H(3) = 5.40$, $p = .145$. We did, however, find significant differences among conditions on the pretest scores of self-efficacy, $H(3) = 9.98$, $p = .019$, and post-hoc test revealed that pretest scores of self-efficacy were significantly lower in the EEEE Condition than in the EPEP Condition ($U = 347$, $p = .002$, $r = 0.473$).

6.1. Does the sequencing of examples and problems affect self-efficacy, perceived competence, and topic interest?

Again, we started with the analysis of whether students' self-efficacy increased from before to after the training phase. We found a main

Table 4

Mean (M), Standard Deviation (SD), and Median (Med) of Pretest (range 0–16), Training Performance (range 0–24), Isomorphic Tasks Performance (range 0–16), Procedural Transfer (range 0–8), Conceptual Transfer (range 0–9), Mental Effort (range 1–9), and Time-on-Task per Condition in Experiment 2.

	EEEE Condition			EPEP Condition			PEPE Condition			PPPP Condition		
	M	SD	Med									
Pretest												
Performance	0.63	1.71	0.00	1.47	3.82	0.00	1.27	3.06	0.00	1.75	2.74	0.00
Training												
Performance				2.82	1.95	3.00	1.98	1.98	1.50	1.28	1.15	1.25
Mental Effort	4.79	1.83	4.50	5.09	1.93	5.00	6.40	1.94	6.25	7.99	1.35	8.50
Time-on-Task	3.58	1.56	3.63	8.16	3.09	7.25	5.03	2.00	4.88	4.75	2.86	3.63
Posttest												
Isomorphic Tasks	3.00	2.64	3.00	4.26	3.28	5.00	1.91	1.97	1.50	0.56	1.03	0.00
Procedural Transfer	0.83	0.28	0.00	0.37	0.68	0.00	0.41	0.80	0.00	0.19	0.54	0.00
Conceptual Transfer	1.46	1.64	1.00	1.63	1.54	1.00	1.41	1.30	1.00	1.13	1.09	1.00
Mental Effort												
Isomorphic Tasks	6.81	1.90	7.00	6.87	2.31	8.00	7.07	2.35	7.50	7.81	2.22	9.00
Procedural Transfer	8.04	1.57	9.00	7.11	2.51	8.00	7.86	2.44	9.00	8.00	2.22	9.00
Conceptual Transfer	7.00	1.87	7.00	6.58	2.63	8.00	7.59	2.46	9.00	7.50	2.13	8.00
Time-on-Task												
Isomorphic Tasks	7.23	5.18	6.75	6.74	5.11	5.50	2.80	2.87	2.50	0.91	1.20	0.25
Procedural Transfer	0.50	0.93	0.00	3.68	4.84	1.00	1.59	2.38	0.50	0.81	1.38	0.00
Conceptual Transfer	3.08	2.28	3.00	5.26	4.07	4.00	2.68	2.06	2.50	3.06	2.17	3.00

effect of Test Moment, $Z = 4.61, p < .001, r = 0.512$. Follow-up tests showed that the self-efficacy scores of the EEEE Condition ($Z = 3.96, p < .001, r = 0.807$) increased significantly over time, whereas the EPEP ($p = .044, r = 0.463$), PEPE ($p = .121, r = 0.331$), and the PPPP Condition ($p = .729, r = 0.087$) did not show a significant increase over time. Regarding the main question of whether there would be differences among instructional conditions on reported self-efficacy measured after the training phase, we found a main effect of Instruction Condition, $H(3) = 16.48, p = .001$. In contrast to Hypothesis 1a (EPEP > EEEE) and Hypothesis 1b (EPEP > PEPE), we found no differences between the EPEP and EEEE Condition ($p = .200, r = 0.195$) and the EPEP and PEPE Condition ($p = .152, r = 0.224$). Further explorations showed that self-efficacy was significantly higher in the EEEE Condition than in the PEPE ($U = 121, p = .001, r = 0.473$) and PPPP Condition ($U = 65, p < .001, r = 0.570$). No other post-hoc comparisons were significant ($ps > 0.056, rs < 0.007$). Note that these results have to be interpreted with caution, because there were pre-existing differences among the conditions on self-efficacy before the training phase (i.e., EEEE > EPEP).

Perceived competence showed the same pattern of results. We found a main effect of Test Moment, $Z = 4.64, p < .001, r = 0.516$, indicating that perceived competence increased significantly over time in the EEEE Condition ($Z = 4.02, p < .001, r = 0.821$) but not in the EPEP ($p = .028, r = 0.505$), PEPE, ($p = .151, r = 0.306$), and PPPP Condition ($p = .593, r = 0.134$). We also found a main effect of Instruction Condition on perceived competence measured after the training phase, $H(3) = 15.08, p = .002$. These results were not in line with our expectations (Hypothesis 1a: EPEP > EEEE; Hypothesis 1b: EPEP > PEPE), because we did not find any differences between the EPEP and EEEE Condition ($p = .641, r = 0.071$) and EPEP and PEPE Condition ($p = .063, r = 0.290$). Further explorations showed that scores were significantly higher in the EEEE Condition than in the PEPE ($U = 131, p = .003, r = 0.436$) and PPPP Condition ($U = 77.5, p = .001, r = 0.508$). No other post-hoc comparisons were significant ($ps > 0.015, rs < 0.002$).

Concerning topic interest, we found that scores did not increase over time, since there was no significant main effect of Test Moment ($p = .196$). Unlike our expectations (i.e., Hypothesis 1a; EPEP > EEEE and Hypothesis 1b; EPEP > PEPE), we found no main effect of Instruction Condition ($p = .562$), meaning that there were no differences among conditions on topic interest measured after the training phase.

6.2. Does the sequencing of examples and problems affect learning and transfer?

Subsequently, we checked whether performance on the test tasks isomorphic to the training phase improved significantly from pretest to posttest. We found a main effect of Test Moment, $Z = 3.76, p < .001, r = 0.418$. Follow-up tests showed that scores increased over time in the EEEE ($Z = 3.51, p < .001, r = 0.717$) and EPEP Condition ($Z = 2.56, p = .010, r = 0.586$), but not in the PEPE ($p = .052, r = 0.414$), and PPPP Condition ($p = .173, r = 0.341$). Regarding our second main aim, namely to examine if there are any differences among conditions on the isomorphic posttest tasks, our results showed a main effect of Instruction Condition, $H(3) = 17.82, p < .001$. In line with our expectations (i.e., Hypothesis 2: EEEE = EPEP > PEPE = PPPP), participants scored significantly higher in the EEEE ($U = 71, p < .001, r = 0.414$) and EPEP Condition ($U = 55.50, p = .001, r = 0.576$) than in the PPPP Condition. However, we did not find any differences between the EEEE and PEPE Condition ($p = .144, r = 0.215$) and EPEP and PEPE Condition ($p = .019, r = 0.366$). As expected, no differences were found between the EEEE and EPEP Condition ($p = .166, r = 0.211$) and PEPE and PPPP Condition ($p = .033, r = 0.377$).

Surprisingly, no significant performance differences were found among conditions on the procedural transfer task ($p = .257$) and the conceptual transfer task ($p = .841$). Hence, the results on our transfer measures also contrasted Hypothesis 4 (EEEE = EPEP > PEPE = PPPP) and Hypothesis 5 (EEEE = EPEP > PEPE = PPPP).

6.3. Does the sequencing of examples and problems affect mental effort and time-on-task in the training phase?

When analyzing self-reported effort investment during the training phase as measure of efficiency, we found a main effect of Instruction Condition, $H(3) = 28.28, p < .001$. In line with Hypothesis 3 (EEEE = EPEP < PEPE = PPPP), reported effort investment was significantly lower in the EEEE ($U = 349, p < .001, r = 0.689$) and EPEP Condition ($U = 277.50, p < .001, r = 0.707$) compared to the PPPP Condition. Moreover, students in the EEEE Condition reported significantly lower effort investment than the PEPE Condition ($U = 394, p = .004, r = 0.422$), but no significant differences were found between the EPEP and PEPE Condition ($p = .059, r = 0.295$). As expected, we found no significant differences in effort investment during the training

phase between the EEEE and EPEP Condition ($p = .470$, $r = 0.110$), but we did find the PEPE Condition to report significant less effort than the PPPP Condition while solving the tasks in the training phase ($U = 271$, $p = .004$, $r = 0.458$).³

Subsequently, we analyzed the average time spent on the tasks during the training phase and found a main effect of Instruction Condition, $H(3) = 26.17$, $p < .001$. We found that average time-on-task while solving the training tasks was significantly longer in the EPEP Condition than in the EEEE ($U = 430.50$, $p < .001$, $r = 0.757$), PEPE ($U = 69$, $p = .001$, $r = 0.573$), and PPPP Condition ($U = 64$, $p = .003$, $r = 0.493$). We found no differences between the EEEE and PPPP Condition ($p = .774$, $r = 0.122$) and PEPE and PPPP Condition ($p = .455$, $r = 0.122$).

6.4. Does the sequencing of examples and problems affect mental effort and time-on-task in the posttest phase?

While exploring the differences among conditions on reported effort investment when solving the posttest tasks, we found no differences among conditions on the isomorphic posttest tasks ($p = .165$), procedural transfer task ($p = .238$), and conceptual transfer task ($p = .201$). We did find a main effect of Instruction Condition for average time invested in the isomorphic posttest tasks, $H(3) = 27.64$, $p < .001$. The average time-on-task was significantly longer in the EEEE Condition than in the PEPE ($U = 117$, $p = .001$, $r = 0.478$) and PPPP Condition ($U = 33$, $p < .001$, $r = 0.699$), and significantly longer in the EPEP Condition compared to the PEPE ($U = 108$, $p = .008$, $r = 0.410$) and PPPP Condition ($U = 41.50$, $p < .001$, $r = 0.629$). No other post-hoc comparisons were significant ($ps > 0.026$, $rs < 0.370$). Concerning the transfer tasks, we found a main effect of Instruction Condition for the procedural transfer task, $H(3) = 9.53$, $p = .023$. Average time-on-task was, however, only significantly longer in the EEEE Condition than in the EPEP Condition ($U = 337.50$, $p = .004$, $r = 0.442$). No other post-hoc comparisons were significant ($ps > 0.048$, $rs < 0.353$). We found no main effect of Instruction Condition for the conceptual transfer task ($p = .086$).

7. Discussion

The main aim of this experiment was to investigate whether the results of Experiment 1 (i.e., EPEP = PEPE) on performance, mental effort invested in the learning tasks, and motivation would be different (i.e., EPEP > PEPE) with a sample of students who are less experienced with these types of mathematical tasks and who would generally be less confident about their mathematical abilities (i.e., teacher training students). Despite the different sample (i.e., primary education teacher training students), the results of Experiment 2 also did not show evidence in favor of our hypotheses (see also Fig. 1); we expected that the EPEP condition would show higher levels of self-efficacy, perceived competence, and topic interest than the EEEE Condition (Hypothesis 1a) and the PEPE condition (Hypothesis 1b), but we found no significant differences between these conditions. As in Experiment 1, no differences were found among conditions regarding topic interest. Results did show that self-efficacy and perceived competence were

³ Like for Experiment 1, we explored whether the mental effort invested in and performance on the two practice problems in the training phase differed between the EPEP and PEPE Condition. As one would expect, the EPEP Condition performed significantly better on the first practice problem ($p = .001$, $r = 0.513$) and invested less effort ($p < .001$, $r = 0.574$) than the PEPE Condition. However, on the second practice problem, we found no difference between the two conditions in terms of performance ($p = .178$, $r = 0.210$) or effort investment ($p = .813$, $r = 0.037$). Again, there was an advantage in favor of the EPEP Condition at the start, but no sign of a lasting disadvantage of starting with a (failed) practice problem solving attempt for those in the PEPE Condition.

significantly higher in the EEEE condition than in the PEPE and PPPP condition (as in Experiment 1).

With regard to cognitive aspects of learning, we partially replicated the results from Van Gog et al., 2011. In line with our expectations on isomorphic posttest performance (i.e., Hypothesis 2; EEEE/EPEP > PEPE/PPPP) and invested mental effort in the training phase (i.e., Hypothesis 3; EEEE/EPEP < PEPE/PPPP), we found that starting with an example (EEEE and EPEP) was more effective and efficient for learning than problem solving only (PPPP). Also, while the differences in isomorphic posttest performance (i.e., mean performance of PEPE seemed higher than PPPP) were not significant, studying PEPE was more efficient than PPPP. In contrast to our expectations, we did not find any significant differences on both variables between the EEEE and EPEP and between the EPEP and PEPE conditions.

When exploring mental effort on the isomorphic posttest tasks, we found no differences among conditions. Our exploration of time-on-task revealed that the EPEP Condition spent significantly more time in the training phase than all the other conditions. In addition, both conditions starting with an example (i.e., EEEE, EPEP) spent significantly more time on the isomorphic posttest tasks than the conditions starting with a problem (i.e., PEPE, PPPP). This might indicate that, considering the performance on the isomorphic posttest tasks, students understood the procedure and therefore spent more time on solving the posttest tasks. In sum, our findings show that all example conditions were more efficient than practice problem solving only, as equal or higher performance on isomorphic posttest problems was attained with less investment of effort in the training phase. Finally, our expectations regarding procedural transfer (i.e., Hypothesis 4; EEEE/EPEP > PEPE/PPPP) and conceptual transfer (i.e., Hypothesis 5; EEEE/EPEP > PEPE/PPPP) were not confirmed. Our results showed no significant differences among conditions on procedural transfer and conceptual transfer.

8. General discussion

Two experiments were conducted to conceptually replicate and extend the study by Van Gog et al. (2011) in order to investigate how different example study and practice problem solving sequences would affect learning and motivation. Our main aim was to investigate how example study only (EEEE), example-problem pairs (EPEP), problem-example pairs (PEPE), and problem-solving only (PPPP) sequences would affect motivational (i.e., self-efficacy, perceived competence, and topic interest) and cognitive (i.e., effectiveness and efficiency) aspects of learning.

First, we were interested in looking from a motivational perspective at the finding by Van Gog et al. (2011) that EEEE did not differ from EPEP in terms of learning outcomes. We expected EPEP to be more motivating for students than passively studying EEEE, as suggested -but not tested- by Sweller and Cooper (1985; see also Trafton & Reiser, 1993). Interestingly, our findings showed that EEEE was not less motivating than EPEP. In Experiment 1, students in the EEEE condition even showed higher self-efficacy (and better performance) than students in the EPEP condition. This finding might indicate that (at least when short training phases are used), the benefits of engaging in practice problem solving instead of further example study, seem limited for both learning and motivation. In general, this calls for further research into the role of practice problem solving in example-based learning, especially as findings from Baars, Van Gog, De Bruin, and Paas (2014) and Van der Meij et al. (2018) showed that even additional problem solving practice did not have a positive effect on learning. However, all of those studies used relatively short training phases. It is possible that motivational differences will start to arise and affect learning when training phases are longer and consist of more training tasks, as students might get bored with studying examples only. The effects of longer sequences should therefore be addressed in future research.

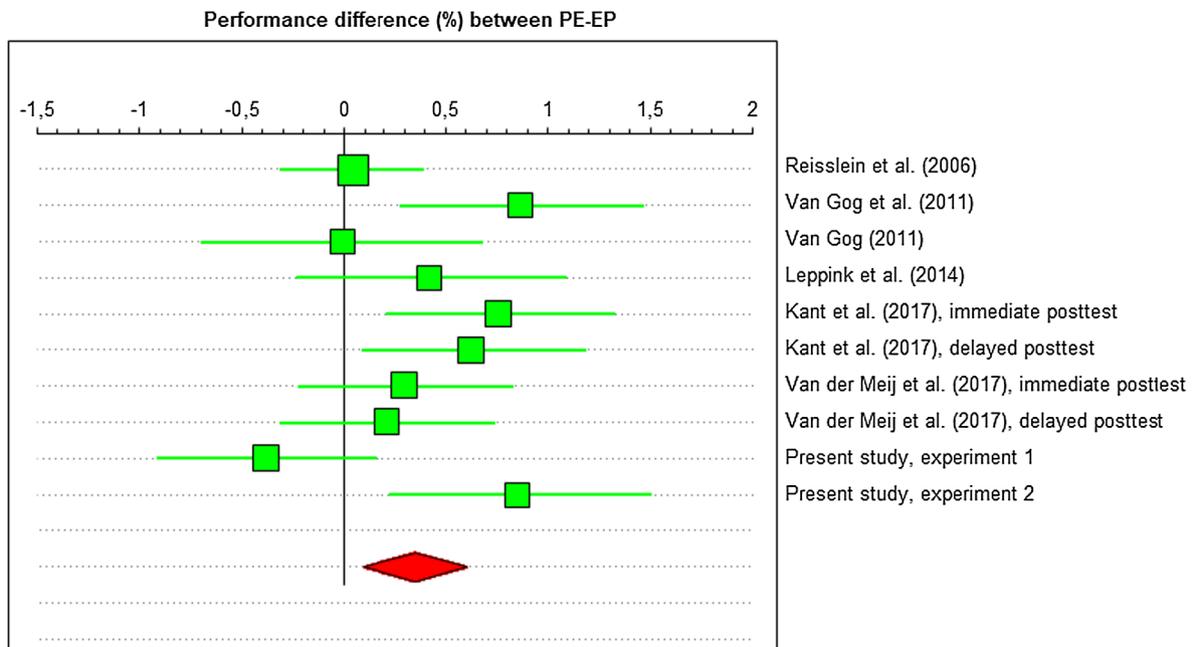


Fig. 2. Results of the meta-analysis.

Second, we aimed to investigate whether motivational aspects of learning could account for the finding by Van Gog et al. (2011) that EPEP led to better test performance with less effort investment in the training phase than PEPE. However, in contrast to our expectations, we did not replicate these findings across two experiments with different populations. We also did not find any significant differences between EPEP and PEPE concerning students' self-efficacy and perceived competence. We thought we had a potential explanation for this null-finding in Experiment 1, because those higher technical education students, despite being novices, presumably had experience with similar types of mathematical problems and these problems were relevant for them (so they would not get frustrated that easily). However, the results of Experiment 2 indicated that this explanation does not hold. In Experiment 2, we again failed to find significant differences in learning or motivation between the EPEP and PEPE condition, even though the sample consisted of student teachers for whom the tasks were less relevant, who had less experience with these types of mathematical tasks, and who felt less confident about their mathematical abilities, as evidenced by the pretest scores on performance, self-efficacy, and perceived competence. Note that, numerically, the EPEP and PEPE conditions did differ on isomorphic posttest performance (i.e., EPEP > PEPE) and average invested mental effort in the training phase as a whole (i.e., EPEP < PEPE). Importantly, exploratory analyses of students' performance on and effort invested in the two practice problems also suggest that starting with a practice problem did not have a demotivating effect in either experiment. Whereas EPEP was more effective and/or efficient than PEPE on the first practice problem (i.e., equal or higher performance attained with less effort), we found no performance or effort advantage of EPEP over PEPE on the second practice problem.

Given that our findings regarding the EPEP vs. PEPE comparison were not in line with other studies, and that the direction of the difference between conditions seemed to vary in our experiments (i.e., we found a non-significant, medium-sized [according to the Cohen's d criterion] negative effect of EP on learning in Experiment 1, and a large but non-significant, positive effect of EP on learning in Experiment 2), we entered all EP-PE comparisons from the published studies we are aware of in a small-scale random effect meta-analysis, to get a better estimate of the EP-PE effect size and its heterogeneity (see Fig. 2). We used Cumming's (2012) ESCI software (www.thenewstatistics.com).

This small-scale meta-analysis showed a significant, medium-sized advantage of EP over PE (Cohen's d of the meta-analytic effect was 0.350). This advantage has to be interpreted with caution because there was substantial heterogeneity among the comparisons (i.e., heavy variation in the results among studies). That is, of the 10 comparisons, 8 showed an EP advantage and 2 showed a PE advantage, and effect sizes varied from -0.397 (the first experiment in this study) to 0.862 (the study by Van Gog et al., 2011).

A possible reason why there is substantial variation in results among the example-problem vs. problem-example comparisons (i.e., EP = PE vs. EP > PE) might lie in study characteristics that vary across studies, such as the learning material, target group, pair type (identical vs. isomorphic pairs), sequence length (two vs. four training tasks), and example format (worked vs. video modeling examples). For instance, when viewing the results of the studies in the small-scale meta-analysis that used video modeling examples, it seems that almost all of these studies did not find any learning differences between EP and PE,⁴ whereas the studies that used worked examples did find EP to be more effective than PE. One could assume that after starting with a problem, demotivated learners would not pay attention to worked examples, but would study video modeling examples. Worked examples can be overwhelming because all the information is presented simultaneously and it might be easy to ignore written text. Video modeling examples, however, present information step-by-step and the combination of dynamic visual information and the model's narration takes the learner by the hand. Thus, it is possible that students might find it more motivating to study a video modeling example after starting with a practice problem than studying a worked example, which might partially explain the differences in findings.

Another factor that might explain the differences in findings is that problem-example pairs may become more effective when the number of training tasks increases. When two tasks are presented in the training phase, a failed practice problem solving attempt means that students in the PE condition only effectively have one task to learn from, whereas those in the EP condition have the opportunity to first build a schema with the example and then learn again from problem solving (and

⁴ Except for the study of Kant et al. (2017) that did find EP to be more effective than PE.

repeat this again). Using four (or more) training tasks means that learners in the PE condition have more opportunities for learning. Note that we checked all studies used in the small-scale meta-analysis on whether one of these factors could explain the variation in results, and found that none of these factors could solely account for the mixed findings. Future research is recommended to investigate which (combination of) factors might moderate the EP-PE effect.

A limitation of the present study is the sample size of Experiment 2. While a power analysis indicated that our sample size was more than sufficient to reliably detect the effect sizes found by Van Gog et al. (2011), our small-scale meta-analysis suggests that the EP-PE effect might be significantly smaller than previously believed. The sample size of Experiment 2 was not sufficient to reliably detect small to medium-sized effects, and therefore, the results of Experiment 2 should be interpreted with caution. A second limitation of this study is that the reliability of our test tasks (i.e., isomorphic posttest tasks and conceptual transfer task) was rather low, particularly in Experiment 2. A possible explanation for the low reliability of the test tasks might be the low scores on the isomorphic posttest tasks and transfer tasks. Together with the high mental effort scores in the training phase, this might indicate that these test tasks were more difficult for the students in Experiment 2 than the students in Experiment 1. Another explanation may lie in the small number of tasks that was used to measure isomorphic posttest performance (i.e., 2 tasks) and conceptual transfer (i.e., 5 open-ended questions).

Nevertheless, our findings are very interesting and relevant for educational practice, where example study and practice problem solving are frequently used to acquire new knowledge and skills (e.g., Atkinson & Renkl, 2007; Van Gog, Verveer, & Verveer, 2014). The results of this study suggest that, when short training phases are used,

studying examples (only) is more preferable than problem solving only. These results complement previous findings on the ‘worked example effect’, that have (also) shown example study to result in higher learning outcomes with less reported effort investment than problem solving only (for reviews, see Atkinson et al., 2000; Renkl, 2014; Sweller et al., 2011; Van Gog & Rummel, 2010). A novel finding, however, is that example study also enhances motivational aspects of learning, such as believing in one’s own competence when mastering a task, whereas problem solving only does not positively affect students’ motivation at all. These results could be used by teachers during their classroom practice when instructing novices on new knowledge or skills, or as guidelines for instructional designers when designing new learning materials (such as books or online learning environments). In addition, students could be given the advice to study examples (only) instead of practice problem solving only when learning new knowledge or skills during self-study (for example when selecting own training tasks in online self-paced learning environments).

Acknowledgments

The authors would like to thank the math teachers of the study programs Mechanical Engineering and Electrical Engineering of Avans University of Applied Sciences and the math teachers of the Primary School Teacher Training Institute of Avans University of Applied Sciences and HZ University of Applied Sciences for facilitating this study. They also would like to thank Jos van Weert, Rob Müller, and Bert Hoeks for their help in developing the materials, Niek van Hoof-Verhagen, Mariette Vissers, Susan van Brussel, Divna van Driel, and Yvette van den Bersselaar for their help with the data collection and Susan Ravensbergen for her help with scoring the data.

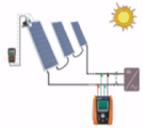
Appendix A. Example of a video modeling example

1

De trapeziumregel
Energimeting

"In this video, we start working with the mathematics topic 'the trapezoidal rule'. The trapezoidal rule is a formula which you can use to approach an area under a graph. We will show you how this works with an example called 'energy measurement'."

2



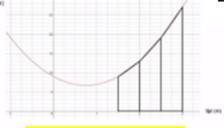
"Jalil has bought a solar cell and wants to know how much energy the solar cell supplies during a certain amount of time. Jalil uses an energy meter to examine how much energy the solar cell produced at a specific moment"

3



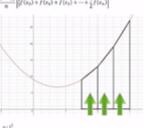
"Jalil has measured the energy at different time points and plotted the results in a graph. The time (in minutes) is plotted on the horizontal axis and the power the solar cell supplies (Joule per minute) is plotted on the vertical axis of the graph. By calculating the area under the graph, Jalil can determine how much energy the solar cell has produced during a certain amount of time"

4



"Approaching the surface can be done by using the trapezoidal rule. At this moment, the formula that belongs to the trapezoidal rule is shown on your screen."

5



"The trapezoidal rule divides the area under a graph into "strips". By adding up the surface of the "strips", you can approach the total surface under the graph. To approach the area under the graph, you need the following information:

a: this is the left x value of the area that has to be approached; in this example 1 1/2

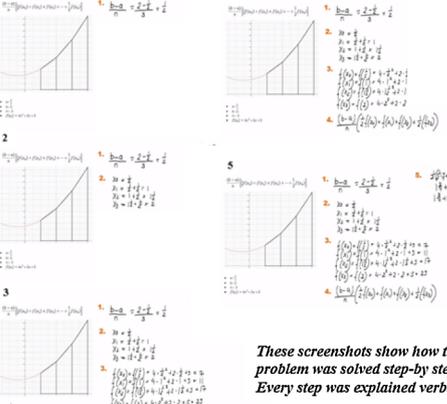
b: this is the right x value of the area that has to be approached; in this example 3

n: this is the number of "strips" in which the area is divided; in this example 3

xi: this is the x-value that belongs to the left- or right border of "strip" and it is calculated using the function that is given under the image. We will now continue to approach the surface under the graph using these different 'components'.

We will now show you how to approach the area under the graph using the information that is given."

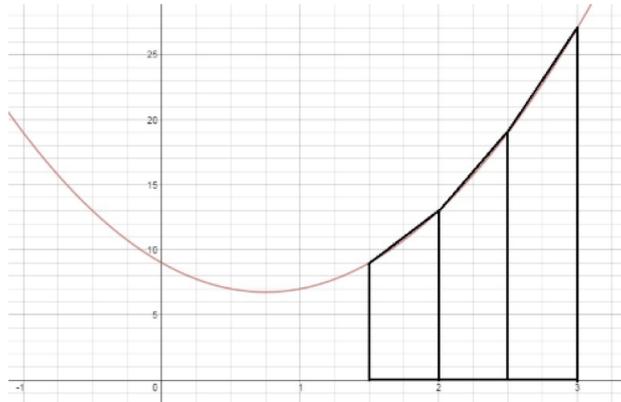
6



These screenshots show how the problem was solved step-by-step. Every step was explained verbally.

Appendix B. Example of a practice problem

Jalil has bought a solar cell and wants to know how much energy the solar cell supplies during a certain amount of time. Jalil has used an energy meter to examine how much energy the solar cell produces during a specific amount of time. Jalil has measured the energy at different time points and plotted the results in a graph. The time (in minutes) is plotted on the horizontal axis and the power the solar cell supplies (Joule per minute) is plotted on the vertical axis of the graph. By calculating the area under the graph, Jalil can determine how much energy the solar cell has produced during a certain amount of time.



Approaching the area under the graph can be done by using the trapezoidal rule:

$$\frac{(b - a)}{n} \left[\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \dots + \frac{1}{2}f(x_n) \right]$$

The trapezoidal rule divides the area under a graph into “strips”. By adding up the surface of the “strips”, you can approach the total area under the graph. To approach the area under the graph, you need the following information:

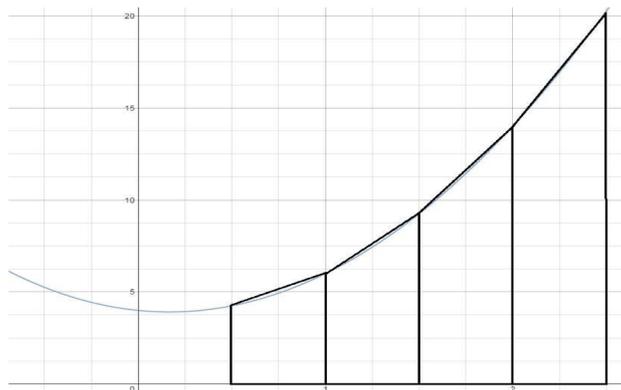
- a: this is the left x value of the area that has to be approached;
- b: this is the right x value of the area that has to be approached;
- n: this is the number of “strips” in which the area is divided;
- xi: this is the x-value that belongs to the left- or right border of a “strip” and it is calculated using the following function:

Approach the area under the graph using the information that is given. Write down all your intermediate steps and calculations.

Appendix C. Examples of isomorphic and transfer posttest tasks

C.1. Isomorphic posttest task

Rachel is an intern at a factory that produces different kinds of perfume. At one point, Rachel's supervisor asks her to examine how many liters of perfume is produced of the brand 'Scents' in two days. Rachel has measured this and plotted the results in a graph. The time (in days) is plotted on the horizontal axis and the liters (liter per day) are plotted on the vertical axis. By approaching the area under the graph, Rachel can determine how much liter has been produced during a certain amount of time.



Approaching the area under the graph can be done by using the trapezoidal rule:

$$\frac{(b - a)}{n} \left[\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \dots + \frac{1}{2}f(x_n) \right]$$

The trapezoidal rule divides the area under a graph into “strips”. By adding up the surface of the “strips”, you can approach the total area under

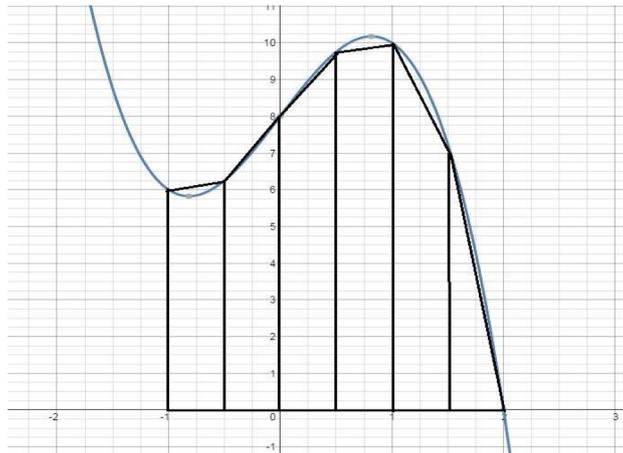
the graph. To approach the area under the graph, you need the following information:

- a: this is the left x value of the area that has to be approached;
- b: this is the right x value of the area that has to be approached;
- n: this is the number of “strips” in which the area is divided;
- xi: this is the x-value that belongs to the left- or right border of a “strips” and it is calculated using the following function:

Approach the area under the graph using the information that is given. Write down all your intermediate steps and calculations.

C.2. Procedural transfer task

It takes energy to stop an elevator at a certain level. This energy is proportional to the distance between the current and desired position. Jimmy wants to determine how much energy is used to stop the lift three levels higher by measuring the distance during a certain amount of time. Jimmy has plotted the results in a graph. The time (in seconds) is plotted on the horizontal axis and the distance (in meters) is plotted on the vertical axis. By approaching the area under the graph, Jimmy can determine the energy that is needed.



Approaching the area under the graph can be done by using the Simpson rule:

$$\int_a^b f(x)dx \approx \frac{b-a}{6} \left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right]$$

The Simpson rule divides the area under a graph into “strips”. By adding up the surface of the “strips”, you can approach the total area under the graph. To approach the area under the graph, you need the following information:

- a: this is the left x value of the area that has to be approached;
- b: this is the right x value of the area that has to be approached;
- n: this is the number of “strip” in which the area is divided;
- xi: this is the x-value that belongs to the left- or right border of a “strip” and it is calculated using the following function:

Approach the area under the graph using the information that is given. Write down all your intermediate steps and calculations.

C.3. Conceptual transfer item

Study the graph below (this is a part of a parabola):



You can approach the area under this graph with help of the trapezoidal rule in two ways:

- A: Left border 2 and right border 4
- B: Left border 7 and right border 9

Which surface will approach the exact surface at best? Choose one of the options and explain your answer.

Appendix D. Topic interest scale

Items (in Dutch)	Translation (in English)
1. Ik vind de opdrachten over de trapeziumregel erg interessant	1. I think that the tasks about the trapezoidal rule are very interesting
2. Weten hoe de trapeziumregel werkt is niet belangrijk voor mij	2. Knowing how the trapezoidal rule works is not important to me
3. Het is gemakkelijk om mijn aandacht bij de opdrachten over de trapeziumregel te houden	3. It is easy to stay focused on tasks about the trapezoidal rule
4. Ik wil meer te weten komen over de trapeziumregel	4. I am keen to learn more about the trapezoidal rule
5. Ik vind de opdrachten over de trapeziumregel niet boeiend	5. I think that the tasks about the trapezoidal rule are uninteresting
6. Ik vind andere wiskunde onderwerpen relevanter dan de trapeziumregel	6. I think that other mathematics topics are more relevant than the trapezoidal rule
7. Ik vind dat tijdens de wiskundelessen aandacht besteed moet worden aan de trapeziumregel	7. I think that during math class, more attention should be paid to the trapezoidal rule

Appendix E. Means, standard deviation, and medians of delayed posttest

Mean (M), Standard Deviation (SD), and Median (Med) of Motivation, Performance, Mental Effort (range 0 to 9), and Time-on-Task per Condition of the Delayed Posttest in Experiment 1.

	EEEE Condition			EPEP Condition			PEPE Condition			PPPP Condition		
	M	SD	Med	M	SD	Med	M	SD	Med	M	SD	Med
Motivation												
Self-efficacy	5.78	1.31	6.00	5.32	2.14	5.00	5.94	1.95	6.00	4.26	2.13	5.00
Perceived Competence	4.57	1.11	4.67	4.12	1.53	4.00	4.77	1.20	4.83	3.55	1.43	3.33
Topic Interest	3.77	1.03	3.79	3.47	1.16	3.43	3.19	0.71	3.14	3.71	0.92	3.86
Performance												
Isomorphic Tasks	10.39	3.52	10.00	8.48	4.49	8.00	9.56	4.15	10.50	8.23	3.95	8.00
Procedural Transfer	2.50	2.71	1.00	2.16	3.12	1.00	2.50	3.31	1.00	2.32	1.98	1.00
Conceptual Transfer	4.72	2.19	4.50	4.32	2.53	5.00	5.69	1.92	5.50	3.47	2.15	3.00
Mental Effort												
Isomorphic Tasks	4.17	1.51	4.50	4.46	1.97	4.50	3.66	1.77	3.25	5.66	2.03	6.00
Procedural Transfer	4.22	2.10	3.00	4.96	2.13	5.00	4.50	2.37	4.00	4.84	2.41	5.00
Conceptual Transfer	3.56	1.46	3.00	4.04	2.13	3.00	3.38	1.54	3.00	4.32	1.81	5.00
Time-on-Task												
Isomorphic Tasks	12.03	3.66	11.25	9.20	3.77	9.50	10.25	3.79	9.75	8.37	3.58	8.00
Procedural Transfer	7.17	3.03	7.00	7.80	4.99	7.00	6.63	2.45	6.50	6.55	3.58	5.00
Conceptual Transfer	5.89	1.60	6.00	6.36	2.98	6.00	7.56	3.44	7.00	6.23	2.67	6.00

References

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning and the psychological processes that mediate their relationship. *Journal of Educational Psychology, 94*(3), 545–561. <https://doi.org/10.1037/0022-0663.94.3.545>.

Atkinson, R. K., & Renkl, A. (2007). Interactive example-based learning environments: using interactive elements to encourage effective processing of worked examples. *Educational Psychology Review, 19*(3), 375–386. <https://doi.org/10.1007/s10648-007-9055-2>.

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*(2), 181–214. <https://doi.org/10.3102/00346543070002181>.

Baars, M., Van Gog, T., De Bruin, A., & Paas, F. (2014). Effects of problem solving after worked example study on primary school children’s monitoring accuracy. *Applied Cognitive Psychology, 28*(3), 382–391. <https://doi.org/10.1002/acp.3008>.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Benton, S. L., Corkill, A. J., Sharp, J. M., Downey, R. G., & Khramtsova, I. (1995). Knowledge, interest, and narrative writing. *Journal of Educational Psychology, 87*(1), 66–79. <https://doi.org/10.1037/0022-0663.87.1.66>.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1–40. <https://doi.org/10.1023/A:1021302408382>.

Carroll, W. M. (1994). Using worked out 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>.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*(4), 347–362. <https://doi.org/10.1037/0022-0663.79.4.347>.

Crippen, K. J., Biesinger, K. D., Muis, K. R., & Orgill, M. (2009). The role of goal orientation and self-efficacy in learning from web-based worked examples. *Journal of Interactive Learning Research, 20*(4), 385–403.

Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research*. New York: University of Rochester Press.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: Sage Publication.

Hausmann, R. G. M., Van de Sande, B., & VanLehn, K. (2008). In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.). *Proceedings of the 30th annual conference of the cognitive science society* (pp. 2369–2374). Austin, TX: Cognitive Science Society.

Hoffman, B., & Schraw, G. (2010). Conceptions of efficiency: Applications in learning and problem solving. *Educational Psychologist, 45*(1), 1–14. <https://doi.org/10.1080/00461520903213618>.

Hoogerheide, V., Loyens, S. 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., Loyens, S. M. M., & Van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective than peers. *Learning and Instruction, 44*, 22–30. <https://doi.org/10.1016/j.lia.2016.05.001>.

- learninstruc.2016.02.004.
- 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.
- Hughes, A., Galbraith, D., & White, D. (2011). Perceived competence: A common core for self-efficacy and self-concept? *Journal of Personality Assessment*, *93*(3), 278–289. <https://doi.org/10.1080/00223891.2011.559390>.
- 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>.
- 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>.
- Klassen, R. M., & Usher, E. L. (2010). Self-efficacy in educational settings: Recent research and emerging directions. In T. C. Urdan, & S. A. Karabenick (Eds.). *The decade ahead: Theoretical perspectives on motivation and achievement* (pp. 1–33). Bingley, UK: Emerald Group Publishing Limited.
- 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*(2), 291–304. <https://doi.org/10.1007/s10648-015-9296-4>.
- Leppink, J., Paas, F., 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>.
- McLaren, B. M., Lim, S., & Koedinger, K. R. (2008). In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.). *Proceedings of the 30th annual conference of the cognitive science society* (pp. 2176–2181). Austin: Cognitive Science Society.
- Mason, L., Gava, M., & Boldrin, A. (2008). On warm conceptual change: The interplay of text, epistemological beliefs, and topic interest. *Journal of Educational Psychology*, *100*(2), 291–309. <https://doi.org/10.1037/0022-0663.100.2.291>.
- Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction*, *16*(2), 173–199. https://doi.org/10.1207/s1532690xc1602_2.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology*, *84*(4), 429–434. <https://doi.org/10.1037/0022-0663.84.4.429>.
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance Cognitive load theory. *Educational Psychologist*, *38*(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8.
- Paas, F., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*(1), 122–133.
- 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*(2), 92–103. <https://doi.org/10.1016/j.learninstruc.2006.02.008>.
- Renkl, A. (2014). Towards an instructionally-oriented theory of example-based learning. *Cognitive Science*, *38*(1), 1–37. <https://doi.org/10.1111/cogs.12086>.
- Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In C. Sansone, & J. M. Harackiewicz (Eds.). *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 375–407). San Diego, CA: Academic Press.
- Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognize designers' styles. *Learning and Instruction*, *19*(2), 185–199. <https://doi.org/10.1016/j.learninstruc.2008.03.006>.
- Schiefele, U., & Krapp, A. (1996). Topic interest and free recall of expository text. *Learning and Individual Differences*, *8*(2), 141–160. [https://doi.org/10.1016/S1041-6080\(96\)90030-8](https://doi.org/10.1016/S1041-6080(96)90030-8).
- Schraw, G., Bruning, R., & Svoboda, C. (1995). Sources of situational interest. *Journal of Reading Behavior*, *27*(1), 1–17. <https://doi.org/10.1080/10862969509547866>.
- Schunk, D. H. (2001). Social cognitive theory and self-regulated learning. In B. J. Zimmerman, & D. H. Schunk (Eds.). *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 125–151). Mahwah, NJ: Erlbaum.
- Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research*, *57*(2), 149–174. <https://doi.org/10.2307/1170234>.
- Stark, R., Gruber, H., Renkl, A., & Mandl, H. (2000). Instruktionale Effekte einer kombinierten Lernmethode: Zahlt sich die Kombination von Lösungsbeispielen und Problemlöseaufgaben aus? [Instructional effects of a combined learning method: Does the combination of worked-out examples and problem-solving tasks pay off?]. *Zeitschrift für Pädagogische Psychologie*, *14*(4), 205–217. <https://doi.org/10.1024/1010-0652.14.4.206>.
- 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*(1), 59–89. https://doi.org/10.1207/s1532690xc0201_3.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–295. <https://doi.org/10.1023/A:1022193728205>.
- Trafton, J. G., & Reiser, B. J. (1993). The contribution of studying examples and solving problems to skill acquisition. *Proceedings of the 15th annual conference of the cognitive science society* (pp. 1017–1022). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tobias, S. (1996). Interest and metacognitive word knowledge. *Journal of Educational Psychology*, *87*(3), 399–405. <https://doi.org/10.1037/0022-0663.87.3.399>.
- Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. G. (2002). Cognitive load theory and aging: Effects of worked examples on training efficiency. *Learning and Instruction*, *12*(1), 87–105. [https://doi.org/10.1016/S0959-4752\(01\)00017-2](https://doi.org/10.1016/S0959-4752(01)00017-2).
- Van Gog, T., Kester, L., Dirx, K., Hoogerheide, V., Boerboom, J., & Verkoijen, P. P. J. L. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educational Psychology Review*, *27*(2), 265–289.
- Van Gog, T., & Kester, L. (2012). A test of the testing effect: Acquiring problem-solving skills from worked examples. *Cognitive Science*, *36*(8), 1532–1541. <https://doi.org/10.1111/cogs.12002>.
- Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, *36*(3), 212–218. <https://doi.org/10.1016/j.cedpsych.2010.10.004>.
- Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, *43*(1), 16–26. <https://doi.org/10.1080/00461520701756248>.
- Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction*, *16*(2), 154–164. <https://doi.org/10.1016/j.learninstruc.2006.02.003>.
- Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, *22*(2), 155–174. <https://doi.org/10.1007/s10648-010-9134-7>.
- Van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education*, *72*, 323–327. <https://doi.org/10.1016/j.compedu.2013.12.004>.
- Van der Meij, H., Rensink, I., & Van der Meij, J. (2018). Effects of practice with videos for software training. *Computers in Human Behavior*, *89*, 439–445. <https://doi.org/10.1016/j.chb.2017.11.029>.
- Williams, G. C., & Deci, E. L. (1996). Internalization of biopsychosocial values by medical students: A test of self-determination theory. *Journal of Personality and Social Psychology*, *70*(4), 767–779. <https://doi.org/10.1037/0022-3514.70.4.767>.
- Williams, G. C., Freedman, Z. R., & Deci, E. L. (1998). Supporting autonomy to motivate glucose control in patients with diabetes. *Diabetes Care*, *21*(10), 1644–1651. <https://doi.org/10.2337/diacare.21.10>.