Received: 4 September 2019 | Revised: 27 February 2020 | Accepted: 28 February 2020
DOI: 10.1002/acp.3649

SPECIAL ISSUE ARTICLE

WILEY

Examples, practice problems, or both? Effects on motivation and learning in shorter and longer sequences

Milou van Harsel[1,2] | Vincent Hoogerheide[2] | Peter Verkoeijen[1,3] | Tamara van Gog[2]

[1]Learning and Innovation Centre, Avans University of Applied Sciences, Breda, The Netherlands

[2]Department of Education, Utrecht University, Breda, The Netherlands

[3]Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Breda, The Netherlands

Correspondence
Milou van Harsel, Learning and Innovation Centre, Avans University of Applied Sciences, PO Box 90116, 4800 RA Breda, the Netherlands.
Email: m.vanharsel@avans.nl

Summary

Research suggests some sequences of examples and problems (i.e., EE, EP) are more effective (higher test performance) and efficient (attained with equal/less mental effort) than others (PP, sometimes also PE). Recent findings suggest this is due to motivational variables (i.e., self-efficacy), but did not test this during the training phase. Moreover, prior research used only short task sequences. Therefore, we investigated effects on motivational variables, effectiveness, and efficiency in a short (Experiment 1; four learning tasks; n = 157) and longer task sequence (Experiment 2; eight learning tasks; n = 105). With short sequences, all example conditions were more effective, efficient, and motivating than PP. With longer sequences, all example conditions were more motivating and efficient than PP, but only EE was more effective than PP. Moreover, EE was most efficient during training, regardless of sequence length. These results suggest that example study (only) is more effective, efficient, and more motivating than PP.

KEYWORDS

example-based learning, mental effort, problem-solving, self-efficacy, video modeling examples

1 | INTRODUCTION

It is well-established that for novices who have little or no prior knowledge of a task, studying worked-out examples of problem solutions—or studying examples alternated with practice problem-solving—is a more effective and efficient instructional strategy than practice problem-solving only (for a review, see Van Gog, Rummel, & Renkl, 2019). Effective means it often results in higher posttest performance, and efficient means that this higher performance is often attained with equal or less effort investment in the learning and test phases. Example study is more effective and efficient for novices than practice problem-solving because it gives novices the opportunity to devote all available cognitive capacity to study the step-by-step explanation of the solution procedure, which helps them to develop a schema on how to solve this type of problem in the future (e.g., Sweller & Cooper, 1985). When solving practice problems, in contrast, novices (lacking prior knowledge) have to resort to weak problem-solving strategies (e.g., via trial-and-error, means-ends analysis), which is very effortful and time consuming, yet hardly contribute to learning (e.g., Sweller, 1988). For learners with higher prior knowledge, however, instructional strategies with a high level of support may be less efficient, because they have already developed proper cognitive schemata to guide their problem-solving (cf. expertise-reversal effect; Kalyuga, Chandler, Tuovinen, & Sweller, 2001; Kalyuga & Sweller, 2004; Kalyuga & Renkl, 2010; Roelle & Berthold, 2013). These learners might gain more from practice problem-solving than example study.

Despite the multitude of studies on *example-based learning*, an important open question that remains is how example study and practice problem-solving should be sequenced to be most effective (i.e., for students' posttest performance), most efficient (i.e., posttest performance considered in light of mental effort investment in the training and test tasks), and most motivating for learning.

2 | SHORT TASK SEQUENCES OF EXAMPLE STUDY AND PRACTICE PROBLEM-SOLVING

Van Gog, Kester, and Paas (2011) were the first to compare the four most commonly used sequences of examples and practice problems to uncover which sequence would be most effective and efficient for learning. Secondary education students (novices) learned how to diagnose a fault in electrical circuits with the help of four training tasks presented as examples only (EEEE), example-problem pairs (EPEP), problem-example pairs (PEPE), or practice problems only (PPPP). Results showed that EEEE and EPEP were more effective and efficient than PEPE and PPPP. No differences were found, however, between the conditions starting with an example (i.e., EEEE and EPEP) and between the conditions starting with a practice problem (i.e., PEPE and PPPP).

Since then, follow-up research has investigated whether these findings would replicate and how they could best be explained. However, studies attempting to replicate the differences between the example-problem pairs (EP-pairs) and problem-example pairs (PE-pairs) conditions showed mixed results (see Table 1 for the characteristics of these studies). Whereas some studies also found that EP-pairs were more effective and efficient for learning than PE-pairs (e.g., Kant et al., 2017; Leppink et al., 2014), others did not find any test performance and/or effort investment differences (e.g., Van Harsel et al., 2019; Coppens et al., 2019; Van der Meij et al., 2018; Van Gog, 2011). A small-scale meta-analysis by Van Harsel et al. (2019) on all (published) studies available at that time showed a significant, small-to-medium meta-analytic advantage of EP over PE on final test performance (Cohen's *d* of 0.350), albeit with a large heterogeneity between effects.

3 | THE ROLE OF MOTIVATION DURING EXAMPLE STUDY AND PRACTICE PROBLEM-SOLVING

An explanation for these mixed findings might lie in motivational aspects of learning. That is, when novices have to learn how to solve a complex task that requires domain-specific knowledge and that is not particularly intrinsically rewarding or enjoyable, then starting the training phase with a practice problem (PE-pairs) might decrease their motivation. Solving such a practice problem could be experienced as so difficult that learners lose interest in the topic of the learning materials (i.e., topic interest) or confidence in their ability to learn the task (e.g., self-efficacy and perceived competence). As a consequence, learners may not be motivated to study the subsequent example (and possibly also the tasks that follow). In this case, PE-pairs are probably less effective for learning than EP-pairs. However, when the complex task is experienced as intrinsically rewarding or enjoyable, starting the training phase with a practice problem (PE) might not have a detrimental effect on students' interest or confidence in their ability to learn the task. In this case, studying EP is probably equally effective for learning as studying PE.

This motivational explanation was tested in two recent studies in which novices learned to solve mathematical problems (i.e., Van Harsel et al., 2019; Coppens et al., 2019). In these studies, aspects of motivation such as topic interest, self-efficacy, and perceived competence were measured before and after the training phase to investigate whether students lose interest in the task (i.e., topic interest) or confidence in their ability to learn the task (i.e., self-efficacy and perceived competence) as a result of starting the training phase with a practice problem. Self-efficacy is defined as a personal judgment of one's own capacities to organize or accomplish a specific task or challenge and has shown to have a positive effect on factors such as academic motivation, study behavior, and learning outcomes (e.g., Bandura, 1997; Schunk, 2001). Perceived competence is related to the construct of self-efficacy, but comprises more general knowledge and perceptions of people's self-concept toward one's own competence (e.g., Deci & Ryan, 2002; Hughes, Galbraith, & White, 2011). Like self-efficacy, perceived competence is also positively linked to factors such as academic motivation and learning outcomes (e.g., Bong & Skaalvik, 2003). Finally, topic interest can be described as personal interest in a domain or activity based on previously acquired knowledge, personal experiences, and emotions (e.g., Ainley, Hidi, & Berndorff, 2002; Renninger, 2000). Topic interest has positive effects on cognitive functioning, (deep) learning, and engagement (e.g., Hidi, 1990; Schiefele & Krapp, 1996; Tobias, 1996).

In contrast to the motivational explanation, Van Harsel et al. (2019) and Coppens et al. (2019) found no differences between EP-pairs and PE-pairs on test performance, or on self-efficacy, perceived competence, and topic interest. However, in these studies, these motivational constructs were only measured before and after the training phase. Measuring self-efficacy after each task in the training phase would be more insightful, because it could reveal whether self-efficacy was not negatively affected at all when starting the training phase with a practice problem or whether it recovered quickly once provided with an example. Another improvement that would allow for a more sensitive test is to use a conceptual pretest rather than a procedural one, as was the case in the study by Van Harsel et al.(2019; i.e., two practice problems isomorphic to the training phase). With such a procedural pretest, one could argue that all participants started with practice problem-solving (also the example conditions: PPEEEE and PPEPEP). Therefore, the first aim of the present study was to investigate students' self-efficacy during the training phase in four task sequences (EEEE, EPEP, PEPE, PPPP). The second aim was to address the open question of how motivational and cognitive aspects of learning would be affected by those task sequences in longer training phases.

TABLE 1 Characteristics of studies investigating the effectiveness and efficiency of EP-pairs and PE-pairs

	Van Harsel, Hoogerheide, Verkoeijen, & Van Gog, 2019; Exp. 1	**Van Harsel, Hoogerheide, Verkoeijen, & Van Gog, 2019; Exp. 2**	**Coppens, Hoogerheide, Snippe, Flunger, & Van Gog, 2019**	**Kant, Scheiter, & Oschatz, 2017**	**Leppink, Paas, Van Gog, Van der Vleuten, & Van Merriënboer, 2014**	**Van der Meij, Rensink, & Van der Meij, 2018**	**Van Gog, 2011**	**Van Gog et al., 2011**
Learner characteristics								
Average age	19.3	19	10.6	12.5	—	11.2	20.2	16.2
Educational level	First-year students from a university of applied sciences, enrolled in an electrical and electronic or mechanical engineering program	First-year students from a university of applied sciences, enrolled in a teacher training program	Elementary school students	Seventh grade students	First-year university students, enrolled in a social and health sciences program	Fifth grade and sixth grade classrooms from elementary school	Students enrolled in programs at the Faculty of Social Sciences	Students in their fourth or fifth year of preuniversity education
Type of knowledge in learning and test materials	Procedural knowledge	Procedural knowledge	Procedural knowledge	Conceptual and procedural knowledge	Procedural knowledge	Procedural knowledge	Procedural knowledge	Procedural knowledge
Topic of learning materials	Mathematics, trapezoidal rule	Mathematics, trapezoidal rule	Mathematics, water jug problems	Science, scientific reasoning, and inquiry tasks	Statistics, application of Bayes' theorem	Software training on word	Mathematics, frog leap	Science, applying Ohm's law to reason about faults in electrical circuits
Learning setting	Classroom experiment at school, not part of the curriculum	Classroom experiment at school, not part of the curriculum	Classroom experiment at school, not part of the curriculum	Computer room experiment at school, not part of the curriculum	Classroom experiment part of statistics course	Computer room experiment at school, not part of the curriculum	Individual experiment in the lab of the university	Classroom experiment at school, not part of the curriculum

4 | LONGER TASKS SEQUENCES OF EXAMPLE STUDY AND PRACTICE PROBLEM-SOLVING

Previous sequencing research often used a small number of training tasks (i.e., two tasks: Van Harsel et al., 2019; Kant et al., 2017; Leppink et al., 2014; four tasks: Van Gog, 2011; Van Gog et al., 2011). In such short sequences, EE was found to be equally or more effective (and efficient) for learning as EP on an immediate posttest (e.g., Van Harsel et al., 2019; Kant et al., 2017; Leppink et al., 2014; Van der Meij et al., 2018) and a delayed posttest (e.g., Leahy, Hanham, & Sweller, 2015; Van Gog et al., 2015; Van Gog & Kester, 2012). Moreover, no differences between EE and EP were found on motivational aspects of learning (i.e., self-efficacy, perceived competence, and topic interest; Van Harsel et al., 2019).

However, in educational practice students may encounter (much) longer study sequences. Because students will gain knowledge as training progresses, longer task sequences may affect motivational and cognitive aspects of learning differently than shorter sequences. That is, studying examples only might not only become boring but also redundant as students gain knowledge from the first few tasks. This in turn might have negative effects on motivational aspects of learning (and performance; see Kalyuga et al., 2001) as compared to sequences in which examples and problems are alternated. It might be more engaging for learners to actively attempt to solve practice problems than to continuously study examples, which is more passive learning (as suggested—but not tested—by Sweller & Cooper, 1985). Examples alternated with practice problems might be more engaging than example study only in longer sequences as the interspersed practice problems give learners the opportunity to actively apply what they have learned and allow them to identify gaps in their knowledge (cf. Baars, Van Gog, De Bruin, & Paas, 2014, 2017), which they can repair when studying subsequent examples.

5 | THE PRESENT STUDY

In sum, the present study aimed to examine how short (i.e., Experiment 1: EEEE, EPEP, PEPE, and PPPP) and longer (i.e., Experiment 2: EEEEEEEE, EPEPEPEP, PEPEPEPE, and PPPPPPPP) task sequences of examples and/or practice problems would affect motivational and cognitive aspects of learning on an immediate posttest. With regard to short sequences, we added a delayed posttest to see whether effects remained stable over time. Furthermore, we measured self-efficacy after each task in the training phase (instead of only before and after the training phase). In this way, we were able to explore whether and how motivation was affected by the order of examples and practice problems in the training phase. Finally, a conceptual pretest was used instead of a procedural pretest as in the study by Van Harsel et al. (2019).

6 | EXPERIMENT 1

In Experiment 1, it was investigated how short task sequences of examples and/or practice problems (i.e., EEEE, EPEP, PEPE, and PPPP) would affect motivational (i.e., self-efficacy, perceived competence, and topic interest measured before and after the training phase) and cognitive aspects of learning (i.e., invested mental effort in the training phase and performance on isomorphic and transfer tasks). We explored effects on time-on-task (training phase and posttest phases) and mental effort (posttest phases), because when combined with test performance, these measures are indicators of the efficiency of the learning process and learning outcomes (Van Gog & Paas, 2008). We also administered a delayed posttest to explore whether the pattern of results would remain stable after a 1 week delay. We expect to replicate the pattern of results found by Van Harsel et al. (2019), because the same materials and population are used (see Table 2 for results found by Van Harsel et al., 2019). Note that we used a conceptual pretest instead of a procedural pretest to rule out the alternative explanation that when a procedural pretest is used (e.g., two practice problems in Van Harsel et al., 2019), one could argue that all participants start with practice problem-solving (also the example conditions: PPEEEE and PPEPEP). As a result, if the motivational explanation would be valid, even students in the example-first conditions would lose interest and confidence in their own abilities before the first example. Therefore, it is possible that EPEP becomes more motivating, effective, and efficient for learning compared to PEPE when using a conceptual pretest (instead of EPEP = PEPE as found by Van Harsel et al., 2019).

Regarding self-efficacy after each training task, it was expected that students in the EEEE and EPEP condition would show significantly higher levels of self-efficacy after the first training task than students in the PEPE and PPPP condition (H1a). We assumed that the PEPE condition would "recover" after receiving an example as second training task (given that prior research with these tasks showed no differences in motivation and learning outcomes after training), and therefore we expected no significant differences on self-efficacy scores among the EEEE, EPEP, and PEPE conditions from the second training task onwards (H1b). Since students in the PPPP condition

TABLE 2 Main results of Experiment 1 of Van Harsel et al. (2019) regarding the effects of short sequences of examples and problems (EEEE, EPEP, PEPE, and PPPP) on isomorphic tasks performance, transfer tasks performance, mental effort, self-efficacy, perceived competence, and topic interest

	Immediate posttest
Training phase	
Mental effort	EE, EP, PP < PP; EE < EP, PE; EP = PE
Immediate posttest phase	
Isomorphic tasks	EE, PE > PP; EE > EP; EP = PE
Procedural transfer task	EE = EP = PE = PP
Conceptual transfer task	EE = EP = PE = PP
Self-efficacy	EE, EP, PP > PP; EE > EP; EP = PE
Perceived competence	EE, EP, PP > PP; EE > EP; EP = PE
Topic interest	EE = EP = PE = PP

Abbreviations: EE, example study only; EP, example-problem pairs; PE, problem-example pairs; PP, problem-solving only.

were not provided with an opportunity to study an example, it was predicted that self-efficacy scores would be significantly higher in the EEEE, EPEP, and PEPE condition than in the PPPP condition from the second training task onwards (H1c).

6.1 | Method

6.1.1 | Participants and design

Participants were 157 Dutch higher education students enrolled in the first year of an electrical and electronic mechanical engineering program (M^{age} = 19.13, *SD* = 1.75; 155 male, 2 female). Participants were randomly assigned to one of four conditions: examples only (*n* = 33; EEEE), example-problem pairs (*n* = 45; EPEP), problem-example pairs (*n* = 40; PEPE), or practice problems only (*n* = 39; PPPP). The experiment consisted of four phases: (a) pretest, (b) training phase, (c) immediate posttest phase, and (d) delayed posttest phase. At the delayed posttest, which was completed after 1 week, 25 participants were absent so these data are based on 132 participants (M^{age} = 19.04, *SD* = 1.71; 130 male, 2 female). Participants were assumed to be novices to the modeled task (i.e., approximating the definite integral of a function using the trapezoidal rule) as this subject had not (yet) been a part of their study program. Participants gave their informed consent prior to their inclusion in the study and received study credits for their participation.

6.1.2 | Materials

All materials were presented using a web-based learning environment. The materials were based on the materials developed by Van Harsel et al. (2019).

Pretest

The pretest was a conceptual prior knowledge test that consisted of seven multiple-choice questions (α = .49)[1] and was developed in collaboration with two math teachers from a higher education institute. This test was used to check whether participants' ability to recognize and name the basic principles of the trapezoidal rule was low and whether prior knowledge did not differ among conditions. An example of a conceptual prior knowledge question was given in Appendix C.

Training phase

The training phase consisted of four tasks that required participants to use the trapezoidal rule. The trapezoidal rule is a numerical integration method that is used to give a quantitative approximation of the region under the graph of a specific function. Each task had its own cover story (i.e., task 1: fitness, task 2: energy measurement, task 3: washing machine, and task 4: soapsuds). To ensure that only the task format differed across conditions, the task order was identical for all participants (i.e., in order: fitness, energy measurement, washing machine, and soapsuds). Each task was part of a task pair (i.e., pair 1: fitness and energy measurement, pair 2: washing machine and soapsuds). Within a task pair, the tasks were isomorphic (i.e., a similar problem-solving procedure, but surface features such as the cover stories and numbers used in functions were slightly different). There was a minor complexity difference between the first and second task pair. The first pair of tasks required Participants to calculate with positive numbers. The second pair was slightly more complex because Participants had to calculate with both positive and negative numbers.

Regarding the design of the tasks, the practice problems started with a short description of the problem state. Then, some additional information was provided on how to solve the problem, such as the trapezoidal rule formula, the graph of a function, the left border and right border of the area to be calculated, and the number of intervals. It was, however, not explained how to use the information to solve the practice problem. At the end of the problem format, participants received the following assignment: "Approach the area under the graph using the information that is given. Write down all your intermediate steps and calculations." Participants could solve the problem by completing the four steps: (a) "compute the step size of each subinterval," (b) "calculate the *x*-values," (c) "calculate the function values for all *x*-values," (d) "enter the function values into the formula and calculate the area." An example of a problem format is given in Appendix A.

Each video modeling example displayed a screen capture of a female model's computer screen, in which she demonstrated in a stepwise manner how to solve a practice problem with the help of the trapezoidal rule. While solving the problem, the model provided verbal explanations and on-screen handwritten notes. At the start of the video, the model first explained the purpose of the trapezoidal rule and then provided an explanation of the problem state. The problem state was exactly the same as in the problem format. Subsequently, the model demonstrated and explained how one could 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, and the trapezoidal rule) and eventually showed how to solve the problem by calculating the four steps listed in the description of the problem format. A screenshot of a video modeling example is given in Appendix B.

Immediate and delayed posttest

The immediate and delayed posttest presented four tasks, two isomorphic and two transfer tasks. Of the two isomorphic tasks (immediate posttest: α = .71; delayed posttest: α = .77), one was isomorphic to the first pair of training tasks and the other to the second pair of training tasks. The third posttest task measured procedural transfer and asked participants to use the Simpson rule instead of the trapezoidal rule to approximate the definite integral under a graph. The Simpson rule is also a numerical method for approximating the integral of a function. The problem-solving procedure of Simpson's rule is comparable to that of the trapezoidal rule, however, Simpson's rule uses a different formula to approximate the definite integral of a function (i.e., with a sequence of quadratic parabolic segments instead of straight lines such as the trapezoidal rule). The fourth posttest task measured conceptual transfer and consisted of five open-ended questions that aimed to measure Participants' understanding of the trapezoidal rule. All five questions comprised a multiple-choice part with four options an "explanation" part (where participants had to justify their chosen answer). Hence, these questions were more complex than the conceptual pretest items, which only required participants to select the correct answer. Unfortunately, the data regarding the conceptual transfer questions had to be excluded from the analyses due to a programming error. An example of an isomorphic posttest task, procedural transfer task and conceptual transfer question can be found in Appendix C.

Mental effort

After each task on the pretest, the training phase, the immediate posttest, and the delayed posttest, participants rated their mental effort on a 9-point mental effort rating scale (Paas, 1992), with answer options ranging from (1) "very, very low mental effort" to (9) "very, very high mental effort."

Self-efficacy, perceived competence, and topic interest

Self-efficacy was measured before, during (i.e., after each training task), and 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" (Van Harsel et al., 2019; adapted from Hoogerheide, Van Wermeskerken, Loyens, & Van Gog, 2016).

Perceived competence was measured using the *Perceived Competence Scale for Learning* (Van Harsel et al, 2019; based on Williams & Deci, 1996; Williams Freedman, & Deci, 1998). This perceived competence scale (immediate posttest: α = .98; delayed posttest: α = .97) consisted of three items: "I feel confident in my ability to learn how to approximate the definite integral of a graph using the trapezoidal rule", "I am capable of approximating the definite integral of a graph using the trapezoidal rule", and "I feel able to meet the challenge of performing well when I have to apply the trapezoidal rule". Participants were asked to rate on a scale of (1) "not at all true" to (7) "very true" to what degree these three items applied to them.

The topic interest scale (Van Harsel et al., 2019; adapted from the topic interest scale by Mason, Gava, & Boldrin, 2008, and the perceived interest scale by Schraw, Bruning, & Svoboda, 1995) were used to measure participants' interest in the topic (i.e., the trapezoidal rule). The topic interest scale (immediate posttest: α = .81; delayed posttest: α = .82) consisted of seven items and participants had to rate on a 7-point scale, ranging from (1) "totally disagree" to (7) "totally agree", to what degree each of the items applied to them. All items are shown in Appendix D.

6.1.3 | Procedure

The experiment was run in 16 sessions (i.e., eight first sessions and eight second sessions) and took place in a computer classroom at the participants' institute of higher education. The number of participants ranged from 2 to 23 per session. Prior to the first session, headsets, pens, and scrap paper (to write down calculations) were distributed. Once participants were seated in the computer classroom, the first session (ca. 106 min) started with a general introduction by the experimenter explaining the aim and procedure of the experiment. Participants were told they could work at their own pace (with a maximum of 135 min) on mathematical tasks in an online learning environment by means of different instructional formats (i.e., examples and/or practice problems). They were instructed to write down as much as possible when solving a training task or test task, and that if they really did not know what to answer, to write an "X". After the instruction, participants received a paper with a link and a password that gave access to the online learning environment.

The learning environment was designed in such a way that each task and questionnaire were presented on a separate page. Participants were unable to go back to previous pages and had to complete each task or questionnaire before they could go to the next page. Time was logged for each task. When participants entered the learning environment, they were assigned to one of the four conditions (i.e., EEEE, EPEP, PEPE, or PPPP). Participants started with a short demographic questionnaire (e.g., age, gender, and preliminary education), followed by the conceptual pretest. After the pretest, participants completed the self-efficacy, perceived competence, and topic interest questionnaires before they started the training phase. During the training phase, participants received four tasks that were presented as examples and/or practice problems (depending on their assigned condition). After each task, participants were asked to indicate their perceived mental effort and self-efficacy. After the training phase, participants completed the self-efficacy, perceived competence, and topic interest questionnaires again. Lastly, participants took the immediate posttest. Participants had to rate their invested mental effort after each posttest task. Participants handed in their scrap paper before working on the posttest phase and received new ones to make notes.

The delayed posttest took place exactly 7 days later (ca. 40 min) and started with a general introduction in which the procedure was explained. Again, participants were told they could work at their own pace, write down everything they could, and note an "X" if they were not able to answer a question. Participants were provided with scrap paper and a password that gave them access to the online learning environment. They first completed the self-efficacy, perceived competence, and topic interest questionnaires. Subsequently, they took the delayed posttest, which consisted of four tasks that were isomorphic to the tasks used in the immediate posttest phase. After each task, participants were asked to indicate their invested mental effort.

6.1.4 | Data analysis

The data was scored by the experimenter (i.e., first author) and a second encoder based on a scoring protocol that was developed by Van Harsel et al. (2019) in collaboration with higher education mathematics teachers. Participants could earn a maximum of eight points per training problem. Two points could be earned for calculating the step size of

TABLE 3 Post hoc comparisons of mental effort, self-efficacy, perceived competence, topic interest, isomorphic tasks performance, and procedural transfer on immediate and delayed posttests in Experiment 1

	EE vs. EP			EE vs. PE			EE vs. PP			EP vs. PE			EP vs. PP			PE vs. PP		
	U	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*
Training																		
Mental effort	1,035.5	.003	.337	1,309	**<.001**	.392	1,179.5	**<.001**	.716	1,160.5	**<.001**	.651	1,447	**<.001**	.588	1,186	**<.001**	.449
Immediate posttest																		
Isomorphic tasks[a]	800.5	.556	.067	690	.738	.039	232	**<.001**	.554	850.5	.662	.047	355.5	**<.001**	.518	233.5	**<.001**	.607
Procedural transfer[b]	774	.742	.037	679	.826	.026	363	**<.001**	.420	874	.810	.026	391.5	**<.001**	.514	332	**<.001**	.539
Self-efficacy[c]	574.5	.082	.765	562	.260	.132	76	**<.001**	.765	997	.381	.095	171	**<.001**	.698	109	**<.001**	.750
Perceived competence[d]	609.5	.175	.154	589	.425	.093	70.5	**<.001**	.769	1,000	.373	.097	151.5	**<.001**	.714	92	**<.001**	.765
Topic interest[e]	618	.207	.143	562.5	.279	.127	391.5	**.004**	.336	942	.711	.040	679.5	.075	.194	574.5	.044	.227
Delayed posttest																		
Isomorphic tasks[a]	503	.986	.005	470	.769	.038	182.5	**.001**	.457	754.5	.719	.041	256.5	**<.001**	.503	206	**<.001**	.544
Procedural transfer[b]	471	.677	.052	442.5	.907	.002	246.5	**.008**	.353	749.5	.743	.038	426	.011	.301	341	**.005**	.371
Self-efficacy[c]	482.5	.808	.030	518.5	.291	.017	101	**<.001**	.642	869.5	.106	.185	146	**<.001**	.652	85	**<.001**	.735
Perceived competence[d]	475.5	.739	.014	503	.432	.013	111.5	**<.001**	.612	832	.241	.135	168	**<.001**	.624	113	**<.001**	.687
Topic interest[e]	464.5	.631	.060	418	.638	.060	333	.368	.120	723	.975	.004	566.5	.534	.074	532	.743	.040

Abbreviations: EE, example study only; EP, example-problem pairs; PE, problem-example pairs; PP, problem-solving only. Significant *p*-values (after correction) are bolded.
[a]Isomorphic task performance did not differ statistically between the immediate and delayed posttest ($Z = 2{,}821.5$, $p = .766$, $r = .026$).
[b]Procedural transfer task performance statistically differed between the immediate and delayed posttest ($Z = 739.5$, $p = .006$, $r = .239$), however, follow-up tests showed that changes within conditions were not significant ($ps > .031$, $rs < .359$).
[c]Self-efficacy statistically differed between the pretest and immediate posttest ($Z = 9.85$, $p < .001$, r = .786) and increased in EE, EP, and PE condition ($ps < .001$), not in the PP condition ($p = .015$). Self-efficacy statistically differed between the immediate and delayed posttest ($Z = -7.14$, $p < .001$, $r = .621$) and decreased in EE, EP, and PE condition ($ps < .001$), not in PP condition ($p = .954$).
[d]Perceived competence statistically differed between pretest and immediate posttest ($Z = 9.52$, $p < .001$, $r = .760$) and increased in EE, EP, and PE condition ($ps < .001$), not in PP condition ($p = .015$). Perceived competence statistically differed between the immediate and delayed posttest ($Z = -6.034$, $p < .001$, $r = .525$) and decreased in EE, EP, and PE condition ($ps < .001$), not in PP Condition ($p = .954$).
[e]Topic interest did not differ statistically between the pretest and immediate posttest ($p = .736$, $r = .325$). Topic interest statistically differed between the immediate and delayed posttest ($Z = -5.32$, $p < .001$, $r = .463$) and decreased in EP and PE Condition ($ps < .011$), but not in EE Condition ($p = .147$) and PP Condition ($p = .030$).

TABLE 4 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	M	*SD*	Med	M	*SD*	Med	M	*SD*	Med
Pretest												
Self-efficacy	2.18	1.84	1.00	2.40	1.86	2.00	2.33	1.31	2.00	2.00	1.34	1.00
Perceived competence	1.77	1.28	1.33	2.17	1.48	1.67	1.98	1.11	1.67	2.07	1.20	1.67
Topic interest	4.57	0.78	4.86	4.43	0.73	4.29	4.45	0.84	4.36	4.23	0.89	4.43
Training												
Self-efficacy	7.09	1.39	7.26	6.06	1.36	6.00	5.53	1.11	5.38	2.72	1.90	2.00
Immediate posttest												
Self-efficacy	7.39	1.27	7.00	6.73	1.64	7.00	7.10	1.28	7.00	2.79	2.19	2.00
Perceived competence	5.83	0.88	6.00	5.35	1.30	5.67	5.66	0.87	6.00	2.29	1.61	2.00
Topic interest	4.68	0.86	4.86	4.45	0.93	4.43	4.50	0.98	4.57	4.03	0.98	4.29
Delayed posttest												
Self-efficacy	5.12	1.59	6.00	5.18	1.66	5.00	5.69	1.17	6.00	2.39	1.69	2.00
Perceived competence	4.37	1.32	4.67	4.42	1.23	4.50	4.69	0.98	4.83	2.24	1.52	1.67
Topic interest	4.26	0.97	4.14	4.13	0.88	4.00	4.11	0.86	4.00	3.95	0.86	4.14

Abbreviations: EE, example study only; EP, example-problem pairs; PE, problem-example pairs; PP, problem-solving only.

each subinterval, two for correctly calculating all *x*-values, two for correctly calculating the function values for all *x*-values, and two for using the correct formula for the area under the graph and providing the correct answer. If half or more of the solution steps were correct in step two, three, and four, then one point was granted. If less than half of the solution steps were correct in step two, three and four, zero points were granted. These scoring standards were also used to score the two isomorphic posttest tasks (i.e., max. score = 16 points) and the procedural transfer problem (i.e., max. score = 8 points). The intraclass correlation coefficient was .98 for the training tasks, .98 for the isomorphic posttest tasks, and .93 for the delayed posttest tasks.

The average mental effort invested in the training phase and on the isomorphic posttest tasks was calculated. In addition, the average self-efficacy, perceived competence, and topic interest ratings were calculated.

6.2 | Results

Nonparametric tests were used to analyze our main research questions and explorative questions, because with the exception of topic interest on pretest and delayed posttest, and self-efficacy and perceived competence on the delayed posttest, none of our main variables were normally distributed (cf. Field, 2009), with either the kurtosis, skewness, or both coefficients being (substantially) below −1.96 or above +1.96. Therefore, effects of Instruction Condition (EEEE, EPEP, PEPE, and PPPP) were tested on motivational (i.e., self-efficacy, perceived competence, and topic interest) and cognitive aspects of learning (i.e., isomorphic test performance, procedural transfer, conceptual transfer, mental effort and time-on-task in learning and posttest phases) with Kruskal–Wallis tests. Significant main effects of Instruction Condition were followed by six Mann–Whitney *U* tests (EEEE vs. EPEP, EEEE vs. PEPE, EEEE vs. PPPP, EPEP vs. PEPE, EPEP vs. PPPP, and PEPE vs. PPPP) with a Bonferroni-corrected significance level of $p < .008$ (i.e., 0.05/6). Results are presented in the main text and Table 3. Effects of Test Moment (Immediate Posttest and Delayed Posttest) for each condition (EEEE, EPEP, PEPE, and PPPP) were tested with Wilcoxon signed-rank tests and we used four Mann–Whitney *U* tests as post hoc tests (see Table 3), with a Bonferroni-corrected significance level of $p < .013$ (i.e., 0.05/4). 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, respectively (Cohen, 1988) for the post hoc tests. The self-efficacy, perceived competence, and topic interest scores can be found in Table 4, and the test performance scores, mental effort scores, and time-on-task scores in Table 5.

Before the differences within and among conditions were analyzed, we checked for prior knowledge differences. Kruskal–Wallis tests showed no significant differences among conditions on pretest performance, $H(3) = 2.58$, $p = .460$, or on pretest scores of self-efficacy, $H(3) = 2.59$, $p = .460$, perceived competence, $H(3) = 2.18$, $p = .536$, and topic interest, $H(3) = 3.22$, $p = .360$.

6.3 | How do short sequences of examples and problems affect self-efficacy, perceived competence, and topic interest?

6.3.1 | Self-efficacy

Self-efficacy ratings measured after each training task are presented in Figure 1. It was analyzed whether participants' self-efficacy

TABLE 5 Mean (M), Standard Deviation (SD), and Median (Med) of Pretest (range 0–16), isomorphic tasks performance (range 0–16), procedural transfer (range 0–8), 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	2.94	2.03	4.00	2.31	1.41	2.00	2.60	1.63	3.00	2.46	1.59	2.00
Training												
Mental effort	2.57	1.05	2.50	3.42	1.18	3.25	4.21	0.96	4.13	6.44	2.41	6.75
Time-on-task	4.35	1.63	4.50	8.68	5.07	11.00	7.67	2.07	7.00	6.27	5.02	5.50
Immediate posttest												
Isomorphic tasks	9.67	4.06	10.00	9.89	5.07	11.00	10.20	3.34	10.50	3.77	4.64	2.00
Procedural transfer	1.91	2.34	1.00	1.73	1.68	1.00	1.63	1.53	1.00	0.33	0.74	0.00
Mental effort												
Isomorphic tasks	4.89	1.52	5.00	4.73	1.69	4.50	4.94	1.38	5.00	6.51	2.56	7.00
Procedural transfer	5.36	2.41	5.00	5.98	2.15	6.00	5.10	2.37	5.00	6.62	2.56	8.00
Time-on-task												
Isomorphic tasks	16.87	6.39	14.50	10.61	4.99	10.50	11.90	3.34	11.25	4.99	4.79	4.00
Procedural transfer	9.27	4.87	9.00	8.38	5.29	8.00	7.88	4.29	7.00	3.87	4.13	2.00
Delayed posttest												
Isomorphic tasks	9.28	5.30	11.00	9.60	4.42	10.00	10.00	4.16	10.50	4.16	4.75	2.00
Procedural transfer	1.32	1.70	1.00	1.15	1.53	1.00	1.08	1.23	1.00	0.52	1.48	0.00
Mental effort												
Isomorphic tasks	4.80	1.90	4.00	4.55	1.52	4.50	4.81	1.65	5.00	6.76	2.00	7.50
Procedural transfer	5.36	2.33	5.00	5.23	2.07	5.00	5.03	2.18	5.00	6.71	2.52	8.00
Time-on-task												
Isomorphic tasks	12.56	4.48	12.00	11.69	4.82	11.50	10.85	4.37	10.50	7.31	5.29	7.50
Procedural transfer	7.52	4.72	6.00	7.45	4.91	7.00	7.53	3.56	8.00	4.71	4.17	5.00

Abbreviations: EE, example study only; EP, example-problem pairs; PE, problem-example pairs; PP, problem-solving only.

reported after each training task differed among conditions (see Table 6 for post hoc comparisons). With regard to the *first training task*, there was a main effect of Instruction Condition, $H(3) = 83.13$, $p < .001$. As predicted (H1a), self-efficacy levels were higher in the EEEE and EPEP Condition than the PEPE and PPPP Condition. No significant differences were found between the EEEE and EPEP Condition or between the PEPE and PPPP Condition.

Regarding self-efficacy from the *second training task onwards*, there was also a main effect of Instruction Condition (task 2: $H(3) = 59.48$, $p < .001$; task 3: $H(3) = 68.37$, $p < .001$; task 4: $H(3) = 68.61$, $p < .001$). As expected (H1b, H1c), results showed that for all three tasks the self-efficacy ratings were higher in the EEEE, EPEP, and PEPE condition compared to the PPPP condition. No differences were found, however, between the EPEP and PEPE Condition. Self-efficacy ratings were also higher after task 2 and task 3 in the EEEE Condition compared to the EPEP and PEPE Condition, but not after training task 4.

Analyses of participants' self-efficacy after the training phase revealed a main effect of Instruction Condition, $H(3) = 66.55$, $p < .001$, and self-efficacy ratings were higher in the EEEE, EPEP, and PEPE condition compared to the PPPP condition. No significant differences were found between the EEEE, EPEP, and PEPE condition. Measuring self-efficacy at the start of the delayed posttest phase revealed the same pattern of results. There was a main effect of Instruction Condition, $H(3) = 46.08$, $p < .001$, and follow-up tests showed that self-efficacy scores were higher in the EEEE, EPEP, and PEPE condition compared to the PPPP Condition. Again, there was no significant difference between EEEE and EPEP or between EPEP and PEPE.

6.3.2 | Perceived competence

Analysis of perceived competence measured after the training phase showed a main effect of Instruction Condition, $H(3) = 67.41$, $p < .001$. Perceived competence was higher in the EEEE, EPEP, and PEPE condition than in the PPPP condition, and scores in the EPEP and PEPE condition did not differ significantly. However, there was no significant difference between the EEEE and EPEP condition. The pattern of results was similar for the delayed posttest. There was a main effect of Instruction Condition, $H(3) = 41.19$, $p < .001$, as perceived competence was higher in the EEEE, EPEP, and PEPE condition than in the PPPP

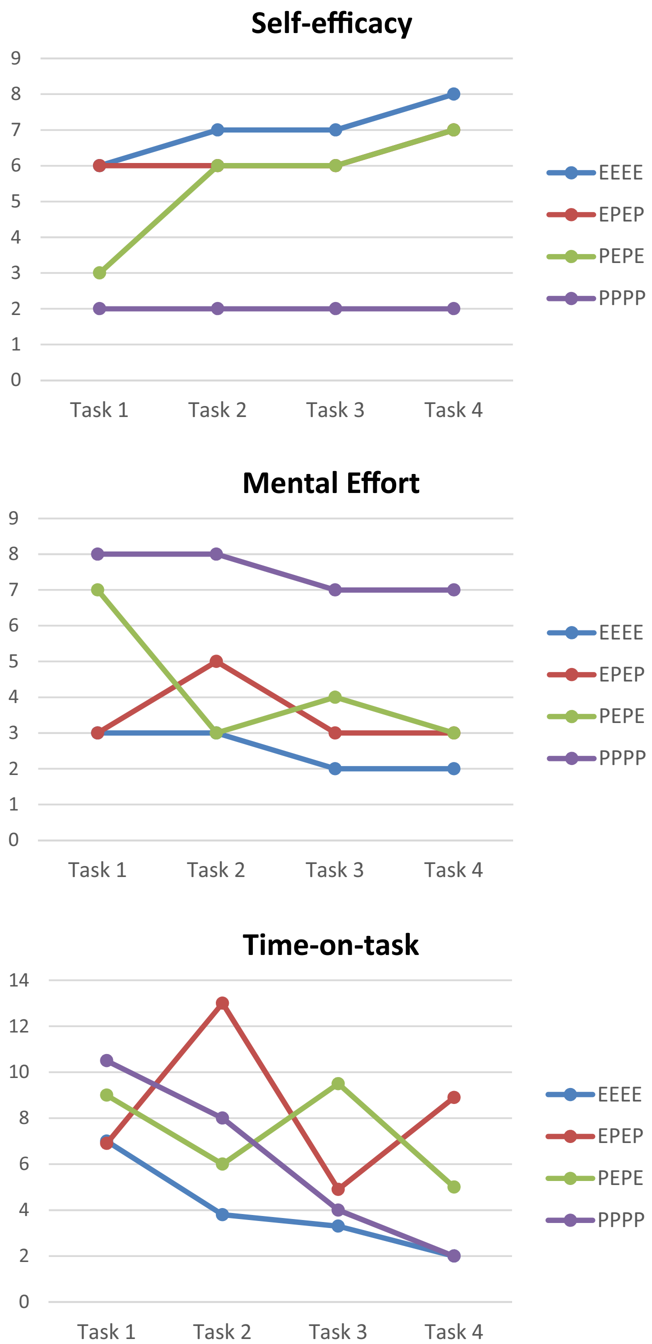

FIGURE 1 Median scores on self-efficacy (top row; range 1–9), mental effort (top row; range 1–9), and time-on-task for each training task in Experiment 1 [Colour figure can be viewed at wileyonlinelibrary.com]

condition. There was no statistically significant difference between the EEEE and EPEP condition or the EPEP and PEPE condition.

6.3.3 | Topic interest

There was a main effect of Instruction Condition, $H(3) = 8.93$, $p = .030$, and there were no differences between the EEEE and EPEP Condition or between the EPEP and PEPE Condition. However, results showed that topic interest scores were lower in the EEEE than in the PPPP Condition. As for topic interest measured before the delayed posttest, there was no main effect of Instruction Condition.

6.4 | How do short sequences of examples and problems affect learning and transfer?

6.4.1 | Isomorphic test tasks

Analyzing whether performance on the isomorphic tasks on the immediate posttest differed among conditions showed a main effect of Instruction Condition, $H(3) = 36.63$, $p < .001$. Results showed that the EEEE, EPEP, and PEPE Condition scored significantly higher than the PPPP Condition. No differences were found between the EEEE and EPEP, EPEP and PEPE, or EEEE and PEPE Condition.

The pattern of results was the same for the isomorphic tasks on the delayed posttest. There was a main effect of Instruction Condition, $H(3) = 24.76$, $p < .001$, and follow-up tests showed that performance on the isomorphic tasks was significantly higher for the EEEE, EPEP, and PEPE Condition than the PPPP Condition. No differences were found between the EEEE and EPEP, EPEP and PEPE Condition, or EEEE and PEPE Condition.

6.4.2 | Procedural transfer task

Analyzing whether performance differed among conditions on the procedural transfer task revealed a main effect of Instruction Condition, $H(3) = 27.41$, $p < .001$. Results showed that the EEEE, EPEP, and PEPE Condition significantly outperformed the PPPP Condition. No differences were found, however, in the other condition comparisons. On the delayed posttest, there was a main effect of Instruction Condition, $H(3) = 10.58$, $p = .014$, and follow-up tests showed that only the EEEE and PEPE Condition, but not the EPEP Condition scored significantly higher than the PPPP Condition on procedural transfer. Again, other comparisons were not significant.

6.5 | How do short sequences of examples and problems affect mental effort and time-on-task in the training phase?

6.5.1 | Mental effort

Mental effort ratings measured after each training task (see Figure 1) were used as a measure of learning efficiency. Results showed a main effect of Instruction Condition for self-reported effort ratings invested in the training tasks, $H(3) = 64.19$, $p < .001$, and the EEEE, EPEP, and PEPE Condition reported less effort during the training phase than the PPPP Condition. Moreover, the EEEE Condition reported less effort than the EPEP and PEPE Condition. Finally, the EPEP Condition also reported significantly less effort than the PEPE Condition.

TABLE 6 Post hoc comparisons of self-efficacy reported after each training task (see Figure 1) in Experiment 1

	EE vs. EP			EE vs. PE			EE vs. PP			EP vs. PE			EP vs. PP			PE vs. PP		
	U	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*
Training task 1	524	.022	.258	80	**<.001**	.761	79	**<.001**	.760	198	**<.001**	.678	191	**<.001**	.680	716	.520	.072
Training task 2	413	**.001**	.384	430	**.008**	.309	96.5	**<.001**	.740	1,041.5	.197	.140	307.5	**<.001**	.567	186.5	**<.001**	.670
Training task 3	441.5	**.002**	.355	359	**.001**	.399	70.5	**<.001**	.772	810.5	.414	.087	175	**<.001**	.698	193	**<.001**	.656
Training task 4	506	.015	.276	479	.039	.242	68	**<.001**	.775	986	.439	.840	173	**<.001**	.696	113.5	**<.001**	.744

Abbreviations: EE, example study only; EP, example-problem pairs; PE, problem-example pairs; PP, problem-solving only. Significant *p*-values (after correction) are bolded.

6.5.2 | Time-on-task

Time-on-task invested in each task in the training phase is presented in Figure 1 and exploratory analyses are presented in Appendix E.

6.5.3 | How do short sequences of examples and problems affect mental effort and time-on-task in the posttest phases?

Exploratory analyses of mental effort and time-on-task invested in the posttest phases are presented in Appendix E.

6.6 | Discussion

Regarding the main aim of uncovering how self-efficacy develops during the training phase, results showed, as expected, that self-efficacy was reported to be significantly higher after the first task for the example-first conditions compared to the problem-first conditions (i.e., EEEE and EPEP > PEPE and PPPP). Throughout the rest of the training phase (i.e., tasks 2 to 4), all example conditions reported significantly higher self-efficacy than the problem-solving only condition, and the EEEE condition reported higher self-efficacy ratings than the EPEP and PEPE condition with regards to training task 2 and 3.

Furthermore, we (partly) replicated the results of Van Harsel et al. (2019) regarding motivational and cognitive aspects of learning measured after the training phase. All example conditions showed higher self-efficacy and perceived competence ratings and test performance (i.e., isomorphic and transfer tasks), while investing less mental effort in the training phase compared to the PPPP condition. All example conditions showed lower effort investment but longer time investment on the isomorphic posttest tasks during the immediate posttest than the PPPP condition. This pattern remained stable on the delayed posttest. Topic interest scores were lower in the EEEE than the PPPP condition on the immediate posttest, but this difference was no longer present on the delayed measurement. There were also no other differences among conditions on topic interest. Importantly, we found no differences on motivational variables (i.e., self-efficacy, perceived competence, or topic interest) or on posttest performance between the EEEE and EPEP, or between the EPEP and PEPE condition. We did find that reported effort investment in the training phase was lower in the EEEE condition than in the EPEP (and PEPE) condition. Effort invested in the training phase was also significantly lower in the EPEP condition than in the PEPE condition.

The results of Experiment 1 provide some evidence for the motivational explanation of differences between EP and PE on learning. Starting the training phase with a practice problem (PE) affected self-efficacy negatively compared to starting with an example. However, this did not lead students in the PE condition to disengage in the present study; they studied the example and after that, their self-efficacy increased to the level of the EP (and EE) condition.

It is an important open question whether the findings on both cognitive and motivational aspects of learning would be different when the training phase is longer (i.e., consists of more training tasks). For example, one might expect that passively studying examples would become redundant and (therefore) boring when task sequences are longer, which in turn might lead to disengagement and lower learning outcomes. Hence, example-problem pairs might be more engaging and effective than example study only, because example-problem pairs provide the benefits of examples but also allow students to actively apply what they have learned. Therefore, a second experiment was conducted with the aim to investigate how motivational and cognitive aspects of learning would be affected by longer task sequences of examples and problems (i.e., eight instead of four tasks: EEEEEEEE, EPEPEPEP, PEPEPEPE, and PPPPPPPP).

7 | EXPERIMENT 2

In Experiment 2, we investigated how longer task sequences of examples and/or practice problems (i.e., EEEEEEEE, EPEPEPEP, PEPEPEPE, and PPPPPPPP) would affect motivational (i.e., self-efficacy, perceived competence, and topic interest measured before and after the training phase) and cognitive aspects of learning (i.e., invested mental effort in the training phase). Time-on-task in the training phase, as well as mental effort and time-on-task in the posttest phases were again measured as (explorative) indicators of efficiency of the learning process and learning outcomes (Van Gog & Paas, 2008). Because example study only might become redundant and boring when task sequences are longer and therefore might lead to disengagement and lower performance scores, we expected that the EPEPEPEP condition would show significantly higher levels of self-efficacy (H2), perceived competence (H3), and topic interest (H4) after the training phase than the EEEEEEEE condition, and that the EPEPEPEP condition would attain higher levels of isomorphic posttest performance (H5), procedural transfer performance (H6), and conceptual transfer performance (H7), while investing less effort in the training phase (H8) compared to the EEEEEEEE condition. All other comparisons were considered exploratory.

7.1 | Method

7.1.1 | Participants and design

Participants were 105 Dutch higher education students in their first year of an electrical and electronic, mechanical engineering, or mechatronics program (M^{age} = 19.30, *SD* = 1.80; 105 male). Participants were randomly assigned to one of four conditions and received eight training tasks: (a) examples only (*n* = 32; EEEEEEEE), (b) example-problem pairs (*n* = 28; EPEPEPEP), (c) problem-example pairs (*n* = 23; PEPEPEPE), or (d) practice problems only (*n* = 22; PPPPPPPP). The experiment consisted of three phases: (a) pretest, (b) training phase, and (c) immediate posttest phase. At the time of the experiment, participants were novices to the modeled task as this subject had not (yet) been a part of their study program. Participants gave their informed consent prior to their inclusion in the study and received study credits for their participation.

7.1.2 | Materials and procedure

The materials were presented using a web-based learning environment. The materials, procedure, and data analysis were the same as in Experiment 1 with the following exceptions. First, the training phase consisted of eight tasks; in addition to the four tasks also used in Experiment 1 two additional pairs of tasks were added. All eight tasks were paired based on their complexity (i.e., pair 1: fitness and energy measurement, pair 2: washing machine and soapsuds, pair 3: drinking water and running, and pair 4: the carousel and coffee consumption). The first pair of tasks required participants to calculate with positive numbers. The second and third pair of tasks were slightly more complex because participants had to calculate with both positive and negative numbers. The fourth pair of tasks was most complex and asked participants to calculate with a cubic function (polynomial of degree 3) instead of the quadratic function (polynomial of degree 2) that was used in the first three task pairs. The design of the formats (i.e., video modeling examples and practice problems) was similar to the formats used in Experiment 1. Second, the immediate posttest consisted of five instead of four tasks as in Experiment 1. Three isomorphic posttest tasks were used (α = .73): one isomorphic to the first pair of training tasks, one to the second and third pair of training tasks, and one to the fourth pair of training tasks. The fourth task was a procedural transfer task (i.e., Simpson rule), followed by the conceptual transfer questions (α = .59).

The procedure was the same as in Experiment 1, with the exception that Experiment 2 did not have a delayed posttest (i.e., in Experiment 1, results were consistent across both test moments and therefore we did not include a delayed posttest). This resulted in 10 single sessions with 2–21 participants per session that lasted ca. 116 min. As for the data analysis, we used the same scoring standards as in Experiment 1 for the training tasks, the three isomorphic posttest tasks (max. Score = 24 points), and the procedural transfer task. Regarding the five conceptual transfer questions, participants could earn a maximum of nine points: one point for the first open-ended question (zero points for an incorrect answer; one point for the correct answer) and two points for the other open-ended questions (zero points for an incorrect answer; one point for the correct answer, two points for the correct answer and a correct explanation).

7.2 | Results

Again, with the exception of pretest performance and topic interest on the immediate posttest, all of the main variables were not normally distributed, with either the kurtosis, skewness, or both coefficients being (substantially) below −1.96 or above +1.96. Again, we used Mann-Whitney *U* tests as post hoc tests (see Table 7). Relevant

TABLE 7 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

	EEEEEEEE condition			EPEPEPEP condition			PEPEPEPE condition			PPPPPPPP condition		
	M	*SD*	Med	M	*SD*	Med	M	*SD*	Med	M	*SD*	Med
Pretest												
Self-efficacy	2.50	1.85	2.00	1.93	1.09	2.00	2.91	1.88	2.00	2.59	1.56	2.50
Perceived competence	2.23	1.41	2.00	1.73	1.00	1.00	2.36	1.54	2.00	1.98	0.91	2.00
Topic interest	4.30	0.87	4.43	4.35	0.70	4.43	4.47	0.91	4.57	4.43	0.81	4.43
Training												
Self-efficacy	6.94	1.45	7.13	6.57	1.19	6.50	6.18	1.57	5.88	3.32	2.24	2.36
Posttest												
Self-efficacy	7.03	1.38	7.00	6.29	1.63	6.00	6.52	1.86	7.00	3.05	2.54	2.00
Perceived competence	5.47	1.94	5.67	5.50	1.40	5.67	5.41	1.25	6.00	2.86	2.04	2.00
Topic interest	4.51	0.69	4.57	4.39	0.68	4.50	3.87	1.00	4.14	4.04	0.95	4.21

Abbreviations: EE, example study only; EP, example-problem pairs; PE, problem-example pairs; PP, problem-solving only.

descriptive statistics of self-efficacy, perceived competence, and topic interest scores are presented in Table 8, and performance scores, mental effort scores, and time-on-task scores are presented in Table 9. Kruskal–Wallis tests showed that there were no significant differences among conditions on pretest performance, $H(3) = 2.86$, $p = .414$, and pretest scores of self-efficacy, $H(3) = 3.94$, $p = .268$, perceived competence, $H(3) = 3.42$, $p = .331$, and topic interest, $H(3) = 1.29$, $p = .731$.

7.3 | How do longer sequences of examples and problems affect self-efficacy, perceived competence, and topic interest?

Self-efficacy

Self-efficacy ratings measured after each training task are presented in Figure 2. First, it was explored whether self-efficacy ratings reported after each training task differed among conditions (see Table 10 for post hoc comparisons). With regard to the *first training task*, there was a main effect of Instruction Condition, $H(3) = 33.45$, $p < .001$, and self-efficacy levels were higher in the EEEEEEEE and EPEPEPEP Condition than the PEPEPEPE and PPPPPPPP Condition. There were no significant differences between the EEEEEEEE and EPEPEPEP Condition or between the PEPEPEPE and PPPPPPPP Condition.

There was also a main effect of Instruction Condition for the *second training task onwards* (task 2: $H(3) = 18.58$, $p < .001$; task 3: $H(3) = 29.12$, $p < .001$; task 4: $H(3) = 32.35$, $p < .001$; task 5: $H(3) = 28.00$, $p < .001$; task 6: $H(3) = 29.52$, $p < .001$; task 7: $H(3) = 30.42$, $p < .001$; task 8: $H(3) = 30.69$, $p < .001$). Results showed that the self-efficacy scores were higher in the EEEEEEEE, EPEPEPEP, and PEPEPEPE Condition compared to the PPPPPPPP Condition. No differences were found, however, between the EPEPEPEP and PEPEPEPE Condition. Also no differences were found between the EEEEEEEE and EPEPEPEP Condition, except for training task 8, where self-efficacy ratings were higher in the EEEEEEEE than EPEPEPEP Condition.

Concerning the main question of whether there would be differences among conditions on self-efficacy ratings measured after the training phase, there was a main effect of Instruction Condition, $H(3) = 29.49$, $p < .001$. Self-efficacy ratings were significantly higher in the EEEEEEEE, EPEPEPEP, and PEPEPEPE Condition compared to the PPPPPPPP Condition. Contrary to our expectations (H2), there were no differences between the EPEPEPEP and EEEEEEEE Condition. Further explorations showed that no other condition comparisons were significant.

Perceived competence

The pattern of results was similar for perceived competence. There was a main effect of Instruction Condition regarding perceived competence measured after the training phase, $H(3) = 23.83$, $p < .001$, and the EEEEEEEE, EPEPEPEP, and PEPEPEPE Condition showed higher perceived competence ratings than the PPPPPPPP Condition. In contrast to our expectations (H3), there was no difference between the EEEEEEEE and EPEPEPEP Condition ($p = .799$, $r = .033$). Further explorations revealed that no other comparisons were significant.

Topic interest

Analyzing whether conditions differed in topic interest scores measured after the training phase revealed a main effect of Instruction Condition, $H(3) = 8.30$, $p = .040$, however, follow-up tests showed no significant differences among any of the condition comparisons (H4).

7.4 | How do longer sequences of examples and problems affect learning and transfer?

Isomorphic test tasks

Analysis revealed a main effect of Instruction Condition for performance on the isomorphic posttest tasks, $H(3) = 12.86$, $p = .005$.

TABLE 8 Mean (M), *SD*, and median (Med) of pretest (range 0–16), isomorphic tasks performance (range 0–24), procedural transfer (range 0–8), conceptual transfer (range 0–9), mental effort (range 1–9), and time-on-task per condition in Experiment 2

	EEEEEEEE condition			EPEPEPEP condition			PEPEPEPE condition			PPPPPPPP condition		
	M	*SD*	Med	M	*SD*	Med	M	*SD*	Med	M	*SD*	Med
Pretest												
Performance	2.03	1.33	2.00	2.00	1.12	2.00	2.74	1.81	3.00	2.36	1.94	2.50
Training												
Mental effort	2.70	1.22	2.56	3.65	1.23	3.81	3.80	1.36	3.75	6.06	2.07	6.31
Time-on-task	2.50	1.29	2.25	7.86	3.16	7.63	5.51	2.51	5.00	6.51	4.26	5.38
Immediate posttest												
Isomorphic tasks	11.94	6.40	12.00	10.43	7.25	11.00	8.22	5.50	8.00	5.63	6.41	5.00
Procedural transfer	2.03	2.56	1.00	1.21	1.97	0.00	2.17	3.23	0.00	0.77	1.97	0.00
Conceptual transfer	3.97	2.48	4.00	3.14	2.66	2.50	4.09	2.02	4.00	3.50	2.72	3.50
Immediate posttest mental effort												
Isomorphic tasks	5.29	1.70	5.67	4.13	1.84	4.17	3.80	1.73	4.00	6.05	2.51	6.33
Procedural transfer	4.78	2.51	5.00	4.82	2.33	5.00	4.00	2.26	5.00	6.59	2.68	7.00
Conceptual transfer	4.22	1.75	5.00	4.00	2.07	3.00	4.13	1.49	5.00	5.18	2.82	5.00
Immediate posttest time-on-task												
Isomorphic tasks	16.13	7.15	16.33	6.69	4.70	6.83	6.07	3.63	4.33	4.00	3.11	3.12
Procedural transfer	5.94	5.12	6.00	2.82	3.17	1.00	3.48	3.36	3.00	2.36	2.98	2.00
Conceptual transfer	7.97	5.43	6.50	4.54	3.42	4.50	5.78	2.75	6.00	5.77	4.02	5.00

Abbreviations: EE, example study only; EP, example-problem pairs; PE, problem-example pairs; PP, problem-solving only.

Results showed that he EEEEEEEE Condition showed significantly higher performance on the isomorphic test tasks than the PPPPPPPP Condition. However, the EPEPEPEP and PEPEPEPE Condition did not significantly differ from the PPPPPPPP Condition. Although we expected EPEPEPEP > EEEEEEEE (H5), there were no performance differences on the isomorphic posttest tasks between the EEEEEEEE and EPEPEPEP Condition. Our explorative analyses showed no other condition comparisons were significant.

Procedural transfer task and conceptual transfer questions

Subsequently, we analyzed whether conditions differed in scores on the procedural transfer task and conceptual transfer questions (H6, H7). Analysis showed there was no main effect of Instruction Condition for the procedural transfer task, $H(3) = 6.04$, $p = .110$, and for the conceptual transfer questions, $H(3) = 2.85$, $p = .415$.

7.5 | How do longer sequences of examples and problems affect mental effort and time-on-task in the training phase?

Mental effort

The average of self-reported effort investment after each task in the training phase (see Figure 2) was analyzed as a measure of efficiency. There was a main effect of Instruction Condition, $H(3) = 34.85$, $p < .001$, and the EEEEEEEE, EPEPEPEP, and PEPEPEPE Condition invested less effort in the training tasks than the PPPPPPPP Condition. As expected (H8), the EEEEEEEE Condition invested significantly less effort in the training tasks compared to the EPEPEPEP Condition, and less effort than the PEPEPEPE Condition. No differences were found between the EPEPEPEP and PEPEPEPE Condition.

Time-on-task

Time-on-task invested in each task in the training phase is presented in Figure 1 and exploratory analyses are presented in Appendix E.

7.5.1 | How do short sequences of examples and problems affect mental effort and time-on-task in the posttest phase?

Exploratory analyses of mental effort and time-on-task invested in the posttest phase are presented in Appendix F.

7.6 | Discussion

The main aim of Experiment 2 was to investigate how longer training task sequences of examples and problems (i.e., EEEEEEEE, EPEPEPEP, PEPEPEPE, and PPPPPPPP) would affect motivational and cognitive variables. It was expected that example study only would result in lower scores on performance and motivational variables than example-problem pairs. In contrast to our hypotheses, however, there were no motivational or test performance differences between the

TABLE 9 Post hoc comparisons of mental effort, self-efficacy, perceived competence, topic interest, isomorphic tasks performance, procedural transfer, and conceptual transfer on the immediate posttest in experiment 2

	EE vs. EP			EE vs. PE			EE vs. PP			EP vs. PE			EP vs. PP			PE vs. PP		
	U	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*
Training																		
Mental effort	652	**.002**	.391	531.5	**.005**	.377	644.5	**<.001**	.701	338.5	.755	.044	502	**.001**	.537	407	**.001**	.522
Immediate posttest																		
Isomorphic tasks	396.5	.444	.099	238	.026	.300	167.5	**.001**	.445	267	.295	.147	188	.018	.335	177.5	.083	.258
Procedural transfer	369	.203	.164	345.5	.677	.056	233	.017	.325	348.5	.568	.080	249	.154	.201	194	.094	.250
Conceptual transfer	361.5	.196	.167	378	.863	.023	314	.500	.092	405.5	.111	.223	331	.650	.064	216	.397	.126
Self-efficacy[a]	328.5	.071	.233	319.5	.397	.114	81.5	**<.001**	.655	360	.464	.103	103	**<.001**	.574	75.5	**<.001**	.606
Perceived Competence[b]	465	.799	.033	368.5	.993	.001	112.5	**<.001**	.577	311	.833	.030	103.5	**<.001**	.569	82	**<.001**	.583
Topic interest[c]	429	.777	.037	221.5	.012	.338	256	.090	.231	204	.025	.314	241.5	.192	.184	281	.524	.095

Abbreviations: EE, example study only; EP, example-problem pairs; PE, problem-example pairs; PP, problem-solving only. Significant *p*-values (after correction) are bolded.

[a]Self-efficacy statistically differed between the pretest and immediate posttest ($Z = 8.16$, $p < .001$, $r = .796$) and increased in EE, EP, and PE Condition ($ps < .001$), not in PP Condition ($p = .303$).

[b]Perceived competence statistically differed between the pretest and immediate posttest ($Z = 8.30$, $p < .001$, $r = .810$) and increased in EE, EP, and PE Condition ($ps < .001$), not in PP Condition ($p = .020$).

[c]Topic interest did not differ statistically between the pretest and immediate posttest ($p = .297$, $r = .102$).

EEEEEEEE and EPEPEPEP condition. As hypothesized, the effort that students reported to invest in the training phase was lower in the EEEEEEEE than the EPEPEPEP condition. However, exploring effort on the posttest phase revealed that levels of perceived effort when solving the isomorphic posttest tasks were higher in EEEEEEEE than EPEPEPEP. This might be explained by the fact that students in the EEEEEEEE condition did not have the opportunity to practice problem-solving in the training phase, whereas the EPEPEPEP condition did have the opportunity to practice problem-solving in the training phase and therefore could apply and automate the procedure several times.

With regard to our exploratory question of how the other conditions would compare to each other, the pattern of results regarding motivational aspects of learning was similar as in Experiment 1. Our exploration of self-efficacy during the training phase showed that there were differences in self-efficacy ratings between the conditions starting with an example and the conditions starting with a practice problem (i.e., EEEEEEEE, EPEPEPEP > PEPEPEPE, PPPPPPPP) regarding the first training task. From the second training task onward, however, self-efficacy ratings in the PEPEPEPE condition increased to the same level as in the conditions starting with an example, whereas self-efficacy in the PPPPPPPP condition remained low. This pattern of results remained stable during and after the training phase, and was also similar for perceived competence. There were no differences among conditions on topic interest.

Regarding performance, only the EEEEEEEE condition significantly outperformed the PPPPPPPP condition on isomorphic test performance, and there was no effect of condition on procedural transfer and conceptual transfer. All example conditions were more efficient in the sense that they reported to invest less effort in the training phase than the PPPPPPPP condition. Again, the EEEEEEEE condition was most efficient considering that they reported to invest the lowest effort levels (and time-on-task) in the training phase. Lastly, no differences in motivational aspects of learning, test performance, or effort investment were found between the EPEPEPEP and PEPEPEPE condition.

8 | GENERAL DISCUSSION

Two experiments were conducted to investigate how different sequences of example study and practice problem-solving (i.e., example study only [EE], example-problem pairs [EP], problem-example pairs [PE], problem-solving only [PP]) would affect motivational (i.e., self-efficacy, perceived competence, and topic interest) and cognitive aspects of learning (i.e., performance on isomorphic and transfer tasks, and mental effort). A short sequence of four training tasks was used in Experiment 1 and a longer sequence of eight training tasks in Experiment 2. We were particularly interested in how participants' self-efficacy would develop during the training phase and whether the pattern of results would remain stable on a delayed posttest (Experiment 1), as well as whether findings would change when the training phase comprised more training tasks (Experiment 2).

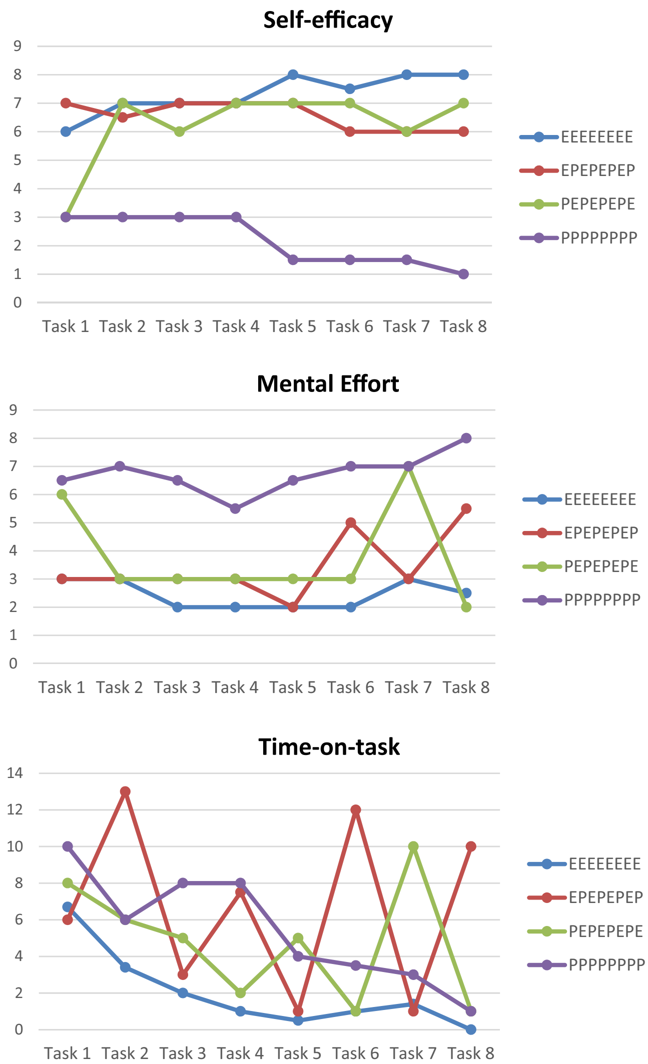

FIGURE 2 Median scores on self-efficacy (top row; range 1–9), mental effort (top row; range 1–9) and time-on-task for each training task in Experiment 2 [Colour figure can be viewed at wileyonlinelibrary.com]

In a training phase with four training tasks, example study (alternated with practice problem-solving) was a more effective (in terms of performance on isomorphic and procedural transfer tasks) and efficient (in terms of mental effort invested in the training and posttest phases) strategy for learning than problem-solving only. We also replicated the findings of Van Harsel et al. (2019): self-efficacy and perceived competence scores were significantly higher after the training phase in all three example conditions compared to problem-solving only. We did find, however, that studying example-problem pairs resulted in lower mental effort investment during the training phase than studying problem-example pairs in Experiment 1. A novel finding is that these effects persisted on a delayed test 1 week later. Experiment 2 showed that with longer sequences, example study (alternated with practice problem-solving) resulted in lower mental effort ratings during the training phase and higher ratings on self-efficacy and perceived competence than problem-solving only. Whereas mental effort was lower during the training phase in the example-problem pairs condition compared to the problem-example pairs condition in Experiment 1, no differences were found between these conditions when sequences were longer as in Experiment 2.

8.1 | Effects of different short sequences on motivation

The findings of Experiment 1 provide evidence for the first part of the motivational explanation regarding the differential effects of EP vs. PE comparisons reported in the literature (cf. Van Harsel et al., 2019; Coppens et al., 2019). That is, starting the training phase with a practice problem (PE, PP) affected self-efficacy negatively compared to starting with an example (EE, EP). However, we found no evidence for the second part of the motivational explanation (i.e., as a consequence of lower self-efficacy levels, students might not be motivated to study subsequent example and probably also the tasks that follow). It seems that in our study, learners did not disengage after starting with a practice problem and studied the example that was provided as a second training task. As a consequence, their levels of self-efficacy increased to the level of the EP (and EE) condition and remained stable during the entire training phase. We must note, though, that using a complex math task might not have resulted in lasting detrimental effects on students' self-efficacy (and perceived competence), because our sample of technical higher education students had experience with similar types of tasks and did not find these tasks unpleasant or uninteresting (topic interest scores were relatively high). Further research is recommended to investigate whether these findings replicate with different learning materials and student populations.

These findings indicate that the benefit of an EP-sequence over a PE-sequence is likely not as large as previously believed (e.g., Van Gog et al., 2011) and may only occur under specific conditions. It is, however, an open question what factor or combination of factors moderate(s) the (small) differential effects of EP versus PE on learning (see small-scale meta-analysis by Van Harsel et al., 2019). In other words, what factors determine whether students will or will not disengage after starting with a practice problem (as they presumably did in prior studies, in which their learning outcomes did not benefit from the examples presented to them; e.g., Kant et al., 2017; Leppink et al., 2014; Van Gog et al., 2011)? It is still possible that other (motivational) variables play a role in determining whether students would disengage. For instance, students in PE conditions might disengage when interest in the learning material is very low, or when the second task consists of a text-based worked example (cf. Van Gog et al., 2011) rather than a video example as used in the present study (which might more easily grab and hold their attention). Hence, future research should further explore what (combination of) factors might moderate the EP-PE effect. We recommend testing larger sample sizes, because a recent meta-analysis indicated that the effectiveness of example-problem pairs as compared to problem-example pairs is rather small (Van Harsel et al., 2019).

TABLE 10 Post hoc comparisons of self-efficacy reported after each training task (see Figure 2) in Experiment 2

	EE vs. EP			EE vs. PE			EE vs. PP			EP vs. PE			EP vs. PP			PE vs. PP		
	U	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*	*U*	*p*	*r*
Training task 1	566	.075	.230	186	**.002**	.423	138.5	**<.001**	.516	105.5	**<.001**	.581	71.5	**<.001**	.661	220.5	.455	.112
Training task 2	419.5	.668	.055	349.5	.749	.043	138.5	**<.001**	.516	323.5	.997	.004	130.5	**<.001**	.496	105.5	**.001**	.505
Training task 3	443.5	.946	.009	274.5	.105	.219	97.5	**<.001**	.615	223.5	.058	.266	74.5	**<.001**	.651	100	**<.001**	.523
Training task 4	426	.738	.043	319	.393	.115	74	**<.001**	.672	287	.494	.096	66	**<.001**	.677	69	**<.001**	.629
Training task 5	405.5	.519	.083	339	.613	.068	83.5	**<.001**	.650	323.5	.977	.004	100	**<.001**	.582	69	**<.001**	.618
Training task 6	279.5	.011	.328	322	.423	.108	85	**<.001**	.630	397.5	.144	.205	110	**<.001**	.557	83.5	**<.001**	.583
Training task 7	274.5	.009	.338	233	.019	.316	83.5	**<.001**	.647	317	.923	.013	112	**<.001**	.548	99.5	**<.001**	.527
Training task 8	265.5	**.006**	.356	299	.230	.162	93.5	**<.001**	.629	403.5	.116	.220	115.5	**<.001**	.553	78.5	**<.001**	.599

Abbreviations: EE, example study only; EP, example-problem pairs; PE, problem-example pairs; PP, problem-solving only. Significant *p*-values (after correction) are bolded.

8.2 | Effects of longer task sequences on performance and motivation

Another noteworthy finding is that longer task sequences did not necessarily result in better learning outcomes when we visually compare the results of Experiments 1 and 2, except in the examples only condition. Studying examples only remained very effective, efficient, and motivating even with longer sequences. This is at first glance surprising in light of the expertise-reversal effect, which proposes that examples become less conducive to learning than practice problems for learners with more prior knowledge (e.g., Kalyuga et al., 2001). Moreover, one might expect that studying examples only, which is more passive, could be less motivating (i.e., more boring) than alternating examples and problems (cf. Sweller & Cooper, 1985), especially with longer sequences. This could, in turn, lead to disengagement and lower learning outcomes, but we found no evidence that this was the case. It should be noted, though, that the training tasks increased in complexity during the training phase (i.e., after the second task in Experiment 1 and 2 and after the sixth task in Experiment 2). Although the problem-solving procedure remained the same, this may have prevented students from experiencing the examples as too repetitive. Moreover, we provided participants the opportunity to study examples and/or solve practice problems in a self-paced instead of a system-paced learning environment. Although participants were instructed to watch the entire example, it was possible to skip (parts of) the video modeling example. As evidenced by the time-on-task data that was obtained during training phase, time spent on the examples decreased as the learning phase progressed, and this control over the video examples may also explain why participants did not disengage during example study only. Further research should investigate whether the overall findings replicate, and under what circumstances studying longer sequences of examples only remains effective, efficient, and motivating for learning.

8.3 | Limitations

There are also some limitations to this study. The first limitation is that we did not directly manipulate sequence length (i.e., four vs. eight training tasks) as a between-subject factor in one single experiment, which would have allowed us to test for interaction effects between the length of the task sequence and the outcome variables. That being said, the pattern of results in Experiment 1 and 2 is highly similar and thus seems to reinforce each other. Secondly, a strength of our study was the use of a conceptual pretest. A procedural pretest (as used in the prior study by Van Harsel et al., 2019) might have led students in the example-first conditions to feel that they started the learning phase with practice problem-solving. Yet, we did not experimentally vary the type of pretest within the present experiments, and therefore cannot draw definite conclusions about the potential effects of a procedural vs. conceptual pretest. That being said, when we compare the findings from the present study (with a conceptual pretest) to the prior study (with a procedural pretest; Van Harsel et al., 2019) the results are highly similar: There is no evidence for an advantage of example-problem pairs. Thirdly, we "only" used two different task length sequences. The findings might be different with short training phases comprised of two tasks, where students provided with a PE-sequence would only have one example to study after starting with failed practice problem. Lastly, it is as of yet an open question whether example study would become less effective and motivating with even longer sequences. Hence, future research is recommended to experimentally manipulate how many tasks students receive during the training phase and to cover a broader range of possible sequence length manipulations which take into account the (increase of) complexity level of the training tasks.

Another limitation concerns the self-efficacy and perceived competence measures. The use of a 9-point scale for the self-efficacy measurement raises the question of whether students are really able to report their task-specific confidence on such a fine-grained level—the same question arises when asking students to report their effort investment on a 9-point scale. A factor that might also have influenced the self-efficacy measurements during the learning phase is whether students could estimate their task-specific confidence based on an actual attempt to solve the problem or just the imagination of doing so after studying the example. Moreover, it has been questioned whether or not (task-specific) self-efficacy and perceived competence are really separate constructs. Literature shows that perceived competence may be a common core component of both self-efficacy and measures of self-concept (e.g., Bong & Skaalvik, 2003; Marsh et al., 2019; Schunk & Pajares, 2005). In line with this notion, the pattern of results on self-efficacy and perceived competence was nearly identical in both experiments and the correlations between these two constructs on the measurement after the training phase were extremely high in Experiment 1 (.96) and Experiment 2 (.92). As such, the use of one of the measures might suffice in future research in this area.

8.4 | Conclusions

In sum, our results have shown that studying examples only—possibly alternated with practice problem-solving—is more effective and efficient for novices' learning than practice problem-solving only. These results were established with higher technical education students and a mathematical problem-solving task. However, based on the large body of research on the worked example effect (see for a review Van Gog et al., 2019), it seems safe to assume that these effects would generalize to other problem-solving tasks and populations as well. A new finding of our study was that examples had clear effects on motivational aspects of learning (i.e., self-efficacy and perceived competence); so far, little is known about the effects of different example and problem sequences on motivation (Renkl, 2014; Sweller, Ayres, & Kalyuga, 2011; Van Gog & Rummel, 2010). Moreover, a new and interesting finding both from a theoretical and practical perspective, is that example study only can remain more effective, efficient, and motivating for learning than solving practice problems only when longer sequences are studied. However, because our study is among the first to examine the effects of different short and longer sequences of examples and problems on student motivation, an open question that needs to be addressed in future research is whether these results generalize to other populations, domains, and materials.

8.5 | Implications for practice

Our results could be interesting and relevant for educators who are instructing new knowledge or skills to novices, for students who have to learn new knowledge or skills through self-study, and also for instructional designers who are designing learning materials. Our results suggest that, when studying short sequences of examples and problems, it is more preferable to study or provide examples (probably alternated with problem-solving) instead of practicing problem-solving only, from both a cognitive and a motivational perspective. Moreover, even with longer sequences, example study remains very effective, efficient and motivating, however, future research should further investigate under what specific conditions example study remains effective in longer learning phases. Secondly, it is advisable to start training phases with an example instead of a problem. Although we did not find any differences in test performance and student motivation between example-problem pairs and problem-example pairs, our results showed that starting the training phase with an example is more efficient for learning than starting with a practice problem.

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 for facilitating this study. The authors would also like to thank Jos van Weert, Rob Müller, and Bert Hoeks for their help in developing the materials, Marlies Otten, Lottie Raaijmakers, Maud Muskens, Gabrielle Peeters, Marleen de Haan, and Hans Slabbekoorn for their help with the data collection, and Susan van Ravensbergen for her help with scoring the data.

ORCID

Milou van Harsel https://orcid.org/0000-0003-4883-7494

ENDNOTE

[1] A possible explanation for the low reliability of the pretest could be the unfamiliarity of the participants with the subject matter (indeed, it was meant as a check that students were indeed novices regarding those tasks). Because the pretest consisted of multiple-choice questions and "I do not know" was not included as an answer option, students would have had to guess, which likely resulted in low reliability of the pretest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author.

ORCID

Milou van Harsel https://orcid.org/0000-0003-4883-7494

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

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

Baars, M., Van Gog, T., De Bruin, A., & Paas, F. (2017). Effects of problem solving after worked example study on secondary school children's monitoring accuracy. *Educational Psychology*, *37*(7), 810–834. https://doi.org/10.1080/01443410.2016.1150419

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

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

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

Coppens, L. C., Hoogerheide, V., Snippe, E. M., Flunger, B., & Van Gog, T. (2019). Effects of problem-example and example-problem pairs on gifted and nongifted primary school students' learning. *Instructional Science*, *47*(3), 279–297. https://doi.org/10.1007/s11251-019-09484-3

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

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

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, *60*(4), 549–571. https://doi.org/10.3102/2F00346543060004549

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.learninstruc.2016.02.004

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

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science*, *38*(3), 209–215. https://doi.org/10.1007/s11251-009-9102-0

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, *96*(3), 558–568. https://doi.org/10.1037/0022-0663.96.3.558

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

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

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology*, *111*(2), 331–353. https://doi.org/10.1037/edu0000281

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

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

Renkl, A. (2014). Towards an instructionally-oriented theory of example-based learning. *Cognitive Science*, *38*, 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.

Roelle, J., & Berthold, K. (2013). The expertise reversal effect in prompting focused processing of instructional explanations. *Instructional Science*, *41*(4), 635–656. https://doi.org/10.1007/s11251-012-9247-0

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

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., & Pajares, F. (2005). Competence perceptions and academic functioning. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 85–104). New York, NY: Guilford Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer. https://doi.org/10.1007/978-1-4419-8126-4

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/s1532690xci0201_3

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 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

Van Gog, T. (2011). Effects of identical example-problem and problem-example pairs on learning. *Computers & Education*, *57*(2), 1775–1779. https://doi.org/10.1016/j.compedu.2011.03.019

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., Dirkx, K., Hoogerheide, V., Boerboom, J., & Verkoeijen, 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. https://doi.org/10.1007/s10648-015-9297-3

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., Rummel, N., & Renkl, A. (2019). Learning how to solve problems by studying examples. In J. Dunlosky & K. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 183–208). New York, NY: Cambridge University Press.

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 Harsel, M., Hoogerheide, V., Verkoeijen, P., & van Gog, T. (2019). Effects of different sequences of examples and problems on motivation and learning. *Contemporary Educational Psychology*, *58*, 260–275. https://doi.org/10.1016/j.cedpsych.2019.03.005

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

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: van Harsel M, Hoogerheide V, Verkoeijen P, van Gog T. Examples, practice problems, or both? Effects on motivation and learning in shorter and longer sequences. *Appl Cognit Psychol*. 2020;34:793–812. https://doi.org/10.1002/acp.3649