ORIGINAL RESEARCH

How do higher education students regulate their learning with video modeling examples, worked examples, and practice problems?

Milou van Harsel[1,2] · Vincent Hoogerheide[2] · Eva Janssen[2] · Peter Verkoeijen[1,3] · Tamara van Gog[2]

Abstract
Presenting novices with examples and problems is an effective and efficient way to acquire new problem-solving skills. Nowadays, examples and problems are increasingly presented in computer-based learning environments, in which learners often have to self-regulate their learning (i.e., choose what type of task to work on and when). Yet, it is questionable how novices self-regulate their learning from examples and problems, and to what extent their choices match with effective principles from instructional design research. In this study, 147 higher education students had to learn how to solve problems on the trapezoidal rule. During self-regulated learning, they were free to select six tasks from a database of 45 tasks that varied in task format (video examples, worked examples, practice problems), complexity level (level 1, 2, 3), and cover story. Almost all students started with (video) example study at the lowest complexity level. The number of examples selected gradually decreased and task complexity gradually increased during the learning phase. However, examples and lowest level tasks remained relatively popular throughout the entire learning phase. There was no relation between students' total score on how well their behavior matched with the instructional design principles and learning outcomes, mental effort, and motivational variables.

✉ Milou van Harsel
   m.vanharsel@avans.nl

1   Learning and Innovation Centre, Avans University of Applied Sciences, PO Box 90116, 4800 RA Breda, The Netherlands

2   Department of Education, Utrecht University, Utrecht, The Netherlands

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

Introduction

Problem solving is important in many curricula, especially in the domains of science, technology, engineering, and mathematics (STEM; Van Gog et al., 2020). Most problems students encounter in (the initial years of) STEM curricula are well-defined (algorithmic) problems, in which students have to learn to perform the procedure to get from an initial state to a described goal state (Newell & Simon, 1972). Problem solving is happening more and more in (online) computer-based learning environments in which different tasks are often embedded (e.g., Roll et al., 2011), such as video modeling examples (i.e., a model demonstrating and possibly explaining the solution procedure step by step on video), worked examples (i.e., a written step-by-step explanation of a full and correct solution procedure of how to solve a problem), and practice problems that students have to try to solve themselves. A popular example of such an environment is Khan Academy (www.khanacademy.org), where students can decide for themselves which type of tasks to work on (i.e., examples or problems), for how long, and in which order.

When acquiring problem-solving skills in computer-based learning environments where little support or guidance is available, it is important that students can adequately self-regulate their learning from examples and problems. Although there are different theoretical models of self-regulated learning (see Panadero, 2017), these models all agree that self-regulated learning requires students to plan, monitor (i.e., track), and control their learning (e.g., Nelson & Narens, 1990; Winne & Hadwin, 1998; Zimmerman, 1990). Monitoring and control are very important factors in self-regulated learning, also of problem-solving skills, as research has shown that learners who accurately monitor and control their learning show higher learning outcomes (e.g., Kostons et al., 2012). The processes of monitoring and control during self-regulated learning are illustrated by the model of Nelson and Narens (1990). In their model (see Fig. 1), an object level (the actual learning process) and a meta level (representation of the learning process) are represented. The object-level informs the meta-level (i.e., monitoring), and in turn, the meta-level can change (i.e., control) the object-level. Translating this model to self-regulated learning of problem-solving tasks, it requires students to monitor their progress while performing the task they have selected during the planning phase and judge their performance after the task is completed (monitor), and to use this as input for deciding what subsequent task to work on (i.e., control; decide which task suits their learning needs best, e.g., De Bruin & Van Gog, 2012; Van Gog et al., 2020).

Fig. 1 The relation between monitoring and control (Nelson & Narens, 1990)
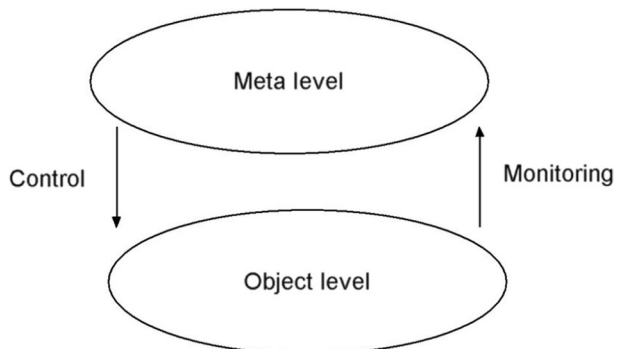
However, little is known about how students regulate their learning of problem-solving tasks, that is, about when they choose examples or practice problems, and why (i.e., what reasons underlie their choices; e.g., Van Gog et al., 2019, 2020). Moreover, it is an open question how well students' choices align with what we know to be effective, efficient, and motivating principles for acquiring new problem-solving skills from many years of instructional design research. Therefore, the present study addresses those questions.

Learning from examples and problems at different complexity levels

Instructional design research has uncovered several principles on how to optimize the acquisition of new problem-solving skills for novices (i.e., students with little, if any, prior knowledge). These principles are concerned with how to ensure that novices work on tasks that provide an optimal level of instructional support and complexity given their current level of knowledge (4C/ID Model; Van Merriënboer, 1997; Van Merriënboer & Kirschner, 2013). With regard to instructional support, the *worked examples principle* (Renkl, 2014; Sweller et al., 2011; Van Gog et al., 2019) states that for novices studying several examples (possibly alternated with solving practice problems) leads to better test performance (i.e., is more effective) attained with less time and/or effort investment (i.e., is more efficient) than practice problem solving only. Since this applies not only to worked examples (i.e., a written step-by-step explanation of how to solve a problem) but also to video modeling examples (i.e., a person demonstrating and/or explaining a problem-solving procedure on video), we call this the *example-based-learning-principle*.

A second robust finding is the *example-first-principle*, which says that novice learners should not only study several examples while learning, but also start the learning phase with an example instead of practice problem solving, because research has consistently shown that sequences of example study only or example-problem pairs –but not problem-example pairs– are more effective and efficient for learning than solving practice problems only (e.g., Kant, et al., 2017; Van Gog et al., 2011; Van Harsel et al., 2020). Recent studies have also shown that example study only, or example study alternated with problem solving, is also more beneficial for motivational aspects of learning (i.e., self-efficacy and perceived competence) than problem solving only (e.g., Coppens et al., 2019; Van Harsel et al., 2019, 2020). We must note, though, that learners do not need the instructional support provided by examples anymore when their knowledge increases. From that point onwards, they learn more from solving problems than from example study (i.e., the expertise reversal effect; Kalyuga et al., 2003).

Finally, problem-solving tasks that are presented in school curricula (either via online computer-based environments or in textbooks/workbooks) often span multiple complexity levels. Task complexity is determined by the number of elements in a learning task and the interaction between those elements (e.g., Sweller & Chandler, 1994). Simple learning tasks consist of a few information elements and a small number of interactions between elements that need to be processed simultaneously in working memory. With increasing numbers of information elements and interactions between elements, task complexity (and working memory load) increases (e.g., Pollock et al., 2002; Van Zundert et al., 2012).

When tasks span multiple complexity levels, learners should not only be working on tasks that provide optimal support given their current level of knowledge, but also on tasks that are at an optimal level of complexity (see the 4C/ID model; Van Merriënboer, 1997; Van Merriënboer & Kirschner, 2013), as learning outcomes and motivation might suffer when tasks are too complex (Van Merriënboer et al., 2003). Therefore, novices should start

with a task at the lowest complexity level (i.e., *lowest-level-first-principle*) and build up tasks in such way that the level of complexity gradually increases (i.e., *simple-to-complex-principle*). When the choice is made to move up a complexity level, learners often need instructional support again (cf. 4C/ID model). Therefore, it becomes important to start each new complexity level with example study (i.e., *start-each-level-with-example-principle*).

Self-regulated learning from examples and problems

These instructional design principles provide clear guidelines on what works best when learning from examples and problems at different complexity levels. However, the question is whether students would spontaneously apply these principles when selecting examples and problems at different complexity levels during self-regulated learning in a computer-based learning environment. As mentioned earlier, for effective self-regulated learning of problem-solving tasks, students need to be able to self-assess their performance on a task just completed and then select a next task with the right level of support and complexity (e.g., De Bruin & Van Gog, 2012). There are, however, both empirical and theoretical reasons to believe that learners will engage in suboptimal task selection when they are left to their own devices (e.g., Azevedo et al., 2008; Niemiec et al., 1996).

Firstly, self-regulated learning research has shown that learners' estimation of their own task performance (or knowledge) is often not in line with their actual performance (e.g., Bjork, 1994, 1999; Kostons et al., 2010, 2012; Rawson & Dunlosky, 2007), particularly for novices (e.g., Dunning et al., 2004; Koriat & Bjork, 2005). Inaccurate self-assessments are a major problem when learners are in control of task selection, because for learning to be effective and efficient, learners need to select a task at an optimal level of instructional support and complexity given their current level of performance. Novices who overestimate their performance might select a task that is too complex and/or does not provide the necessary instructional support, while those who underestimate their performance might select a task that is too easy (e.g., Dunlosky & Rawson, 2012). As a result, learners will end up working on tasks that are not aligned with their learning needs, which might negatively affect their performance on domain specific knowledge or skills and motivation.

Secondly, research has shown that novices often experience difficulties discerning which task aspects are relevant for learning when selecting their own learning tasks (e.g., Quilici & Mayer, 2002), probably because they lack domain knowledge and/or task-selection skills (i.e., knowing about relevant task-selection aspects and combining this with characteristics of available learning tasks; e.g., Van Merriënboer et al., 2006). As a consequence, novices might select tasks based on surface features used to exemplify the problem-solving procedure (e.g., cover story) rather than structural features that are (more) relevant for learning (e.g., the level of complexity and instructional support; Corbalan et al., 2008).

A recent study conducted by Foster et al. (2018) provided some evidence for the idea that learners also show suboptimal behavior when they can select their own task format in the form of examples and practice problems. In their study, university students (novices) learned how to solve probability calculation problems in an (online) computer-based learning environment. Students received 12 (Experiment 1) or 24 (Experiment 2 and 3) probability problems and could decide whether they wanted to study it in the form of a worked example or a practice problem or opt for completion problems (Experiment 2 and 3), which are partially worked-out examples that provide a medium level of support but require learners to complete some steps themselves (e.g., Paas, 1992; Renkl & Atkinson, 2003; Van Merriënboer et al., 2002). One would expect that novices learn most when they

select more examples than problems and start the learning phase with a worked example rather than a (completion) problem. However, results of all experiments showed that problems were selected more frequently on average than examples (at odds with the example-based-learning-principle), and that students rarely chose a worked example as a first task.

In sum, little research has been conducted on how learners regulate their learning from examples and problems (at different complexity levels) and how well their selection behavior matches with evidence-based principles from instructional design research. The few studies available, suggest that learners' task-selection behavior does not align with these principles (Foster et al., 2018). It is important to get more insight in what learners do (and why) when they can determine themselves how to learn new problem-solving skills and to what extent there decisions align with principles that are effective for learning, as self-regulated learning of problem-solving skills in computer based learning environments is becoming more common. Having this information can help teachers and instructional designers to determine whether and what instructional support or advice novices might need to optimally self-regulate their learning from examples and practice problems (at different complexity levels).

The present study

This study investigated how higher education students (novices on the to-be-learned topic) regulate their learning in a computer-based learning environment when they select their own learning tasks from a task database comprising video modeling examples, worked examples, and practice problems of varying levels of complexity and with different cover stories. We decided to provide the option of worked examples *and* video modeling examples because both are widely used in computer-based learning environments, yet it is largely unclear which example types students prefer (at which phase in the learning process). For example, Hoogerheide et al. (2014) compared the effects of worked examples to video modeling examples with two samples of secondary education students. Although they found no differences between the two example formats on cognitive (i.e., test performance, mental effort) and motivational aspects of learning (i.e., self-efficacy, perceived competence), there was an effect on the degree to which students preferred to receive instruction in a similar manner in the future. When only one example was studied (Experiment 2), the video modeling example condition gave a higher preference rating (at least numerically; $p = .07$), but when two examples were studied (Experiment 1), the worked example condition gave a higher preference rating ($p = .03$).

These findings might suggest that students would prefer video modeling examples at the beginning of a learning phase and worked examples later in the training phase. A possible explanation for why students would prefer to start with a video modeling example instead of a worked example could be that in video modeling examples, information is demonstrated in a step-by-step manner and that the combination of dynamic visual information and the model's narration take the learner by the hand. Worked examples can be overwhelming because all the information is presented simultaneously, and it might be easy to ignore written text. However, because all the information is presented simultaneously, worked examples do allow for efficiently looking up difficult problem-solving steps more easily than video modeling examples (in which the information is presented in succession).

This study had three research questions. First, it was explored what tasks technical higher education students select and why, when learning from examples and problems at different complexity levels? Second, it was explored to what extent students' task

selections match with instructional design principles derived from research on optimizing the acquisition of new problem-solving skills for novices. Given the paucity of research on what learners do when they are in charge of learning a new problem-solving skill with the help of (different) examples and problems at different complexity levels, we refrain from formulating explicit hypotheses and consider this study as exploratory in nature. Third, we investigated whether there is a positive relation between the extent to which students choices match with the instructional design principles and their scores on learning outcomes, mental effort, and motivational variables. Given how much evidence there is for the instructional design principles, one would expect a positive relationship between the extent to which learners' choices match with these principles and scores on learning outcomes (i.e., isomorphic tasks, procedural transfer task, and conceptual questions) and motivational aspects of learning (i.e., self-efficacy, perceived competence, and topic interest), and a negative relationship with mental effort during the learning phase.

Method

Participants and design

Participants were 180 Dutch higher education students enrolled in the first year of an electrical and electronic mechanical engineering program ($M_{age} = 19.00$, $SD = 1.64$; 169 male, 11 female). All participants were assigned to a computer-based learning environment in which they had to learn a mathematical problem that required them to approximate the region under a graph using the trapezoidal rule. The trapezoidal rule is an integration method that can be used to approximate the area under a curve by dividing that area into trapezoids or "strips" (rather than using rectangles). By adding up the surface of the "strips", one can approach the total area under that curve (for more information, see https://en.wikipedia.org/wiki/Trapezoidal_rule). The environment consisted of three phases: (1) pretest, (2) self-regulated learning phase, (3) and posttest. We excluded 10 participants who did not finish the isomorphic (and transfer) items on the posttest on time, and seven participants of whom (part of the) learning phase data was missing due to a programming error. Because we were interested in the task-selection behavior of novice learners, we also excluded 16 participants who had too much prior knowledge, indicated by a score of 5 or more (out of 10) on the prior knowledge test. Therefore, the final sample consisted of 147 participants ($M_{age} = 18.90$, $SD = 1.64$; 139 male, 8 female). Participants gave their informed consent in the learning environment. The experiments were scheduled as regular math classes, because integration is one of the math topics in the curriculum of first year electrical and electronic mechanical engineering students and therefore the trapezoidal rule was of direct use to these students in their study programs. However, attendance was optional. Students who did attend received study credits. Moreover, students' participation in the study was voluntary, so they could opt out of having their data used for research purposes.

Materials

The materials were based on the materials developed by Van Harsel et al. (2019, 2020) and presented in a web-based learning environment.

Learning tasks

The task database contained 45 learning tasks. These tasks varied in complexity level, task format, and cover story (for an overview, see Fig. 2).

Complexity level Tasks could be selected at three levels of complexity. *Level 1* tasks required participants to use the trapezoidal rule in problems that always contained a polynomial degree of 2 (i.e., quadratic function). These problems also required participants to calculate more than two intervals and calculate with fractions and positive numbers only. *Level 2* tasks were more complex than *Level 1* tasks, because participants were asked to calculate with negative numbers. The negative number changes the relation between information elements. That is, calculating with negative numbers requires students to take into account an additional rule (i.e., relation between elements) compared to calculating with positive numbers (i.e., subtracting a negative number from a positive number, turns the two signs into a plus sign; $5 - 7 = 12$). Moreover, in more complex calculations with negative numbers (i.e., large functions using brackets, exponents, different arithmetic operations), the order of the arithmetic operations is important. *Level 3* tasks were, in turn, more complex than *Level 2* tasks, because students had to calculate with a cubic polynomial (i.e., cubic function) instead of a quadratic polynomial. A cubic polynomial has a term more than a quadratic polynomial, which increases the number of information elements and relations students have to calculate with.

Task format Within each complexity level, participants could choose from three task formats, namely, video modeling examples, worked examples, and conventional practice problems (see Fig. 2). Each *video modeling example* displayed a computer screen recording of a female model who demonstrated (with handwritten notes) and verbally explained how to solve a mathematical problem step-by-step, using the trapezoidal rule. The screen recording started with a brief introduction on the trapezoidal rule and an explanation of a specific problem state. Subsequently, the model explained how to interpret the information that was given to solve the problem (i.e., the graph of a function, the left border and right border of the area, the number of intervals, and the formula of the trapezoidal rule). Finally, she demonstrated and explained how to solve the problem by undergoing four steps: (1) 'compute the step size of each subinterval', (2) 'calculate the x-values', (3) 'calculate the function values for all x-values', and 4) 'enter the function values into the formula and calculate the

Level 1			Level 2			Level 3		
Video modeling examples	Worked examples	Practice problems	Video modeling examples	Worked examples	Practice problems	Video modeling examples	Worked examples	Practice problems
Drinking beer	Drinking beer	Drinking beer	Running	Running	Running	Carousel	Carousel	Carousel
Energy measurement	Energy measurement	Energy measurement	Drinking water	Drinking water	Drinking water	Rowing	Rowing	Rowing
Fitness	Fitness	Fitness	Washing machine	Washing machine	Washing machine	Perfume	Perfume	Perfume
Fuel consumption	Fuel consumption	Fuel consumption	Soapsuds	Soapsuds	Soapsuds	Coffee consumption	Coffee consumption	Coffee consumption
Traffic	Traffic	Traffic	Plasterer	Plasterer	Plasterer	Chocolate party	Chocolate party	Chocolate party

Fig. 2 Screenshot of the task database

area'. The written information on previously completed steps remained visible on the screen while the model worked on and explained the next step.

Each *worked example* was presented on one page. The worked examples also started with a short description of the problem state and participants received some additional information that was needed to solve the problem (i.e., the graph of a function, the left border and right border of the area, the number of intervals, and the formula of the trapezoidal rule). This was followed by the written out solution procedure that showed students how to solve each step of the problem (the problem state, additional information, and written explanations and correct answers on all steps were simultaneously visible on the screen).

The *practice problems* were presented on one page and consisted of the problem state and some additional information on how to solve the problem (i.e., the graph of a function, the left border and right border of the area, the number of intervals, and the formula of the trapezoidal rule), with by the following assignment: "Approach the area under the graph using the information that is given. Write down all your intermediate steps and calculations". Participants did not receive any feedback on their answers. A screenshot of a practice problem, video modeling example, and worked example are given in online Appendix A, B, and C.

Cover story In addition to selecting a complexity level and instructional format, students could also choose their own cover story. At each complexity level, participants could choose between five different cover stories (see Fig. 2). For example, they could approximate how many liters of beer is tapped within a certain amount of time (i.e., drinking beer) or approximate how often the circular platform of a carousel rotates in a given period of time (i.e., carousel). The cover stories were similar for each task format that was provided within a complexity level (e.g., drinking beer could be selected as video modeling example, worked example, and practice problem), yet the numbers used differed per task format.

Dependent variables

Test tasks

The *pretest* was a conceptual prior knowledge test that consisted of five questions (i.e., multiple choice questions with explanation part) that aimed to measure participants' understanding of the trapezoidal rule. Cronbach's alpha in the current sample was 0.33. Each multiple-choice question had four answer options (i.e., a, b, c, and d) and an 'explanation' part where participants had to explain their answer. The *posttest* consisted of five tasks. The first three tasks were isomorphic to the tasks in the self-regulated learning phase (i.e., a level 1, 2, and 3 task). Cronbach's alpha in the current sample was 0.74. The fourth task was a procedural transfer task that required participants to use the Simpson rule (instead of the trapezoidal rule) to approximate the definite integral under a graph. Simpson's rule is also a numerical integration method to approximate the integral of a function. Although both procedures look almost similar, Simpsons' rule uses quadratic polynomials (instead of the straight line segments). The fifth task consisted of five open-ended conceptual questions that aimed to measure participants' understanding of the trapezoidal rule, and these were isomorphic to the questions in the pretest. Cronbach's alpha in the current sample was 0.44. An example of a conceptual pretest item, an isomorphic posttest task, a procedural transfer task, and a conceptual posttest item is shown in Online Appendix D.

Mental effort

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". Mental effort was rated after each task in the self-regulated learning phase and the posttest phase, except for the five conceptual posttest questions (where it was rated only once after the last question).

Self-efficacy

After the pretest, during the self-regulated learning phase (i.e., after each learning task), and before the posttest, participants were asked to rate to what extent they were confident that they could approximate the definite integral of a graph using the trapezoidal rule. A 9-point rating scale was used, ranging from (1) "very, very unconfident" to (9) "very, very confident" (Van Harsel et al., 2020; adapted from Hoogerheide et al., 2016).

Perceived competence

Perceived competence was measured using the *Perceived Competence Scale for Learning* (Van Harsel et al., 2019, 2020; based on Williams & Deci, 1996; Williams et al., 1998). This perceived competence scale 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. Cronbach's alpha in the current sample was 0.93.

Topic interest

Finally, participants' interest in the topic was measured with a topic interest scale, comprised of seven items (Van Harsel et al., 2020; adapted from the topic interest scale by Mason et al., 2008, and the perceived interest scale by Schraw et al., 1995). Participants were asked to rate to what degree each of the items applied to them on a 7-point scale (1: totally disagree, to 7: totally agree). Cronbach's alpha in the current sample was 0.82. All items are shown in online Appendix E.

Task-selection questionnaire

To shed light on why participants selected the learning tasks that they did, we developed a questionnaire. This questionnaire consisted of five questions, each with a multiple-choice (mc) and open-answer part, namely: (1) What was the format of the first task you chose (mc: video modeling example, worked example, practice problem) and why (open answer)?, (2) What was the level of complexity of the first task (mc: level 1, level 2, level 3) and why (open answer)?, (3) What was the format of the second task you chose (mc: video modeling example, worked example, practice problem) and why (open answer)?", (4) What was the level of complexity of the second task and why (open answer)?, and (5) Which task format did you choose most often and why (open answer)?

Procedure

The study was run in sixteen sessions with 7 to 28 participants per session. The sessions lasted 116 min on average and took place in a computer classroom at participants' higher education institute. Each participant received a headset, pen, and scrap paper to write down their calculations. The session started with the experimenter explaining the aim and procedure of the study. Participants were told that they were going to learn a mathematical task in the learning environment by selecting their own learning tasks. Participants were also instructed that they could work at their own pace (with a maximum of 135 min). Moreover, they received the instructions to write down as much as possible and to write an "X" if they really did not know what to answer.

After the instructions, participants entered the learning environment. In the environment, tasks and questionnaires were presented on a separate page, and participants were unable to go back to the previous pages or to look forward to the next page before completing the current task or questionnaire. Time was logged for each task. Participants were first presented with, in order, a short demographic questionnaire (e.g., age, gender, and prior education), the pretest, and the self-efficacy, perceived competence, and topic interest questionnaire. Then, participants entered the self-regulated learning phase. To ensure that participants had some knowledge of the task database and how to select their own tasks, they first received an explanation of the task database. A picture of the task database was presented on the screen. Participants were instructed to select six learning tasks of their own choice from a task database containing 45 tasks that differed in format (video modeling examples, worked examples, and practice problems), complexity level (level 1, 2, and 3) and cover story. They were also told that each task could only be selected once, and that there was a maximum of 10 min to watch, study, or solve each task.

After the self-regulated learning phase, participants completed the self-efficacy, perceived competence, and topic interest questionnaires again. Participants were instructed to turn their scrap paper upside down and given a new scrap paper to use during the posttest. After each task on the posttest, participants rated their mental effort. Lastly, participants completed the task-selection questionnaire.

Data analysis

What tasks do technical higher education's students select and why?

To shed light on participants' task-selection behavior, we counted the task format (video modeling example, worked example, practice problem) and complexity level (1, 2, or 3) of the six learning tasks each participant had selected and converted these scores into percentages. Then, we counted the task formats and complexity levels participants *said* they selected on the first and second learning task, and the task format participants *said* they selected most often during the entire learning phase. We used Chi-Square Tests to analyze whether there was a significant relation between task format or complexity level and the order of the learning tasks.

To evaluate participants' answers on the task-selection questionnaire, we coded their explanations (open coding) and grouped these codes into categories (axial coding). Two coders scored about 20% of the data and the interrater reliability of their scores was assessed by calculating Cohen's Kappa (Cohen, 1960). A Kappa value of 0 would mean no

agreement, values between 0.01 and 0.20 slight agreement, values between 0.21 and 0.40 fair agreement, values between 0.41 and 0.60 moderate agreement, values between 0.61 and 0.80 substantial agreement, and values between 0.81 and 1.00 almost perfect agreement (Landis & Koch, 1977). The agreement between the coders was moderate to almost perfect: Cohen's Kappa was 0.65 for question 1, 0.82 for question 2, 0.59 for question 3, 0.84 for question 4, and 0.95 for question 5.

How do novices' task selections match with instructional design principles?

We scored for each participant whether their task-selection behavior matched with the instructional design principles (i.e., example-based-learning-principle, example-first-principle, simple-to-complex-principle, lowest-level-first-principle, and start-each-complexity-with-example-principle). For each of these principles, participants could earn 1 point in total. More specifically, for the example-based-learning-principle, simple-to-complex-principle, and start-each-complexity-with-example-principle, 1 point was awarded when students' choices matched the principle entirely, 0.5 points when their choices matched the principle only partially, and 0 points when their choices did not match the principle at all. For the example-first-principle and lowest-level-first-principle, 1 point was awarded when students' choices matched the principle entirely and 0 points when their choices did not meet the principle at all. For each participant, a total score was computed (maximum: 5 points). For an extended version of the scoring protocol and an example, see Online Appendix F. Two coders scored about 20% of the data and the interrater reliability of their scores was assessed by calculating a two-way mixed, consistency, single-measures intraclass correlation (ICC; McGraw & Wong, 1996). According to Cicchetti (1994), ICC values that are below 0.40 are classified as poor, values between 0.40 and 0.59 are classified as fair, values between 0.60 and 0.74 are classified as good, and values between 0.75 and 1.0 are classified as excellent. With a score of 0.96, the ICC was in the good and excellent range for the principles.

Is there a positive relation between the extent to which students choices match with the instructional design principles and their scores on learning outcomes, mental effort, and motivational variables?

Lastly, we explored the extent to which students choices match with the instructional design principles correlated with cognitive (i.e., performance on the isomorphic posttest tasks, procedural transfer task, and conceptual questions, and mental effort) or motivational aspects of learning (i.e., self-efficacy, perceived competence, and topic interest).

We computed averages for the perceived competence and topic interest measurements before and after the learning phase, as well as for the reported effort invested in the learning tasks and the isomorphic posttest tasks. Test performance was scored by the first author and the third author based on a scoring protocol that was developed in collaboration with higher education mathematics teachers by Van Harsel et al. (2019). On the conceptual pretest and conceptual posttest items, participants could earn a maximum of 9 points. One point could be earned for the first open-ended question (1 point for the correct answer, 0 points for an incorrect answer) and 2 points for the other open-ended questions. Participants were rewarded with the maximum of 2 points when they got the answer right and provided correct explanations. Only 1 point was awarded if the answer was correct, but the explanation was incorrect or missing, and 0 points were given when both the answer

and explanation were incorrect. On the isomorphic posttest items, a maximum of 8 points could be earned for each task (i.e., three tasks, max. score=24 points), with 2 points for calculating each step correctly: (1) the step size of each subinterval, (2) all x-values, (3) the function values for all x-values, and (4) using the correct formula for the area under the graph and providing the correct answer. In step two, three, and four, one point was granted if half or more of the solution steps were correct and zero points were granted if less than half of the solution steps were correct. The same scoring standard was used to score the procedural transfer task (i.e., max. score=8 points). Again, two coders scored about 20% of the data and the interrater reliability of their scores was assessed by calculating a two-way mixed, consistency, single-measures intra-class correlation (ICC; McGraw & Wong, 1996). According to Cicchetti (1994), all our ICCs were all in the excellent range, with a score of 0.77 for the conceptual pretest tasks, 0.95 for the conceptual posttest tasks, 0.99 for the isomorphic posttest tasks, and 0.91 for the procedural transfer task.

Results

To answer our research questions on what tasks students selected and why (i.e., question 1) and how well their behavior matched with evidence-based principles from instructional design research (i.e., question 2), we report descriptive statistics. Regarding the correlational analyses (i.e., question 3), the effect size of Pearson r correlation is reported with values of 0.10, 0.30, and 0.50 representing a small, medium, and large effect size, respectively (Cohen, 1988). We must note, though, that we used an uncorrected significance level ($p < .05$) for the correlational analyses reported in this paper and that significant findings should be regarded with caution as we could not control the false-positive rate in the present study.

What tasks do technical higher education's students select and why?

Participants' task-selection behavior during the learning phase was explored. The percentages of selected formats are presented in Fig. 3, and the percentages of selected complexity levels in Fig. 4. The results of the Chi-Squared Tests are presented in the text.
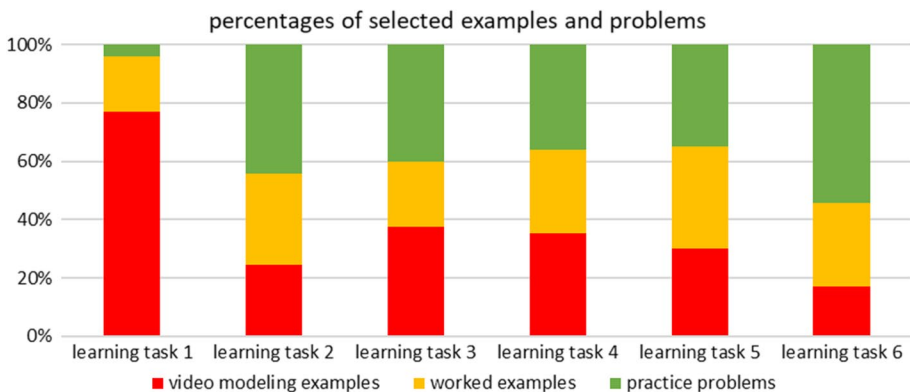

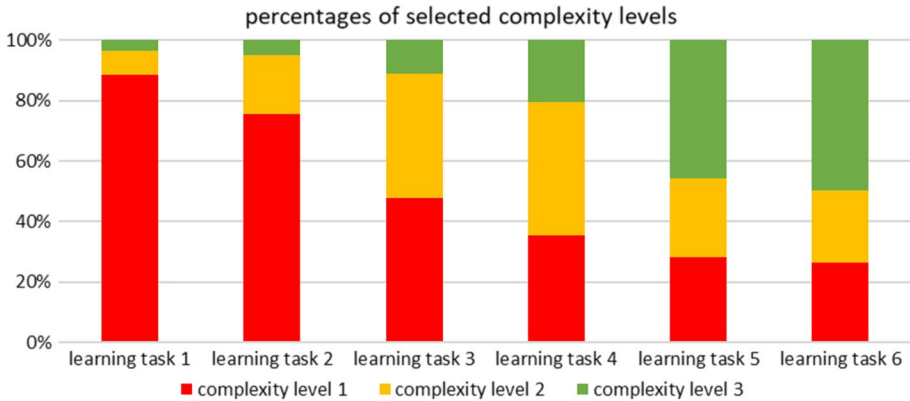**Fig. 3** Percentages of selected examples and problems in the self-regulated learning phase

Fig. 4 Percentages of selected complexity levels in the self-regulated learning phase

Examples and problems

On average, participants selected more examples to study (64.3%) than practice problems to solve (35.7%). The large majority of participants started the learning phase with an example instead of a practice problem. However, there was a strong decrease in the number examples selected from task 1 to task 2, and a strong increase in the selection of practice problems. Surprisingly, example study remained the preferred task format on task 2, 3, 4, and 5 (55% or higher). Only on the last learning task (i.e., task 6) did more participants select a practice problem than an example. Chi-Squared Tests revealed that the proportion selected examples depends on the order of the tasks, χ^2 (5) = 92.48, p < .001. This suggests that students selected fewer examples (and more problems) as the learning phase progressed.

Example format

Results showed that participants, on average, selected more video modeling examples (36.9%) than worked examples (25.8%) during the learning phase. The majority of the participants selected a video example as the first learning task. However, the percentage of selected video modeling examples dropped considerably on the second learning task. This percentage remained relatively stable up to and including the fifth learning task, but decreased further on the last learning task. The percentage of selected worked examples increased from the first to the second learning task and stayed relatively constant during the rest of the learning phase. Chi-Squared Tests revealed that the proportion selected video modeling examples depends on the order of the tasks χ^2 (5) = 52.48, p < .001, meaning that students selected fewer video modeling examples (and more worked examples) as the learning phase progressed.

Complexity level

The results showed that the many participants started the learning phase with a task at the lowest complexity level. The complexity of the selected tasks seemed to increase as the

learning phase progressed. That is, the percentage of level 1 tasks was highest on the first and second learning task, but declined from the third learning task onwards. The percentage of selected level 2 learning tasks, on the other hand, was relatively low on the first and second learning task, was highest on the third and fourth learning task and declined again on the fifth and sixth learning task. Level 3 tasks were selected seldomly during the first half of the learning phase, and were selected most often on the last two learning tasks. Surprisingly, results also showed that during the second half of the learning phase (i.e., learning task 4, 5, and 6), almost one third of the total sample still selected tasks at the lowest complexity level. Chi-Squared Tests revealed that the proportion selected lowest level tasks (i.e., level 1) depends on the order of the tasks $\chi^2 (10) = 285.92$, $p < .001$, suggesting that students selected fewer level 1 tasks (and more level 2 or level 3 tasks) as the learning phase progressed.

Reasons for task selections

We also analyzed participants' answers to the questions that asked them which tasks they selected and why. As shown in Tables 1 and 2, participants reported that they predominantly started the learning phase with a video modeling example, because this format was most comfortable for them or provided the most support. The reason why it was common to start with tasks of the lowest complexity level is that participants believed that this could help them build up their level of expertise or that this suited their current level of expertise. Note that these were also the most common reasons why participants chose the lowest complexity level as a second learning task. Regarding the format of the second learning task, participants selected practice problems most often, followed by worked examples. They mentioned that these formats helped them to assess their level of expertise or provided the most support (this reason was especially mentioned by those who selected worked examples). Finally, video modeling examples (followed by practice problems) were preferred most on average during the learning phase. Student often said it was the most comfortable way of learning (especially for video modeling examples because this format was most familiar, suited their learning preference, was most clear, etc.) or said they learned most from these format (especially for practice problems because participants felt practice helped them master the procedure). Note that participants' memory regarding what format and level they selected for the first task matched their actual choice, but for the second task participants only correctly remembered the task format and not the complexity level.

Do novices' task selections match with the instructional design principles?

Thirdly, we analyzed how well students' behavior matched the principles known to be effective and efficient based on instructional design research. Results showed that students' choices matched with many of the principles when selecting their own learning tasks. As shown in Table 4, the majority of the students (80.9%) had a total score of 4 or higher (out of 5), which means that their choices matched with (almost all of) the principles. When exploring how well participants' task selections matched with the individual principles (see Table 3), results showed that most of participants' choices matched with the example-based-learning-principle, example-first-principle, and lowest-level-first-principle. Moreover, the majority of the students started each complexity level with an example and another 21.1% did this only partially. Finally, only half of participants' choices aligned with the

Table 1 Students' answers to the open questions regarding the selection of different formats in the self-regulated learning phase

	What was the format of the first task you chose and why?	What was the format of the second task you chose and why?	Which task format did you choose most often and why?
Format			
Video modeling example	75.5%	24.5%	40.0%
Worked example	17.7%	35.4%	15.7%
Practice problem	2.0%	38.1%	29.3%
All formats equally often	–	–	19.0%
No answer	4.8%	2.0%	2.0%
Reason			
This way of learning is (most) comfortable	34.7%	12.9%	20.4%
This way of learning is (most) effective	2.7%	2.0%	25.9%
This way of learning is (most) supportive	36.7%	17.0%	10.2%
This way, I can assess my level of expertise	1.4%	36.1%	4.8%
This way, I can build up the level of complexity	–	1.4%	–
It fits my current level of expertise	–	2.0%	1.4%
It fits the order in the task database	2.7%	4.1%	–
Unclear reason	9.6%	10.2%	16.3%
No reason	12.2%	14.3%	21.1%

Table 2 Students' answers to the open questions regarding the selection of complexity levels in the self-regulated learning phase

	What was the level of the first task you chose and why?	What was the level of the second task you chose and why?
Complexity level		
Level 1	87.1%	57.1%
Level 2	9.5%	30.7%
Level 3	2.7%	10.2%
No answer	0.7%	2.0%
Reason		
This way of learning is (most) comfortable	–	–
This way of learning is (most) effective	–	–
This way of learning is (most) supportive	0.7%	1.4%
This way, I can assess my level of expertise	9.5%	18.4%
This way, I can build up the level of complexity	33.3%	21.7%
It fits my current level of expertise	35.4%	21.8%
It fits the order in the task database	1.4%	2.7%
Unclear reason	4.7%	12.2%
No reason	15.0%	21.8%

Table 3 Percentages of principles from example-based learning research applied in the self-regulated learning phase

	Total sample		
	Fully (%)	Partially	Not at all (%)
Example-based learning principle	89.8	9.5%	0.7
Example first principle	95.9	X	4.1
Lowest level first principle	88.4	X	11.6
Simple-to-complex principle	49.0	26.5%	24.5
Start each level with example principle	76.9	21.1%	2.0

X = not a scoring option for this principle

simple-to-complex-principle entirely, and more than a quarter of participants' choices aligned with this principle only partially (Table 4).

Is there a positive relation between the extent to which students choices match with the instructional design principles and their scores on learning outcomes, mental effort, and motivational variables?

Participants' scores on the cognitive and motivational aspects of learning are presented in Table 5. We explored how whether the degree of spontaneously applying the instructional design principles correlated with cognitive or motivational aspects of learning.

Table 4 Percentage of total score principles applied in the self-regulated learning phase

Total sample	All principles were applied Score 5.0	Score 4.50	Score 4.0	Score 3.5	Score 3.0	Score 2.5	Score 2.0	Score 1.5	Score 1.0	Score 0.5	No principles were applied Score 0.0
Total score	39.4%	19.7%	21.8%	7.5%	4.8%	3.4%	2.7%	0.7%	0%	0%	0%

Maximum total score = 5.0

Table 5 Mean (M), standard
deviation (SD), and median
(Med) of conceptual questions
(range 0 to 9), isomorphic tasks
(range 0 to 24), procedural
transfer task (range 0 to 8),
mental effort (range 1 to 9), self-
efficacy (range 1 to 9), perceived
competence (range 1 to 7), and
topic interest (range 1 to 7)

	Total sample		
	M	SD	Med
Pretest			
Conceptual questions	1.71	1.40	2.00
Self-efficacy	2.35	1.70	2.00
Perceived competence	2.04	1.41	1.33
Topic interest	4.26	0.096	4.43
Learning phase			
Mental effort	4.03	1.32	4.00
Self-efficacy	5.76	1.43	6.00
Posttest			
Conceptual questions	3.05	1.94	3.00
Isomorphic tasks	8.07	6.30	8.00
Procedural transfer task	2.07	2.89	0.00
Mental effort			
Conceptual questions	4.44	1.89	5.00
Mental effort			
Isomorphic tasks	5.56	1.99	6.00
Mental effort			
Procedural transfer task	6.36	2.40	7.00
Self-efficacy	6.12	1.65	6.00
Perceived competence	5.08	1.27	5.33
Topic interest	3.98	1.02	4.14

As shown in Table 6, total scores of how well students task selections matched with the principles did not correlate with any of the cognitive or motivational variables. However, some of the individual principles did correlate with some of the cognitive or motivational variables. Firstly, there was a positive relation between spontaneously applying the example-first-principle and average scores on self-efficacy in the learning phase ($r = .183$). Secondly, spontaneously applying the lowest-level-first-principle negatively correlated with average scores of self-efficacy in the learning phase ($r = -.257$) and self-efficacy ($r = -.268$) and perceived competence ($r = -.219$) after the learning phase. The − lowest-level-first-principle also negatively correlated with the scores on the procedural transfer task ($r = -.235$). A positive correlation was shown, however, between the lowest-level-first-principle and average ratings of mental effort invested in the conceptual questions in the posttest ($r = .232$). Finally, spontaneously applying the simple-to-complex-principle positively correlated with average scores of mental effort invested in the learning ($r = -.180$) and isomorphic posttest tasks ($r = -.183$). We must note, though, that the strength of these correlations can be referred to as small (because the absolute values of r are below .30; Cohen, 1988) and that significant findings should be regarded with caution as we could not control the false-positive rate in the present study.

Table 6 Correlation coefficients (r) from Pearson correlation analysis between principles from example-based learning and different cognitive and motivational variables

	Example-based learning principle	Example first principle	Lowest level first principle	Simple-to-complex principle	Start each level with example principle	Total score principles
	r	r	r	r	r	r
Learning phase						
Mental effort	−0.002	−0.069	0.147	−0.180*	−0.114	−0.084
Self-efficacy	−0.045	0.183*	−0.257*	0.136	0.072	0.025
Posttest						
Conceptual questions	0.094	0.148	0.032	0.022	0.037	0.090
Isomorphic tasks	0.004	0.052	−0.064	0.067	0.019	0.028
Procedural transfer task	0.037	0.123	−0.235*	0.117	0.165	0.050
Mental effort						
Conceptual questions	0.023	−0.134	0.232*	0.031	−0.020	0.073
Isomorphic tasks	0.085	−0.046	0.134	−0.183*	0.000	−0.033
Procedural transfer task	0.101	0.017	0.161	0.004	−0.010	0.087
Self-efficacy	−0.085	0.038	−0.268*	0.081	0.000	−0.073
Perceived competence	−0.056	0.076	−0.219*	0.115	0.036	−0.010
Topic interest	−0.086	−0.004	0.089	0.149	0.084	0.117

*$p < .05$. We used an uncorrected significance level ($p < .05$) for the correlational analyses reported in this paper and that significant findings should be regarded with caution as we could not control the false-positive rate in the present study

Discussion

The aim of this study was to explore the task-selection choices of first year higher education students (i.e., novices to the learning materials) when engaging in self-regulated learning in a computer-based learning environment. Students had to learn how to solve problems using the trapezoid rule and could select learning tasks from a database comprising different task formats (i.e., video modeling examples, worked examples, and practice problems), levels of complexity (i.e., three levels), and cover stories. We were particularly interested in which tasks students would choose and why (Research Question 1) and how students' task-selection decisions would adhere to the robust principles from instructional design research (Research Question 2). Finally, we were interested in whether there is a positive relation between the extent to which students choices match with the instructional design principles and their scores on learning outcomes, mental effort, and motivational variables (Research Question 3).

With regard to Research Question 1, results showed that the selection of video modeling examples significantly decreased and the selection of worked examples and problems significantly increased during the learning phase. In addition, the selection of lowest level tasks significantly decreased, whereas the selection of level 2 and level 3 tasks increased. Also, findings showed that students' choices matched quite well with the principles derived from instructional design research on the effectiveness and efficiency of different fixed sequences of examples and problems (Research Question 2). The vast majority of students selected many examples during the learning phase (i.e., example-based-learning-principle) and started the learning phase with an example instead of a problem (i.e., example-first-principle). Although the choices of approximately half our sample aligned with the simple-to-complex principle, almost all participants started the learning phase with a task at the lowest complexity level (i.e., lowest-level-principle). Moreover, most participants started each complexity level with example study (i.e., start-each-level-with-example-principle). There was a lack of correlations between spontaneously applying the instructional design principles and cognitive or motivational aspects of learning (Research Question 3).

That students spontaneously applied almost all of the instructional design principles (with the exception of the simple-to-complex-principle) is surprising. Although there is relatively little research on this issue, the available evidence suggested that novices underutilize example study with respect to the amount (i.e., about 40% worked examples versus 60% practice problems) and timing (i.e., students rarely started the learning phase with example study) of their use (e.g., Foster et al., 2018).

There are several possible reasons for why students' choices were so well aligned with the instructional design principles in our study compared to the study of Foster et al. (2018). Firstly, our sample consisted of technical higher education students instead of a mixed group of students obtained from the university's participant pool (as in the study of Foster et al., 2018). In the study programs of our sample, mathematics is an important subject and as a result, students might have already had experience with learning new mathematical problem-solving skills with the help of examples (since examples are frequently used to learn new mathematical procedures; e.g., Hoogerheide & Roelle, 2020). Moreover, this prior experience might have helped students decide how much support they needed or what complexity level they should work on (as individuals with extensive prior knowledge are better able to identify their knowledge needs and make their task-selections accordingly; e.g., Corbalan et al., 2006).

Secondly, video modeling examples might be more preferred at the beginning of a learning phase (compared to worked examples), because information is demonstrated in a step-by-step manner and the combination of dynamic visual information and the model's narration takes the learner by the hand. In contrast, worked examples might be preferred later in the learning phase, because they allow for efficiently looking up difficult problem-solving steps. Indeed, our findings revealed that almost all students started with a video modeling example. However, this number rapidly decreased while the selection of worked examples gradually increased. These results correspond with the results of Hoogerheide et al. (2014) that worked examples were also more preferred than video modeling examples when more tasks had to be studied. In addition, students said in their answers to the open questions that they selected worked examples as a second task mostly because this format provided them the opportunity to assess their level of expertise and easily check to what extent they had understood the procedure.

Lastly, a more likely and more practical explanation for why our sample relied so heavily on example study is that we provided the opportunity to choose between two example formats (i.e., video modeling examples and worked examples) next to practice problems. As a result, two thirds of the learning tasks were examples (i.e., 67%) and only one third were practice problems (i.e., 33%). This might have increased the likelihood of selecting an example rather than a practice problem. In the study of Foster et al. (2018), students could only choose between worked examples next to (completion) problems.

These possible explanations provide several interesting avenues for future research in self-regulated learning settings. For instance, it would be interesting to investigate in further detail whether or not familiarity with example study in one domain would affect the degree to which novices opt for example study relative to practice problem solving in the same or in a different domain. Moreover, as comparisons between (different sequences of) worked and video modeling examples are scarce (e.g., Hefter et al., 2019; Hoogerheide et al., 2014), future research could investigate whether starting the learning phase with a video modeling example and switching to worked examples is not only a more preferred way of learning but also more effective, efficient, and motivating than the other way around. Finally, although our sample already relied heavily on example study, it might be interesting to investigate whether the likelihood of selecting an example instead of a practice problem would further increase in this student population when using male models. That is, the model-observer similarity (MOS) hypothesis suggests that male learners might prefer and learn more from a male model, because learners who perceive themselves to be more similar to the model would show greater self-efficacy and learning gains. We must note though, that there is little evidence for the model-observer similarity hypothesis in recent research with video modeling examples: when content is kept equal, the gender of the model does not seem to affect test performance, self-efficacy, or perceived competence (e.g., Hoogerheide et al., 2016).

Limitations

This study also has some limitations. First, because performance on the practice problems was not logged, it was not possible to examine the degree to which students' task-selections were adaptive to their needs. The optimal sequence (length) of examples and problems differs for each individual learner because there is variance in the speed to which students learn (e.g., due to differences in cognitive abilities). If we had access to practice problem solving performance, we could score whether students made accurate decisions following

a practice problem. For instance, students who just failed to solve a problem should ideally select an example or another practice problem of the same complexity level, while students should select a more complex task after successfully solving a practice problem. That task-selections should ideally be tailored to individual progress could also explain the lack of correlations between the extent to which students' task selections matched with the instructional design principles and learning outcomes (Research Question 3). Moreover, had we successfully logged performance on the practice problems, we could determine whether students' knowledge during the learning phase was so high that they would benefit more from problem solving than example study (cf. the expertise reversal effect; Kalyuga et al., 2003). However, it is unlikely that this expertise-reversal effect occurred, because performance on the isomorphic posttest was not that high and students were only allowed to select six learning tasks while having to learn three different complexity levels.

A second limitation of the present study concerns the measurement of self-efficacy and perceived competence. There is research showing overlap between these two constructs, and more specifically that perceived competence may be a common core component of both self-efficacy and self-concept (e.g., Marsh et al., 2019; Schunk & Pajares, 2005). The results of our study confirm this idea, as correlational analyses of these two constructs measured after the self-regulated learning phase revealed a score of $r = .86$. One could wonder to what extent both measures differ or measure the same general feeling of competence regarding to what has been learned and how well someone considers him/herself capable of solving a similar task. Therefore, it might be sufficient for future research in this area to use of one of the questionnaires.

Conclusions and practical implications

In sum, our exploratory study showed that novices' task-selection patterns corresponded fairly well with principles derived from instructional design research. This seems promising, because it would mean that students know quite well how to use examples and problems (at different complexity levels) when learning new problem-solving skills and therefore might need little support. However, given the paucity of research on self-regulated learning of examples and problems (at different levels of complexity), the mixed findings regarding the use of examples and problems (i.e., Foster et al., 2018 vs. our study), and the open question of whether students' task selections are adapted to their levels of expertise, we cannot say this with absolute certainty. Moreover, regarding the task selections (of some of the students) and test performance scores in our study and the study of Foster et al. (2018), there seems to be (some) room for improvement in how novices regulate their learning from examples and problems. Therefore, more research is needed to gain insight in how and how well (novice) learners regulate their learning from examples and practice problems, and whether and how they can benefit from support. Moreover, future research should investigate to what extent the findings of this study regarding students' task selections are problem-specific or generalizable, for example by using similar procedures but different problem-solving tasks.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Informed consent All participants gave their informed consent in the online learning environment.

References

Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development, 56*, 45–72. https://doi.org/10.1007/s11423-007-9067-0

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). MIT Press.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284. https://doi.org/10.1037/1040-3590.6.4.284

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). 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

Corbalan, G., Kester, L., & Van Merriënboer, J. J. (2006). Towards a personalized task selection model with shared instructional control. *Instructional Science, 34*(5), 399–422. https://doi.org/10.1007/s11251-005-5774-2

Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2008). Selecting learning tasks: Effects of adaptation and shared control on learning efficiency and task involvement. *Contemporary Educational Psychology, 33*(4), 733–756. https://doi.org/10.1016/j.cedpsych.2008.02.003

De Bruin, A. B. H., & Van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction, 22*, 245–252. https://doi.org/10.1016/j.learninstruc.2012.01.003

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*(3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x

Foster, N. L., Rawson, K. A., & Dunlosky, J. (2018). Self-regulated learning of principle based concepts: Do students prefer worked examples, faded examples, or problem solving? *Learning and Instruction, 55*, 124–138. https://doi.org/10.1016/j.learninstruc.2017.10.002

Hefter, M. H., ten Hagen, I., Krense, C., Berthold, K., & Renkl, A. (2019). Effective and efficient acquisition of argumentation knowledge by self-explaining examples: Videos, texts, or graphic novels? *Journal of Educational Psychology, 111*(8), 1396–1405. https://doi.org/10.1037/edu0000350

Hoogerheide, V., Loyens, S. M., & Van Gog, T. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior, 41*, 80–91. https://doi.org/10.1016/j.chb.2014.09.013

Hoogerheide, V., & Roelle, J. (2020). Example-based learning: New theoretical perspectives and use-inspired advances to a contemporary instructional approach. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.3706

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

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

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

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology, 31*, 187–194. https://doi.org/10.1037/0278-7393.31.2.187

Kostons, D., Van Gog, T., & Paas, F. (2010). Self-assessment and task selection in learner controlled instruction: Differences between effective and ineffective learners. *Computers & Education, 54*(4), 932–940. https://doi.org/10.1016/j.compedu.2009.09.025

Kostons, D., Van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction, 22*, 121–132. https://doi.org/10.1016/j.learninstruc.2011.08.004

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

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

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–173). Academic Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.

Niemiec, R. P., Sikorski, C., & Walberg, H. J. (1996). Learner-control effects: A review of reviews and a meta-analysis. *Journal of Educational Computing Research, 15*(2), 157–174. https://doi.org/10.2190/JV1U-EQ5P-X2PB-PDBA

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

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, 422. https://doi.org/10.3389/fpsyg.2017.00422

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*(1), 61–86. https://doi.org/10.1016/S0959-4752(01)00016-0

Quilici, J. L., & Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology, 16*(3), 325–342. https://doi.org/10.1002/acp.796

Rawson, K. A., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology, 19*(4–5), 559–579. https://doi.org/10.1080/09541440701326022

Renkl, A. (2014). Towards an instructionally-oriented theory of example-based learning. *Cognitive Science, 38*, 1–37. https://doi.org/10.1111/cogs.12086

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*(1), 15–22. https://doi.org/10.1207/S15326985EP3801_3

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21*(2), 267–280. https://doi.org/10.1016/j.learninstruc.2010.07.004

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

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

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185–233. https://doi.org/10.1207/s1532690xci1203_1

Van Gog, T., Hoogerheide, V., & Van Harsel, M. (2020). The role of mental effort in fostering self-regulated learning with problem-solving tasks. *Educational Psychology Review, 32*, 1055–1072. https://doi.org/10.1007/s10648-020-09544-y

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., 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). Cambridge University Press.

Van Harsel, M., Hoogerheide, V., Verkoeijen, P. P. J. L., & 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

Van Harsel, M., Hoogerheide, V., Verkoeijen, P. P. J. L., & Van Gog, T. (2020). Examples, practice problems, or both? Effects on motivation and learning in shorter and longer sequences. *Applied Cognitive Psychology, 34*(4), 793–812. https://doi.org/10.1002/acp.3649

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Educational Technology Publications.

Van Merriënboer, J. J. G., & Kirschner, P. A. (2013). *Ten steps to complex learning: A systematic approach to four-component instructional design* (2nd ed.). Taylor & Francis.

Van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*(1), 5–13. https://doi.org/10.1207/S15326985EP3801_2

Van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance, and training efficiency. *Learning and Instruction, 12*(1), 11–37. https://doi.org/10.1016/S0959-4752(01)00020-2

Van Merriënboer, J. J. G., Sluijsmans, D. M. A., Corbalan, G., Kalyuga, S., Paas, F., & Tattersall, C. (2006). Performance assessment and learning task selection in environments for complex learning. In J. Elen & R. E. Clark (Eds.), *Handling complexity in learning environments: Theory and research (Advances in learning and instruction series)* (pp. 201–220). Elsevier.

Van Zundert, M. J., Könings, K. D., Sluijsmans, D. M. A., & Van Merriënboer, J. J. G. (2012). Teaching domain-specific skills before peer assessment skills is superior to teaching them simultaneously. *Educational Studies, 38*(5), 541–557. https://doi.org/10.1080/03055698.2012.654920

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

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 279–306). Erlbaum.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*, 3–17. https://doi.org/10.1207/s15326985ep2501_2