CrossMark

Training self-regulated learning skills with video modeling examples: Do task-selection skills transfer?

Steven F. Raaijmakers[1,2] · Martine Baars[2] · Lydia Schaap[2,3] ·
Fred Paas[2,4] · Jeroen van Merriënboer[5] · Tamara van Gog[1,2]

Abstract Self-assessment and task-selection skills are crucial in self-regulated learning situations in which students can choose their own tasks. Prior research suggested that training with video modeling examples, in which another person (the model) demonstrates and explains the cyclical process of problem-solving task performance, self-assessment, and task-selection, is effective for improving adolescents' problem-solving posttest performance after self-regulated learning. In these examples, the models used a specific task-selection algorithm in which perceived mental effort and self-assessed performance scores were combined to determine the complexity and support level of the next task, selected from a task database. In the present study we aimed to replicate prior findings and to investigate whether transfer of task-selection skills would be facilitated even more by a more general, heuristic task-selection training than the task-specific algorithm. Transfer of task-selection skills was assessed by having students select a new task in another domain for a fictitious peer student. Results showed that both heuristic and algorithmic training of self-assessment and task-selection skills improved problem-solving posttest performance after a self-regulated learning phase, as well as transfer of task-selection skills. Heuristic training was not more effective for transfer than algorithmic training. These findings show that example-based self-assessment and task-selection training can be an effective and

---

✉ Steven F. Raaijmakers
s.f.raaijmakers@uu.nl

1   Department of Education, Utrecht University, P.O. Box 80140, 3584 CS Utrecht, The Netherlands

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

3   Learning and Innovation Centre, Avans University of Applied Sciences, Breda, The Netherlands

4   Early Start Research Institute, University of Wollongong, Wollongong, Australia

5   Department of Educational Development and Research, Maastricht University, Maastricht, The Netherlands

relatively easy to implement method for improving students' self-regulated learning outcomes. Importantly, our data suggest that the effect on task-selection skills may transfer beyond the trained tasks, although future research should establish whether this also applies when trained students perform novel tasks themselves.

Keywords Self-regulated learning · Problem solving · Self-assessment · Task selection · Example-based learning

Fostering students' ability to engage in effective self-regulated learning is an important goal of secondary education, as it prepares students for the demands of higher education or workplace learning. Self-regulated learning abilities are also associated with better academic achievement across childhood and adolescence (Dent and Koenka 2016). Meta-analyses of self-regulation training programs show that self-regulated learning skills can be improved through training programs in both primary and secondary education (Dignath and Büttner 2008). Two pivotal processes of self-regulated learning are monitoring and control, which figure prominently in most models of self-regulated learning (Nelson and Narens 1990; Pintrich and De Groot 1990; Winne and Hadwin 1998; Zimmerman and Schunk 2001). Monitoring can be defined as evaluating one's own cognitive processes concurrently (e.g., while performing a task), retrospectively (e.g., self-assessing performance on a task that was just completed), or prospectively (e.g., estimating one's own performance on a similar task in the future; Bjork et al. 2013). This evaluation can serve as input for control processes, that is, the regulation of current or subsequent study behavior (Zimmerman and Schunk 2001). By focusing on monitoring and control, instructional strategies to foster self-regulated learning could be made more effective (Schraw 1998).

Monitoring and controlling learning processes

Monitoring and control have been studied at different levels. At the *item* level, learners typically monitor how well they have memorized a word-pair or comprehended a text by predicting their performance on a future test, and decide whether they would have to restudy that item or text (e.g., Thiede et al. 2003). At the *task* or *topic* level, learners typically monitor their understanding while they are engaged in a study task, for instance, text passages presented in a hypermedia learning environment, and decide which parts to select for study and how long to spend on them (e.g., Azevedo and Cromley 2004). Finally, at the *task-sequence* level, learners typically monitor how well they performed a learning task after completing it (referred to as 'self-assessment' here, to distinguish it from monitoring during task performance), and then select a suitable next task (e.g., Corbalan et al. 2008).[1]

Research on each of those levels of granularity has shown that learners' monitoring and control are often inaccurate (item: Rawson and Dunlosky 2007; task/topic: Bannert and Reimann 2012; task-sequence: Kostons et al. 2010, 2012). This is problematic, because both monitoring and control need to be accurate for effective self-regulated learning. Monitoring accuracy can be regarded as a necessary but not sufficient condition for accurate control. That is, without accurate monitoring, it is unlikely that students would select subsequent tasks that are suitable for their level of knowledge or skill (see e.g., Dunlosky and Rawson 2012; Thiede et al. 2003). However, improving monitoring accuracy would only aid control when students actually consider their monitoring judgments when making study decisions. Moreover, accurate control requires other factors to be considered as well (e.g., the learning goals,

---

[1] Note that some authors refer to this as self-*directed* learning (Loyens et al. 2008).

complexity of the task, et cetera). The accuracy of monitoring and control has been associated with better item recall, text comprehension, and problem-solving performance (item: Rawson et al. 2011; task/topic: Panadero and Romero 2014; task-sequence: Kostons et al. 2010, 2012). These skills tend to develop over time with 12 year-olds demonstrating better monitoring and control than 9 year-olds (Roebers et al. 2009).

Researchers in educational and (applied) cognitive psychology have been concerned with finding means to support or scaffold monitoring and control *during* self-regulated learning (e.g., Azevedo and Hadwin 2005; Bannert 2006; Dabbagh and Kitsantas 2005; Kramarski and Gutman 2006; Winne et al. 2006), as well as to train monitoring and control *prior to* self-regulated learning (e.g., Azevedo and Cromley 2004; Costa Ferreira et al. 2015; Kostons et al., 2012; Leidinger & Perels, 2012; Perels et al. 2005) in order to enhance students' learning.

The present study is concerned with improving secondary education students' self-regulated learning at the task-sequence level. As higher education requires adequate self-regulated learning skills and because these skills are fully in development during secondary education (i.e., during adolescence), it is important to investigate whether secondary education students benefit from such interventions. This study's aim is to replicate and extend prior research that suggested that secondary education students' self-assessment (monitoring) and task-selection (control) skills can be trained by means of video modeling examples (Kostons et al. 2012).

Training self-assessment and task selection

Kostons and colleagues trained self-assessment and task selection in the domain of biology with video modeling examples based on principles from both social learning theory (Bandura 1977) and example-based learning (Renkl 2014; Van Gog and Rummel 2010). In these examples, another person (the model) demonstrated and explained the cyclical process of task performance, self-assessment, and task selection. Self-assessment was trained by modeling how to assign one point for each correctly solved step of the problem-solving task and task selection was trained by modeling how to use an algorithm that combines subjective ratings of effort and performance into an advice for selecting a next task from a task database that contained tasks at different levels of complexity and with different levels of support (cf. Corbalan et al. 2008; Salden et al. 2006).

In their first experiment, Kostons et al. (2012) compared the effectiveness of video modeling examples that provided self-assessment training, task-selection training, or both, to a condition that received no self-assessment or task-selection training (but students did see the model's performance on the problem-solving task). Results showed that the self-assessment training led to more accurate self-assessment, but not to more accurate task selection, whereas the task-selection training led to more accurate task selection, but not to more accurate self-assessment.[2] Therefore, Kostons et al. (2012) concluded that both aspects should be explicitly modeled in the training. In the second experiment, participants received self-assessment and task-selection training, after which they engaged in a self-regulated

---

[2] The finding that task selection became more accurate while there was no improvement in self-assessment accuracy seems counterintuitive, but is presumably related to the way in which task-selection accuracy was computed, namely by using the self-assessed performance rather than the actual performance. This allowed for checking whether the task-selection procedure had been learned, independent of how accurate self-assessment was.

learning phase in which they could select and perform eight problem-solving tasks from a database with tasks at five different complexity levels and three levels of support (i.e., high support: 4/5 steps already worked out, low support: 2/5 steps worked out, or no support: no steps worked out) at each complexity level. Problem-solving performance on a posttest after the self-regulated learning phase was higher for trained participants compared to a control group, indicating that training self-assessment and task selection skills indeed enhanced the effectiveness of self-regulated learning. An important open question, however, is whether the trained skills would transfer beyond the trained tasks.

Transfer of trained self-assessment and task-selection skills

Transfer is one of the key goals of scaffolding or training self-regulated learning skills, but we still know relatively little about the re-use of those scaffolded/trained skills in new environments and domains (Koedinger et al. 2009; Roll et al. 2014). Although the degree to which self-regulated learning processes are domain-general has been discussed (Greene et al. 2015; Poitras and Lajoie, 2013), the issue has not been extensively addressed through empirical investigation (Alexander et al. 2011), and we are not aware of any studies that addressed it experimentally on the task-sequence level. If task-selection is a domain-general, higher-order skill (as suggested in Van Merriënboer 1997; Van Merriënboer and Kirschner 2018), then learners should—in theory—benefit from having acquired task-selection skills on tasks in one domain when they subsequently engage in self-regulated learning in another domain.

For example, would students know how to decide what a suitable next learning task would be in mathematics, when they have acquired task-selection skills in the context of biology problems? It is unlikely that transfer would happen spontaneously as students often experience difficulties transferring learned skills to different domains (i.e., far transfer; Donovan et al. 1999). One reason for these difficulties might be that successful transfer between domains would require that learners abstract the more general underlying principle or heuristic from the learning material and recognize similarities in the 'old' and 'new' contexts that warrant application of the principle (i.e., bridging; Salomon and Perkins 1989). For problem-solving task performance, transfer has been shown to improve when learners are able to (or helped to) abstract a general rule from the learning material, or are presented with that general rule (Kimball and Holyoak 2000).

Although findings regarding transfer of problem-solving rules might not necessarily apply to transfer of task-selection rules, similar mechanisms might be at work. The task-selection algorithm that Kostons et al. (2012) used to teach students task-selection skills was specifically tailored to their five-step biology problem-solving tasks. In order for transfer to another domain to occur, students would first have to realize that it is applicable to every task that consists of 5 steps, regardless of whether it is biology, math, or chemistry. Second, as the tasks in other domains are not likely to have the exact same number of problem-solving steps, learners would have to mentally transform the algorithm, which requires additional effort. A more general task-selection heuristic would eliminate the additional effort required for the transformation of the task-selection algorithm (Shah and Oppenheimer 2008). Therefore, analogous to the findings on problem-solving skills, we hypothesized that training students using this more general, heuristic, task-selection rule ("when performance is high and effort is low, choose a more complex task"), would lead to improved transfer of task-selection skills compared to training them using a specific

algorithm. As, the heuristic task-selection rule uses a relative indication of performance and effort (i.e., in terms of low, high, average) instead of mentioning a specific score or rating, transfer to different types of tasks should be easier. In essence, the heuristic task-selection rule explicitly teaches students the underlying principle of the algorithm (i.e., choosing a task that is neither too complex nor too easy for the student given the level of performance and invested mental effort on a previous task).

In sum, the first aim of the present study was to replicate the findings by Kostons et al. (2012) that training self-assessment and task-selection skills prior to a self-regulated learning phase, will lead to better problem-solving performance after the self-regulated learning phase. The second aim was to extend those prior findings by adding a heuristic task-selection training condition and a task-selection transfer task. This allowed us to investigate whether the heuristic training would also be effective for improving self-regulated learning outcomes, whether both types of training would lead to better transfer of task-selection skills than no training, and whether heuristic training would lead to better transfer of task-selection skills than algorithmic training.

The present study

To be able to test whether training self-assessment and task-selection skills prior to a self-regulated learning phase, will lead to better self-regulated learning outcomes (as measured by problem-solving performance on a posttest) and to better transfer of task-selection skills than no training, we used the following experimental design. First, students took a pretest, after which they received training consisting of observing modeling examples in which the models: (1) performed a genetics problem-solving task (control condition) (2) performed the problem-solving task, rated their perceived mental effort, assessed their performance, and applied an algorithm to select a next task (algorithmic training condition), or (3) performed the problem-solving task, rated their perceived mental effort, assessed their performance, and applied a heuristic to select a next task (heuristic training condition). Then, students engaged in self-regulated learning, being allowed to choose the problem-solving tasks they wanted to work on from a task database that contained tasks at different levels of complexity and different levels of support at each complexity level (cf. principles of the 4C/ID model, the four component instructional design model of Van Merriënboer 1997; see also Van Merriënboer and Kirschner 2018). After working on eight self-selected tasks from the database, students took a problem-solving posttest. They were also asked to rate perceived mental effort, assess their performance, and choose a next task (which they did not actually get, as the posttest was the same for all participants) so that we could explore differences in self-assessment and task-selection accuracy at posttest. This was followed by a transfer test to assess whether they could also apply task-selection skills to select appropriate new tasks for a fictitious student in the domain of math, when the number of steps in a task differed from the trained tasks or when the task database had a different layout.

Initially, we did not have a dedicated learning environment (similar to the one used by Kostons et al. 2012) at our disposal, and used Qualtrics survey software to deliver the problem-solving tasks, as this allowed for participants to choose tasks to work on. To ensure that the modeling examples were congruent with how the tasks appeared in this environment, we created screen-recordings of the model performing the tasks in the Qualtrics environment, in which the model read the text aloud, thought aloud, and clicked on or typed in an answer at each step. These modeling examples were rather different from

those of Kostons et al. (2012), in which the model was writing out each step of the problem-solving procedure while thinking aloud. Regrettably, this did not turn out to be a wise design choice, as the modeling examples turned out to be ineffective (see Experiment A and B in the supplementary materials); potentially because modeling examples are more effective at guiding and maintaining attention when information is drawn out by the model than when it is already present (cf. Fiorella and Mayer 2016). Indeed, redesigning the modeling examples to be more similar to the ones used by Kostons et al., made them more effective (see Experiment C in supplementary materials). The experiment reported here is the one we originally set out to perform, but now conducted with the effective video modeling examples and a dedicated learning environment.

It was hypothesized that self-assessment and task-selection training would result in an increase in posttest problem-solving performance after a self-regulated learning phase (i.e., algorithmic and heuristic training > no training; Hypothesis 1), as well as in better transfer of task-selection skills (i.e., algorithmic and heuristic training > no training; Hypothesis 2a), and that the heuristic condition would show better transfer than the algorithmic group (Hypothesis 2b).

Method

Participants and design

A total of 125 Dutch students in their second year of higher general secondary education (middle level of secondary education in the Netherlands with a 5 year duration) participated in this study. Three participants who did not manage to finish the experiment, were excluded from the data set, leaving 122 participants ($M_{age} = 13.65$ years, $SD_{age} = 0.64$; 64 boys and 58 girls). Participants had been randomly assigned to one of the three conditions (in brackets the numbers after exclusion): (1) the control condition ($n = 43$) in which the 'training' showed the models performing problem-solving tasks (2) the algorithmic training condition ($n = 42$), in which the models performed problem-solving tasks, rated invested mental effort, self-assessed their performance, and used an algorithm to select a next task, and (3) the heuristic training condition ($n = 37$), in which the models performed problem-solving tasks, rated invested mental effort, self-assessed their performance, and used a heuristic to select a next task. The study took place at a point in the biology curriculum (obligatory for all students) at which participants were assumed to have little if any prior knowledge about monohybrid crossing problems (the problem-solving tasks used in this experiment), and this was verified by their pretest performance.

Materials

The problem-solving tasks, algorithmic training, and control condition were similar to those used by Kostons et al. (2012). All materials were web-based, using a dedicated online learning environment designed for the present study.

Problem-solving tasks

The problem-solving tasks (adapted from Kostons et al. 2012) were monohybrid cross problems[3] in the domain of biology (Mendel's law of heredity). The procedure for solving these problems consisted of five steps: (1) translating the information given in the cover story into genotypes, (2) putting this information in a family tree, (3) determining the number of required Punnett squares, (4) filling in the Punnett square(s), and (5) finding the answer(s) in the Punnett square(s). An example of a task is shown in Appendix 1.

Task database

The task database (cf. Kostons et al. 2012) contained 75 problems at five levels of complexity, with three levels of support within each complexity level (see Fig. 1). The levels of complexity (top row of Fig. 1) increased by changing: the possibility of multiple answers (complexity level 2), the type of reasoning used to solve the problem (complexity level 3), the number of generations (complexity level 4), and the number of unknowns (complexity level 5). At each level of complexity, there were three different levels of support, depending on the amount of steps that were already worked-out for the learner (see the second row in Fig. 1). At the high support level, the first 4 steps were worked out and the learner had to complete the fifth step; at the low support level the first 2 steps were worked out and the learner had to complete steps 3 through 5; and at the no support level no steps were worked out and the learner had to complete all five steps. Note that this allowed learners to move from high levels of instructional guidance to low or no guidance within a level of complexity, in line with the completion strategy (Paas 1992; Van Merriënboer et al. 2002) or fading guidance strategy (Renkl and Atkinson 2003), which have proven to be effective for acquiring problem-solving skills. Allowing learners to move from lower complexity tasks to higher complexity tasks, with high instructional guidance fading to lower guidance at each level, is in line with the recommendations of the 4C/ID model (Van Merriënboer 1997; Van Merriënboer and Kirschner 2018). The combination of five levels of complexity and three levels of support created 15 columns in which the tasks are organized. In each column, five isomorphic tasks (one per row) were presented (resulting in a total of 75 tasks), which were structurally equivalent but had different surface features (i.e., cover stories).

Pretest

The pretest was administered to check whether students were indeed novices regarding the topic at hand. It consisted of three problem-solving tasks without support (one task at complexity level 1, one task at complexity level 2, and one task at complexity level 3, in this order). These tasks had the same structure as the tasks in the database, but contained different surface features.

---

[3] In a monohybrid cross problem you can work out what traits offspring might have of parents who only differ in one trait (i.e., monohybrid). By combining the alleles (different forms of a gene coding for a trait) you can find out what traits the offspring could possibly have.

Complexity level	Complexity 1: -2 generations -1 unknown -Single answer -Deductive			Complexity 2: -2 generations -1 unknown -Multiple answers -Deductive			Complexity 3: -2 generations -1 unknown -Multiple answers -Inductive			Complexity 4: -3 generations -2 unknown -Multiple answers -Both ways			Complexity 5: -3 generations -2 unknowns -Multiple answers -Both ways		
Support level	High support	Low support	No support	High support	Low support	No support	High support	Low support	No support	High support	Low support	No support	High support	Low support	No support
Learning tasks	Eye color	Eye color	Eye color	Eye Color	Eye color	Eye color	Eye color	Eye color	Eye color	Eye color	Eye color	Eye color	Eye color	Eye color	Eye color
	Hair structure	Hair structure	Hair structure	Hair structure	Hair structure	Hair structure	Hair structure	Hair structure	Hair structure	Hair structure	Hair structure	Hair structure	Hair structure	Hair structure	Hair structure
	Cat fur shape	Sickle cell	Hunting-ton	Flower color	Cat fur shap	Tongue curling	Fruit flies	Dog tail length	Freckles	Albinism	Fruit flies	Pea plant	Milk allergy	Dog tail length	Cystic fibrosis
	Apple tree	Chicken beak	Milk allergy	Widow's peak	Albinism	Apple tree	Chicken beak	Apple tree	Flower color	Cat fur shape	Tongue curling	Dimples	Depres-sion	Wolfram	Fruit flies
	Depres-sion	Guinea pigs	Cleft lip	P.R.A.	Pea plant	Fruit flies	Wolfram	Milk allergy	Earlobes	Fruit flies	Flower color	Depres-sion	Huntingt on	Flower color	Rat fur

Fig. 1 Task database containing the 75 problem-solving tasks showing the different levels of complexity, different levels of support, and the different surface features of the learning tasks

Training phase

The training consisted of an introductory video, in which the main concepts were explained (i.e., dominant/recessive, homozygous/heterozygous), followed by four video modeling examples based on principles from both social learning theory (Bandura 1977) and example-based learning (Renkl 2014; Van Gog and Rummel 2010). The video modeling examples were recorded using Camtasia Studio 8 (http://www.techsmith.com/camtasia/; cf. Kostons et al. 2012) and showed human models solving the problem as if on paper. This way the participants were able to follow the pen-movements as the model went through the steps of the problem-solving procedure. The problem statement, as well as all the steps that were performed by the models, were visible on the screen at all times, but the steps were gradually built up while the model was thinking-aloud during problem solving. The video modeling examples showed the model (male or female, see Table 1) performing a problem-solving task (at the first or second level of complexity, see Table 1). A problem statement was given at the top of the screen, and the model started by reading this aloud, and then continued to think aloud while problem solving. The model wrote out the solution to each step that s/he was able to complete (see Fig. 2).

To create variability in performance across video modeling examples (necessary for variability in self-assessment and task selection), models were unable to complete the task in two cases (see Table 1). After completing the problem-solving task, the model proceeded to the next page (in the case of the experimental conditions; explained below), where s/he rated how much mental effort s/he invested in solving that problem, by circling the answer on a scale of 1 to 9 (Paas 1992). The scale was presented horizontally, with labels at the uneven numbers: (1) very, very little effort, (3) little effort, (5) neither little nor much effort, (7) much effort, and (9) very, very much effort. The model then rated how many steps s/he thought s/he had performed correctly (i.e., self-assessment) on a scale ranging from 0 to 5, explaining that s/he assigned one point for each step correctly completed and then summed the points (the model's self-assessment was always correct, i.e. if she had performed 4 steps correctly, she would give a rating of 4). Finally, the model selected an appropriate subsequent task from the task database (while thinking aloud, using

Table 1 Features of the video modeling examples

Example #	Model	Complexity level	# Steps correct/self-assessment	Effort	Task-selection
1	Female	Level 1	5 Steps	2	+2
2	Male	Level 1	5 Steps	5	+1
3	Female	Level 2	4 Steps	7	0
4	Male	Level 2	3 Steps	8	−1

Fig. 2 Still from one of the video modeling examples showing how the model is able to write out the problem-solving steps underneath the problem statement

either the task-selection algorithm or a heuristic, depending on the condition; explained below) by circling that task.

Participants in all three conditions saw an introductory video defining the main concepts and observed the models performing the problem-solving task. However, only participants in the training conditions saw the subsequent effort rating, self-assessment, and task-selection parts (depending on their respective condition) of the video modeling examples. In the algorithmic condition, participants observed the model select a subsequent task (while thinking aloud) using the algorithm that was also used by Kostons et al. (2012). This algorithm combines scores on self-assessed performance and mental effort into an appropriate and specific task selection advice (see Fig. 3 and Table 1). For example, a performance rating of 4 combined with a mental effort rating of 6 would result in a task selection advice to go 1 step forward (i.e., one column to the right) in the task database, leading to a task with lower support at the same level of complexity, or a task with high

Fig. 3 Algorithm used for task-selection advice showing the jump size and direction for each of the combinations of self-assessed performance and mental effort

Performance			
4-5	+2	+1	0
2-3	+1	0	-1
0-1	0	-1	-2
	1-3	4-6	7-9 Effort

support at a higher level of complexity.[4] Participants in the heuristic condition saw the model select the exact same subsequent task, but using a general heuristic (underlying the algorithm) that used relative categories (e.g., low, medium, high performance) rather than absolute categories (e.g., zero or one step correct is low performance, two or three steps correct is medium performance, etc.) with regard to the (self-assessed) problem-solving score and mental effort ratings. For instance, in the above example (with a performance of 4 and effort of 6, the model would say "I attained a high score on performance with a medium amount of effort, so I am ready for a more difficult task or one with less support." During the time that participants in the training conditions observed the models self-assess their performance and select new tasks (approximately 90s), participants in the control condition were asked to remember and restate what was explained in the modeling example by typing in (step-by-step) what should be done at each step (cf. Stark et al. 2002).

Self-regulation phase

In the self-regulated learning phase, participants were presented with the task database and were instructed that they could work on eight tasks of their own choice. They were aware that a posttest would follow to assess what they had learned from those tasks. After each task, they were asked to rate how much mental effort they invested in solving the problem on the same 9-point rating scale (Paas 1992) as the models used. Then participants were asked to assess their own performance on a 6-point rating scale ranging from 0 to 5 and to select the next task from the task database that they wished to work on (after the eight task had been completed, they would automatically go on to the posttest).

Posttest

The posttest consisted of five problems, one of each complexity level. These problems had the same structure as the tasks in the database, but contained different surface features. The posttest problems were the same for all conditions, but after solving each problem participants were asked to rate how much mental effort they invested, to self-assess their performance, and to indicate what a suitable next task would be, to be able to assess self-assessment and task-selection accuracy (they knew though, that they would not actually get that task).

---

[4] Note that this algorithm proved to be effective for fostering learning outcomes in the Kostons et al. (2012) study and that other studies have shown that adaptive computer-regulated task selection using similar algorithms leads to better learning outcomes than fixed task sequences (e.g., Corbalan et al. 2008).

Transfer test

To test students' ability to transfer the trained task-selection skills, eight scenarios were used in which students had to indicate what a suitable next task would be for a peer student who had just completed a problem-solving task in a different domain (math), which could have the same or a different number of steps compared to the biology problems, and the database from which they had to select the new task could have the same or a different structure. The scenarios presented participants with information on a fictitious peer student's performance and invested effort on a math problem, and they could see the complexity and support level (shown in the task database) of that problem in the task database. Based on that information they had to indicate what new task that student should select, by clicking on that new task in the task database that was depicted below the problem. An example of a transfer test task is: "Eve has just performed a math problem of complexity level 2 without any support, consisting of eight steps. She rated her invested mental effort with a 2 on a scale from 1 to 9. She performed 1 step incorrectly. What kind of task should Eve select next from this task database?" The problem that the peer student had just performed was highlighted in the task database. Of these hypothetical scenarios, half contained a task database with 75 math tasks (layout cf. Fig. 1, but in math), and half contained a task database with 32 math tasks across 4 complexity levels with 2 support levels (Fig. 4). For each type of task database, half of the scenarios concerned 5-step problems, the other half concerned 8-step problems.

Procedure

Students were tested in five groups of around 25 participants per group, in a session that took approximately 100 min (two class periods). All three conditions were tested in each

Complexity level	Complexity 1		Complexity 2		Complexity 3		Complexity 4	
Support level	High support	No support	High support	No support	High support	No support	High support	No support
Learning tasks	Marbles	Marbles	Marbles	Marbles	Marbles	Marbles	Marbles	Marbles
	Balloons	Balloons	Balloons	Balloons	Balloons	Balloons	Balloons	Balloons
	Candy bars	Candy bars	Candy bars	Candy bars	Candy bars	Candy bars	Candy bars	Candy bars
	Blocks	Blocks	Blocks	Blocks	Blocks	Blocks	Blocks	Blocks

Fig. 4 Transfer task database version with 32 tasks in the mathematics domain

group. Participants were randomly assigned to conditions (by randomly handing out login codes that took them to the different conditions). First, students were given a general instruction and explanation about the procedure of the experiment. Then they received the pretest. In the subsequent training phase, students watched the four modeling examples (format depending on their assigned condition). This was followed by a self-regulated learning phase (SRL-phase) in which participants went through the following routine eight times: they chose and performed a problem-solving task, rated their mental effort and performance on this task and choose a next task, which they then would receive to repeat this process. After the SRL-phase participants received the posttest followed by the transfer test.

Data analysis

Performance on the pretest and posttest problem-solving tasks was scored by assigning one point for each correct step (i.e., range per problem: 0-5 points); performance on the pretest and posttest was then calculated by averaging the scores on the three (pretest) and five (posttest) problems (i.e., range: 0–5 points). Scores on the pretest and posttest were scored automatically by the online learning environment. Transfer test performance (i.e. task-selection accuracy on the scenarios) was determined by calculating the absolute difference between the task-selection step size that would be recommended based on the algorithm (for the 5-step problems) or adapted algorithm (for the 8-step problems) and the actual task level chosen by the participant. In the adapted algorithm, for the 8-step problems, low performance was 0, 1, or 2 steps performed correctly, medium performance was 3, 4, or 5 steps performed correctly and high performance was 6, 7, or 8 steps performed correctly.

   Thus, for the peer student (Eve) in the scenario that is given as an example above, who performed an 8-step math problem of complexity level 2 and without any support (i.e., column # 4), completing 7 steps correctly while investing very little effort (rating of 2), the next task that should be selected from the mathematics database according to the algorithm would be 2 steps ahead (i.e., to the right, to column # 6). If a student would instead choose a task that is 4 steps to the right (column 8), the task-selection accuracy score (i.e., deviation between recommended and chosen task) would be 2. The closer to 0 the average deviation score across the 8 tasks is, the more accurate a participant was.

Results

Table 2 shows the pretest, posttest, and transfer test data per condition. Unfortunately, mental effort, self-assessment, and task-selection data on the posttest were lost due to a logging error in the environment, so we could only analyze problem-solving performance on the posttest, not self-assessment or task-selection accuracy. Data were analyzed with ANOVAs or Kruskal–Wallis Tests when the assumption of normality was violated. Partial eta-squared (η_p^2) and Pearson's correlation (r) are reported as measures of effect size for ANOVA and Kruskal–Wallis Tests, respectively. The cutoffs for small, medium, and large effects are .01, .06, and .14, respectively, for partial eta-squared, and .10, .30, and .50, respectively, for Pearson's correlation.

Table 2 Mean (and SD) of performance on the pretest, posttest, and transfer test (i.e., task-selection accuracy) per condition

	Control (n = 43)		Heuristic (n = 37)		Algorithmic (n = 42)	
	M	SD	M	SD	M	SD
Pretest performance (max. 5)	0.45	0.38	0.56	0.36	0.48	0.37
Posttest performance (max. 5)	2.30	1.19	2.70	1.05	2.69	1.30
Task-selection transfer test (max. 9)[a]	2.00	0.84	1.19	0.79	1.39	1.03

[a]Lower = better

Pretest (randomization check)

Overall, average performance on the pretest problems was very low ($M = 0.49$ out of 5, $SD = 0.37$). A Shapiro-Wilk's test showed that pretest performance scores were not normally distributed and therefore an Independent-Samples Kruskal-Wallis Test was conducted, which revealed no significant differences between conditions, $\chi^2(2) = 2.152$, $p = .341$, $r = .13$.

Posttest

Average score on the posttest problems across conditions was 2.55 (out of 5; $SD = 1.19$). Because of significant bimodality in the dataset (Hartigan's Dip Test of Unimodality: $D = 0.049$, $p = .028$; Hartigan and Hartigan 1985; Bimodality Coefficient: $BC = 0.563 > BC_{crit}$; SAS Institute Inc. 1990), an Independent-Samples Median Test was performed, which showed a significant difference between conditions, $\chi^2(2) = 6.641$, $p = .036$, $r = .23$. Post-hoc Median Tests showed that in line with Hypothesis 1, both heuristic training, $\chi^2(1) = 4.732$, $p = .030$, $r = .24$, and algorithmic training, $\chi^2(1) = 5.195$, $p = .023$, $r = .25$, led to significantly better problem-solving performance on the posttest than the control condition. However, problem-solving performance did not differ between training conditions, $\chi^2(1) = 0.658$, $p = .417$, $r = .09$.

Transfer test

An Independent-Samples Kruskal-Wallis Test revealed that task-selection accuracy on the transfer test differed between conditions, $\chi^2(2) = 19.228$, $p < .001$, $r = .40$. Post-hoc Mann–Whitney U Tests showed that in line with Hypothesis 2a, both the heuristic training, $U = 355.5$, $p < .001$, $r = .48$, and algorithmic training, $U = 537.5$, $p = .001$, $r = .35$, showed significantly better task-selection accuracy on the transfer test than the control group (i.e., lower scores indicate better task-selection accuracy, see Table 2). However, in contrast to Hypothesis 2b, the heuristic and algorithmic training conditions did not differ from each other, $U = 724.0$, $p = .602$, $r = .06$.

Discussion

The first aim of this study was to replicate prior findings (Kostons et al., 2012) showing that the effectiveness of self-regulated learning in an online environment on biology problem-solving tasks is enhanced when self-assessment and task-selection are first trained through video modeling examples. In line with our first hypothesis, students who had received self-assessment and task-selection training (i.e., in both training conditions) performed significantly better on the posttest than students who did not, showing that these students gained more knowledge during the self-regulated learning phase. There was, however, no significant difference in posttest performance between the two training conditions.

The second aim was to test whether we could find indications that the trained task-selection skills would transfer to a different domain, in our case to mathematics. In line with Hypothesis 2a, students who had received self-assessment and task-selection training were better at selecting tasks in a different domain than students who did not. The third aim was to investigate whether a heuristic training condition would be more effective at fostering task-selection transfer than the original algorithmic training condition from Kostons et al. (2012). In contrast to Hypothesis 2b, the heuristic training condition was not better at selecting tasks in a different domain than the algorithmic training condition. This does not seem to agree with the idea that bridging (i.e., abstracting a general principle and recognizing when to use it; Salomon and Perkins 1989) fosters successful transfer. However, it is possible that students in the algorithmic training condition inferred the principle underlying the algorithm. In other words, learners in the algorithmic condition might have been able to transform the specific algorithm into a less effortful heuristic (Shah and Oppenheimer 2008), which would explain why we found no difference between the heuristic and algorithmic training condition. Future research could address this possibility by interviewing learners after the training, or by asking them to think aloud during task selection.

A limitation of the present study is that the self-assessment and task-selection data at posttest were lost and that there was some bimodality in the posttest problem-solving performance data. This bimodality suggests that a substantial number of learners benefitted very little from engaging in self-regulated learning whereas a substantial number of others did gain a lot from the self-regulated learning phase. It is possible that individual differences such as motivation or achievement goals influenced either how much students learned from the modeling examples, or whether they applied what they learned from the examples during self-regulated learning (Winne and Hadwin 2008). For instance, motivation might be influenced by students' preference for a learning environment (i.e., online vs. face-to-face; Johnson et al. 2000) and motivation may, consequently, affect self-regulated learning. Future research could take into account the potential influence of individual differences in motivation or achievement goals on the effect of self-regulated learning skills training.

Another potential limitation is the way in which we operationalized our transfer test. We measured the transfer of task-selection skills by means of scenarios, in which students had to select a new task for a fictitious peer student in a different domain (math instead of biology as in the trained tasks), in which the problems sometimes differed in the number of steps (8 instead of 5 as in the trained tasks), and the task database sometimes had a different layout (with 32 instead of 75 problems). Although these scenarios did measure whether a learner had understood the task-selection rule and could apply it in a different domain, the degree of transfer required is arguably rather limited. Learners were given the

input they needed to make a decision and could fully devote their attention to task-selection, which is much less cognitively demanding than having to engage in performing these novel math tasks and having to self-assess performance and select a new task for yourself from a different-looking database.

A second limitation of the task-selection transfer test concerns the scoring of the task-selection accuracy. The task-selection accuracy of the heuristic condition on the transfer test was evaluated against the correct implementation of the algorithmic task-selection rule (which had proven, in the Kostons et al. 2012, study, to be effective for improving self-regulated learning as assessed by posttest problem-solving performance). This might have negatively biased the scoring of the task-selection accuracy in the heuristic condition. Nevertheless, task-selection accuracy in the heuristic condition was comparable to the algorithmic condition and there were no differences in posttest problem-solving performance, which one would have expected had the heuristic condition actually been more accurate in task-selection (i.e., by another accuracy measure).

Nevertheless, our findings are promising, and provide an important first step towards determining whether task-selection skills would have to be trained anew for every type of task (which would be highly impractical), or whether they can transfer to other types of tasks in other domains. Our findings are also relevant for educational practice, in that they show that a relatively simple intervention can help students gain more from self-regulated learning. It should be noted though, that our study—as demonstrated by the experiments presented in the supplementary materials—also revealed that the way in which the video modeling examples are designed can have a significant impact on the success of the training. This should be kept in mind when implementing video modeling examples for training self-regulation skills.

Appendix 1

Example of problem-solving task used in pretest and posttest (first level of complexity)

Fur color

A guinea pig's fur color is determined by a gene, which expresses itself as black in its dominant form (F) and white in its recessive form (f). Two guinea pigs, who are both black and homozygote for that trait, produce offspring. What are the possible genotypes for this offspring?

Step 1 Translate information from text into genotypes.

– Both guinea pigs are homozygote for the dominant allele, so both genotypes are FF.

Step 2 Fill in a family tree.

Step 3 Determine number of Punnett squares by deciding if problem is to be solved deductively or inductively.

– Both parents are given, so we can solve the problem deductively. Solving problems deductively only requires one Punnett square.

Step 4 Fill in the Punnett square.

	F	F
F	FF	FF
F	FF	FF

Step 5 Find the answer in the Punnett square.

The only possible genotype for the offspring is FF.

References

Alexander, P. A., Dinsmore, D. L., Parkinson, M. M., & Winters, F. I. (2011). Self-regulated learning in academic domains. In B. J. Zimmerman & D. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 393–407). New York, NY: Routledge.

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96,* 523–535. https://doi.org/10.1037/0022-0663.96.3.523.

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition: Implications for the design of computer-based scaffolds. *Instructional Science, 33,* 367–379. https://doi.org/10.1007/s11251-005-1272-9.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Bannert, M. (2006). Effects of reflection prompts when learning with hypermedia. *Educational Computing Research, 35,* 359–375. https://doi.org/10.2190/94V6-R58H-3367-G388.

Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science, 40,* 193–211. https://doi.org/10.1007/s11251-011-9167-4.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64,* 417–444. https://doi.org/10.1146/annurev-psych-113011-143823.

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

Costa Ferreira, P., Veiga Simão, A. M., & Lopes da Silva, A. (2015). Does training in how to regulate one's learning affect how students report self-regulated learning in diary tasks? *Metacognition and Learning, 10,* 199–230. https://doi.org/10.1007/s11409-014-9121-3.

Dabbagh, N., & Kitsantas, A. (2005). Using web-based pedagogical tools as scaffolds for self-regulated learning. *Instructional Science, 33,* 513–540. https://doi.org/10.1007/s11251-005-1278-3.

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review, 28,* 425–474. https://doi.org/10.1007/s10648-015-9320-8.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3,* 231–264. https://doi.org/10.1007/s11409-008-9029-x.

Donovan, M. S., Bransford, J. D., & Pellegrino, J. W. (1999). *How people learn: Bridging research and practice*. Washington, DC: National Academies Press.

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

Fiorella, L., & Mayer, R. E. (2016). Observing the instructor draw diagrams on learning from multimedia messages. *Journal of Educational Psychology, 108,* 528–546. https://doi.org/10.1037/edu0000065.

Greene, J. A., Bolick, C. M., Jackson, W. P., Caprino, A. M., Oswald, C., & McVea, M. (2015). Domain-specificity of self-regulated learning processing in science and history. *Contemporary Educational Psychology, 42,* 111–128. https://doi.org/10.1016/j.cedpsych.2015.06.001.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics, 13,* 70–84. https://doi.org/10.1214/aos/1176346577.

Johnson, S. D., Aragon, S. R., Shaik, N., & Palma-Rivas, N. (2000). Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments. *Journal of Interactive Learning Research, 11,* 29–49.

Kimball, D., & Holyoak, K. (2000). Transfer and expertise. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 109–122). Oxford: Oxford University Press.

Koedinger, K. R., Aleven, V., Roll, I., & Baker, R. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 897–964). New York, NY: Routledge.

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,* 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.

Kramarski, B., & Gutman, M. (2006). How can self-regulated learning be supported in mathematical E-learning environments. *Journal of Computer Assisted Learning, 22,* 24–33. https://doi.org/10.1111/j.1365-2729.2006.00157.x.

Leidinger, M., & Perels, F. (2012). Training self-regulated learning in the classroom: Development and evaluation of learning materials to train self-regulated learning during regular mathematics lessons at primary school. *Education Research International, 2012,* 1–14. https://doi.org/10.1155/2012/735790.

Loyens, S. M. M., Magda, J., & Rikers, R. M. J. P. (2008). Self-directed learning in problem-based learning and its relationships with self-regulated learning. *Educational Psychology Review, 20,* 411–427. https://doi.org/10.1007/s10648-008-9082-7.

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). New York, NY: Academic Press.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84,* 429–434. https://doi.org/10.1037/0022-0663.84.4.429.

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice, 21,* 133–148. https://doi.org/10.1080/0969594X.2013.877872.

Perels, F., Gürtler, T., & Schmitz, B. (2005). Training of self-regulatory and problem-solving competence. *Learning and Instruction, 15,* 123–139. https://doi.org/10.1016/j.learninstruc.2005.04.010.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82,* 33–40. https://doi.org/10.1037/0022-0663.82.1.33.

Poitras, E. G., & Lajoie, S. P. (2013). A domain-specific account of self-regulated learning: The cognitive and metacognitive activities involved in learning through historical inquiry. *Metacognition and Learning, 8,* 213–234. https://doi.org/10.1007/s11409-013-9104-9.

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

Rawson, K. A., O'Neil, R., & Dunlosky, J. (2011). Accurate monitoring leads to effective control and greater learning of patient education materials. *Journal of Experimental Psychology: Applied, 17,* 288–302. https://doi.org/10.1037/a0024749.

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

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

Roebers, C. M., Schmid, C., & Roderer, T. (2009). Metacognitive monitoring and control processes involved in primary school children's test performance. *British Journal of Educational Psychology, 79,* 749–767. https://doi.org/10.1348/978185409X429842.

Roll, I., Wiese, E. S., Long, Y., Aleven, V., & Koedinger, K. (2014). Tutoring self- and co-regulation with intelligent tutoring systems to help students acquire better learning skills. In R. A. Sottilare, A. C. Graesser, X. Hu, & B. S. Goldberg (Eds.), *Design recommendations for intelligent tutoring systems* (Vol. 2, pp. 169–182). Orlando, FL: US Army Research.

Salden, R. J., Paas, F., & Van Merriënboer, J. J. (2006). Personalised adaptive task selection in air traffic control: Effects on training efficiency and transfer. *Learning and Instruction, 16,* 350–362. https://doi.org/10.1080/09541440701326022.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanism of a neglected phenomenon. *Educational Psychologist, 24,* 113–142. https://doi.org/10.1207/s15326985ep2402_1.

SAS Institute Inc. (1990). *SAS/STAT User's Guide, Version 6* (4th ed.). Cary, NC: SAS Institute.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26,* 113–125. https://doi.org/10.1023/A:100304423.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin, 134,* 207–222. https://doi.org/10.1037/0033-2909.134.2.207.

Stark, R., Mandl, H., Gruber, H., & Renkl, A. (2002). Conditions and effects of example elaboration. *Learning and Instruction, 12,* 39–60. https://doi.org/10.1016/S0959-4752(01)00015-9.

Thiede, K. W., Anderson, M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95,* 66–73. https://doi.org/10.1037/0022-0663.95.1.66.

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22,* 155–174. https://doi.org/10.1007/s10648-010-9134-7.

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

Van Merriënboer, J. J. G., Clark, R. E., & De Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development, 50,* 39–61. https://doi.org/10.1007/BF02504993.

Van Merriënboer, J. J. G., & Kirschner, P. A. (2018). *Ten steps to complex learning: A systematic approach to four-component instructional design* (3rd ed.). New York, NY: Routledge.

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). Hillsdale, NJ: Erlbaum.

Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). New York, NY: Taylor & Francis.

Winne, P. H., Nesbit, J. C., Kumar, V., Hadwin, A. F., Lajoie, S. P., Azevedo, R., et al. (2006). Supporting self-regulated learning with gStudy software: The learning kit project. *Technology, Instruction, Cognition, and Learning, 3,* 105–113.

Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*. Mahwah, NJ: Erlbaum.