SHORT PAPER - 5000 WORDS OR LESS

WILEY

Training self-assessment and task-selection skills to foster self-regulated learning: Do trained skills transfer across domains?

Steven F. Raaijmakers[1] | Martine Baars[2] | Fred Paas[2,3] | Jeroen J. G. van Merriënboer[4] | Tamara van Gog[1]

[1] Department of Education, Utrecht University, Utrecht, The Netherlands

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

[3] Early Start Research Institute, University of Wollongong, Wollongong, Australia

[4] Department of Educational Development and Research, Maastricht University, Maastricht, The Netherlands

Correspondence
Steven Raaijmakers, Department of Education, Utrecht University, PO Box 80140, 3584 CS Utrecht, The Netherlands.
Email: s.f.raaijmakers@uu.nl

Summary

Students' ability to accurately self-assess their performance and select a suitable subsequent learning task in response is imperative for effective self-regulated learning. Video modeling examples have proven effective for training self-assessment and task-selection skills, and—importantly—such training fostered self-regulated learning outcomes. It is unclear, however, whether trained skills would transfer across domains. We investigated whether skills acquired from training with either a specific, algorithmic task-selection rule or a more general heuristic task-selection rule in biology would transfer to self-regulated learning in math. A manipulation check performed after the training confirmed that both algorithmic and heuristic training improved task-selection skills on the biology problems compared with the control condition. However, we found no evidence that students subsequently applied the acquired skills during self-regulated learning in math. Future research should investigate how to support transfer of task-selection skills across domains.

KEYWORDS

example-based learning, self-assessment, self-regulated learning, task selection, transfer

1 | INTRODUCTION

Developing self-regulated learning skills is important to prepare secondary education students for future learning in higher education and workplace settings (Bransford, Brown, & Cocking, 2000). In learning environments in which students get the freedom to choose their own learning tasks, two skills are crucial for self-regulated learning to be effective: self-assessment and task selection. When students are not able to accurately evaluate their own performance (self-assessment) and select an appropriate new learning task in response (task selection), learning outcomes in the domain will be suboptimal, as students will end up working on learning tasks that are either too easy or too difficult for them. Because accurate self-assessment and task selection is difficult (e.g., Bjork, Dunlosky, & Kornell, 2013), researchers in educational and (applied) cognitive psychology have been concerned with finding means to support or scaffold self-assessment and task selection *during* self-regulated learning (e.g., Azevedo & Hadwin, 2005; Bannert, 2006; Dabbagh & Kitsantas, 2005; Kramarski &

Gutman, 2006; Winne et al., 2006) or to train those skills *prior to* self-regulated learning (e.g., Azevedo & Cromley, 2004; Costa Ferreira, Veiga Simão, & Lopes da Silva, 2015; Kostons, Van Gog, & Paas, 2012; Leidinger & Perels, 2012; Perels, Gürtler, & Schmitz, 2005) with the aim of enhancing students' learning outcomes.

One training method that has proven effective for training self-assessment and task-selection skills, and for fostering domain-specific learning outcomes, is the use of video modeling examples (Kostons et al., 2012; Raaijmakers et al., 2017). In those video modeling examples, another person (the model) first performed the task (i.e., a problem-solving task in the domain of biology), then assessed his or her own performance on the task (i.e., self-assessment, by assigning a point for each correctly performed problem-solving step), and, finally, chose a suitable subsequent task from a database with tasks of different levels of complexity and support (i.e., task selection, determining whether to select a task with higher, equal, or lower levels of support or complexity, based on a combination of self-assessed performance and mental effort invested to reach that performance). Students who were trained with

these video modeling examples showed better domain-specific learning outcomes, as well as more accurate self-assessments and task selections after a self-regulated learning phase (i.e., they had to work on eight biology problem-solving tasks that they could freely choose from the task database) than students in the control condition (Kostons et al., 2012).

An important open question, however, is whether the trained self-regulated learning skills would transfer to other domains, other environments, or other types of tasks (Koedinger, Aleven, Roll, & Baker, 2009; Roll, Wiese, Long, Aleven, & Koedinger, 2014). 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? Raaijmakers et al. (2017) started to address this question. Next to a no-training control condition[1] and the algorithmic task-selection training condition used by Kostons et al. (2012), which combined self-assessed performance and mental effort into a specific task-selection advice (i.e., should a student go forward or backward in support/complexity and how far), they implemented a more general, heuristic task-selection training (e.g., "when performance is high and invested effort is low, select a task that offers less support or is more complex"). They expected that the heuristic training condition would be more conducive to transfer as it is less dependent on the specific task details or database characteristics (number of support/complexity levels). Using a similar design and materials as Kostons et al. (2012), students first engaged in training, then engaged in self-regulated learning, followed by a problem-solving posttest (biology problems similar to the learning phase), and finally a transfer test (i.e., task selection in a different domain). Transfer of task-selection skills was assessed 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 problem-solving steps (eight instead of five as in the trained tasks), and the task database sometimes had a different layout (with 32 instead of 75 problems with different complexity and support levels). Results showed that both the heuristic and algorithmic training of self-assessment and task-selection skills improved posttest performance on the biology problem-solving tasks after a self-regulated learning phase (replicating and extending the findings of Kostons et al., 2012). Importantly, both training conditions also showed better transfer of task-selection skills (i.e., self-assessment skills were not measured during the transfer test) than the control condition (however, they did not differ from each other).

Note though, that 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 was arguably rather limited. Learners were given the input they needed to make a decision (self-assessed performance, self-assessed invested mental effort, and the complexity and support level of the previously performed task) 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 select a new task for yourself from a different-looking database. According to cognitive load theory, a secondary task (such as monitoring performance or

thinking about task-selection rules *while* working on the primary, problem-solving task) can harm performance because the additional load involved in processing the secondary task would exceed the limited capacity of our working memory (Van Gog, Kester, & Paas, 2011; Van Merriënboer & Sweller, 2005). Having to perform a task yourself *and* having to select new tasks might therefore harm performance. Finally, having to assess your own performance and select a new task for yourself may differ from doing this for a peer (Panadero, Brown, & Strijbos, 2016). For decision making in general, the consensus is that decisions made for oneself are more risk averse (or less risk prone) than decisions made for others (Polman, 2012). Applied to task selection, this would mean that learners would be more inclined to select easy tasks when selecting tasks for themselves and would be more inclined to select difficult tasks when selecting tasks for others (i.e., choosing difficult tasks increases the risk of failing). Thus, task selection on the transfer tasks in Raaijmakers et al. (2017) using scenarios with fictitious peers might be biased due to these factors. Therefore, it is crucial to determine if task-selection skills also transfer when learners select tasks for themselves (after performing the problem-solving task).

1.1 | The present study

In the present study, we investigated if self-assessment and task-selection skills trained with video modeling examples (cf. Kostons et al., 2012; Raaijmakers et al., 2017) in one domain (biology) would transfer to a different domain (mathematics) in the sense that learners would reach higher domain-specific learning outcomes after self-regulated learning in math when they were trained compared with when they were not trained. We included two training conditions: one in which the model used the algorithm for task selection (i.e., specific advice on complexity/support level) and one in which the model used a more general heuristic (e.g., when performance is high and perceived effort is low, select a task that offers less support or is more complex). We hypothesized that self-assessment and task-selection training in the context of biology problem solving would result in better task-selection skills (i.e., on a transfer test without self-assessment; cf. Raaijmakers et al. 2017) in the same context (Hypothesis 1) and would result in higher posttest problem-solving performance in math after a self-regulated learning phase in math (i.e., algorithmic and heuristic training > no training; Hypothesis 2a) as well as higher self-assessment and task-selection accuracy on the posttest (Hypothesis 3a).

Presumably, the general heuristic will function as a bridge between the two contexts (i.e., biology domain and mathematics domain; Salomon & Perkins, 1989) that allows for better transfer of task-selection skills. However, as Raaijmakers et al. (2017) did not find this advantage of the heuristic over the algorithmic rule with students selecting tasks for fictitious peers, it is still an open question if transfer could be improved by the heuristic. Having to solve problems yourself might increase the cognitive load to such an extent that the heuristic will become necessary to lower the cognitive load. If this is correct, we would expect that the heuristic task-selection training condition would show better transfer of task-selection skills, as evidenced by better performance on the math posttest, than the algorithmic group (Hypothesis 2b) and higher self-assessment and task-selection accuracy on the posttest (Hypothesis 3b).

---

[1] Students in the control condition observed the performance phase of the modeling examples, but not the self-assessment and task-selection phase.

2 | METHOD

2.1 | Participants and design

A total of 84 students in their first year of Dutch secondary education (senior general secondary and preuniversity education) participated in this study. Six participants did not manage to finish the experiment within the two available lesson periods and had to be excluded, leaving 78 participants (mean$_{age}$ = 12.22 years, standard deviation [SD]$_{age}$ = 0.44; 28 boys and 50 girls). Participants were randomly assigned to one out of three conditions: (a) a control condition (n = 26): video modeling examples in which the model performed a problem-solving task, (b) an algorithmic training condition (n = 24): video modeling examples in which the model performed a problem-solving task, rated invested mental effort, self-assessed performance, and used an algorithm to select a subsequent task, and (c) a heuristic training condition (n = 28): video modeling examples in which the model performed a problem-solving task, rated invested mental effort, self-assessed performance, and used a heuristic to select a subsequent task. The timing in the curriculum was chosen, so the participants had no prior knowledge of the specific problem-solving procedure of both the math problems and the biology problems but were ready to acquire it.

2.2 | Materials

All materials were presented online through a learning environment specifically designed for this study.

2.2.1 | Training phase: Video modeling examples and manipulation check (biology)

The training consisted of four video modeling examples previously used in Raaijmakers et al. (2017) and based on Kostons et al. (2012). The video modeling examples were screen recordings that showed the model (male or female, see Table 1) performing a problem-solving task in the domain of biology (i.e., monohybrid cross problem at the first or second level of complexity; see Table 1), self-assessing their performance and invested mental effort (an indicator of experienced cognitive load), and selecting a subsequent task. The biology task database consisted of five levels of complexity, with three levels of support within each complexity level, and five isomorphic tasks for each combination of complexity and support, creating a total of 75 problems. The biology tasks were five-step problem-solving tasks that were solvable using a set procedure. The models used the same problem-solving procedure in each video modeling example. Performance was rated by the models using a 6-point scale ranging from 0 to 5, and mental effort was rated using a 9-point scale with labels at the uneven numbers: (1) *very, very little mental effort*, (3) *little mental effort*, (5) *neither little nor*

much mental effort, (7) *much mental effort*, and (9) *very, very much mental effort* (Paas, 1992). To ensure variability in self-assessment and task selection between the video modeling examples, performance was varied (i.e., models did not complete the task in two cases; see Table 1).

Participants in the algorithmic condition were shown a model selecting a subsequent task—while thinking aloud—using the algorithm also used by Kostons et al. (2012) and Raaijmakers et al. (2017). This algorithm combines scores on self-assessed performance and mental effort into a specific task-selection advice (see Figure 1 and Table 1). For example, if a learner gave his or her performance a self-assessed score of 4 and the invested mental effort a rating of 2, this would result in a task-selection advice of +2 (i.e., go two columns to the right in the task database).

Participants in the heuristic condition were shown a model selecting the same subsequent task, but now, task selection was based on a general heuristic (underlying the algorithm). Using the above example (self-assessed performance of 4 and invested mental effort of 2), the model would say "I attained a high score on performance with a very low amount of effort, so I am ready for a more difficult task or one with less support." During the time that participants in the experimental conditions watched the second and third part of the video modeling examples, participants in the control condition were asked to describe what the five steps of the problem-solving task in the video modeling example were (i.e., typing them in step-by-step; cf. Stark, Mandl, Gruber, & Renkl, 2002).

To check if participants had indeed acquired task-selection skills from the video modeling examples, eight scenarios were used in which students had to indicate what a suitable next task would be for a fictitious peer student who had just completed a five-step heredity problem-solving task (i.e., similar to the tasks used in the video modeling

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

FIGURE 1 Algorithm used for task-selection advice used in the video modeling examples (for the five-step biology problems) showing the jump size and direction in the task database (− indicates one or two rows to the left and + indicates one or two rows to the right) for the combination of self-assessed performance and mental effort

TABLE 1 Features of the video modeling examples

Example #	Model	Complexity level	# of 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

examples). Participants were given a fictitious peer student's self-assessed performance and invested effort on a problem from the task database and were shown the complexity and support level of that problem (highlighted in the task database). With this information, participants had to select an appropriate subsequent task for the fictitious peer by clicking on that task in the task database. An example of a scenario is

> Eve has just performed a biology problem of complexity level 2 without any support, consisting of 5 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?

2.2.2 | Math problem-solving tasks

The tasks used to assess whether the self-assessment and task-selection skills acquired from the training in the context of biology problems would transfer to another domain were math problems in which students needed to find the linear equation of a given line. The problems could be solved in three steps: (a) identify the slope using two known points, (b) find the y-intercept, and (c) use the slope and the y-intercept to complete the equation. Appendix 1 shows an example of a task from the third level of complexity with high support.

The task database with math problems contained 40 problems at five levels of complexity, with two levels of support within each complexity level (see Figure 2b). The levels of complexity (top row of Figure 2b) were designed in collaboration with mathematics teachers and pilot-tested on a separate group of participants. Tasks at Complexity Level 1 only contained an intercept. Tasks at Complexity Level 2 only contained positive slopes. Tasks on Complexity Level 3 contained both an intercept and positive slopes. Tasks at Complexity Level 4 introduced negative slopes. Finally, in tasks at Complexity Level 5, the y axis was not visible in the graph, which meant that the intercept could not be read of off the graph directly and had to be deduced. Each level of complexity contained two levels of support (see the second row in Figure 2b): high support, where the first two steps were worked out, leaving one step for the learner to complete, and no support, where no steps were worked out and the learner had to solve the entire problem without any assistance. The combination of five levels of complexity and two levels of support created 10 columns in which the tasks were organized. In each column, four isomorphic tasks were presented, resulting in a total of 40 tasks.

2.2.3 | Pretest

The math pretest was used to check whether students were indeed novices regarding the topic at hand. It consisted of three problem-solving tasks without support (one task at each of the first three complexity levels, ordered from low to high complexity). These tasks had the same structure as the tasks in the database but contained different slopes and intercepts. After each problem, participants were asked to rate their invested mental effort and to self-assess their performance.

2.2.4 | Self-regulated learning phase

In the self-regulated learning phase, participants were instructed to select and perform (successively) eight tasks of their own choice from the math task database. Participants were asked to rate how much mental effort they invested in solving the problem on the 9-point rating scale (Paas, 1992) and how well they performed on a 4-point rating scale ranging from 0 to 3. Next, they selected the task they would work on next from the math task database. When they had chosen and performed eight tasks, participants automatically went through to the posttest.

2.2.5 | Posttest

The math posttest consisted of five different problems, one from each level of complexity. The problems were structurally similar to the tasks in the database but contained different surface features. After each problem, participants were asked to rate their invested mental effort, self-assess their performance, and indicate what a suitable subsequent task would be. They did not actually get this task, and the posttest was the same for all participants, but this allowed us to calculate task-selection accuracy on the posttest. Participants were informed that they would not actually get the selected tasks on the posttest.

2.3 | Procedure

Test sessions took approximately 90 min (two lesson periods). Four classes participated, and participants were randomly assigned to conditions within each class by means of login codes that allocated them to the different conditions (i.e., all conditions were present in each class). During a session, the experimenter first explained the general procedure of the experiment after which students were allowed to log in to the learning environment. After performing the pretest, students watched the four video modeling examples. Which specific parts of the videos the participants got to see depended on their assigned condition. After the videos, participants received the task-selection skills training check, and this was followed by a self-regulated learning phase. During the self-regulated learning phase, students repeated the following cycle eight times: They chose and performed a problem-solving task, rated their mental effort and performance on this task, and chose a next task, which they would receive to repeat the cycle. Finally, students completed the posttest.

2.4 | Data analysis

Performance on the math pretest and posttest problem-solving tasks was scored by assigning 1 point for each correct step (i.e., range per problem: 0–3 points); a performance score on the pretest and posttest was then calculated by averaging the scores on the three (pretest) and five (posttest) problems (i.e., performance score range: 0–3).

Task-selection accuracy on the manipulation check scenarios was determined by calculating the absolute difference between the task that should have been selected based on the algorithm (for the five-step biology problems; see Figure 1) and the actual task selected by the participant in the biology task database (i.e., a self-assessed performance of 5 and a mental effort of 1 result in a task-selection advice of +2). On the posttest, task-selection accuracy was calculated in a similar manner but with an algorithm adapted for the three-step problems. Low performance was defined as zero steps performed correctly, medium performance as one or two steps performed correctly, and high performance as three steps performed correctly

(a)

(b)

FIGURE 2 (a) Biology task database containing the 75 problem-solving tasks showing the different levels of complexity, the different levels of support, and the different surface features of the learning tasks. (b) Mathematics task database containing the 40 math problem-solving tasks showing the different levels of complexity and the different levels of support

(i.e., a self-assessed performance of 3 and a mental effort of 1 would result in a task-selection advice of +2).

3 | RESULTS

Table 2 shows the pretest, posttest, and transfer test data per condition. Data were analyzed with analyses of variance (ANOVAs) or with Kruskal–Wallis tests when Shapiro–Wilk's test showed that 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 (Cohen, 1988).

TABLE 2 Mean (M and standard deviation [SD]) of performance and mental effort on the math pretest, the task-selection manipulation check (i.e., task-selection accuracy on biology scenarios after training), and performance, mental effort, self-assessment accuracy, and task-selection accuracy on the math posttest, per condition

	Control (n = 26)		Heuristic (n = 28)		Algorithmic (n = 24)	
	M	SD	M	SD	M	SD
Pretest performance (range: 0–3)	0.10	0.23	0.02	0.09	0.10	0.21
Pretest mental effort (range: 1–9)	6.56	2.13	5.45	2.45	6.50	1.58
Task-selection check (range: 0–9)[a]	2.69	1.00	1.67	0.81	1.28	1.11
Posttest performance (range: 0–3)	0.63	0.60	0.76	0.82	0.98	0.61
Posttest mental effort (range: 1–9)	5.77	2.34	5.23	2.62	4.93	2.51
Posttest SA accuracy (range: 0–3)[a]	1.21	0.92	1.24	1.05	1.44	0.77
Posttest TS accuracy (range: 0–9)[a]	2.87	1.10	2.25	1.40	2.25	1.28

[a]Lower = better; SA = self-assessment; TS = task selection.

3.1 | Task-selection skills training manipulation check

To determine whether the training with video modeling examples had been effective, we compared performance on the task-selection manipulation check (biology problem scenarios), expecting the training conditions to outperform the no-training condition (Hypothesis 1). An independent-samples Kruskal–Wallis test showed that the task-selection accuracy scores differed between conditions, $\chi^2(2) = 21.639$, $p < .001$, $r = .53$. Post hoc Mann–Whitney U tests showed that the heuristic training, $U = 152.0$, $p < .001$, $r = .50$, and algorithmic training, $U = 111.0$, $p < .001$, $r = .55$, showed significantly higher task-selection accuracy than the control group (i.e., lower scores indicate better task-selection accuracy, see Table 2). Moreover, algorithmic training led to significantly higher task-selection accuracy than heuristic training, $U = 219.5$, $p = .032$, $r = .30$.

3.2 | Math pretest performance (randomization and prior knowledge check)

As expected, overall performance on the math pretest problems was very low (mean = 0.07 out of 3, SD = 0.18). As a randomization check, pretest performance and self-reported mental effort invested in the pretest were compared between conditions. An independent-samples Kruskal–Wallis test revealed no significant differences between conditions in pretest performance, $\chi^2(2) = 2.500$, $p = .287$, $r = .18$, or effort investment, $\chi^2(2) = 3.437$, $p = .179$, $r = .21$.

3.3 | Math posttest

Average performance on the math posttest problems across conditions was 0.78 (out of 3; SD = 0.70). To test if posttest performance differed between conditions, with performance being higher in the trained conditions than in the control condition (Hypothesis 2a) and higher in the heuristic than in the algorithmic training condition (Hypothesis 2b), an independent-samples Kruskal–Wallis test was performed, which revealed that posttest performance did not differ significantly between conditions, $\chi^2(2) = 3.512$, $p = .173$, $r = .21$. Moreover, mental effort ratings did not differ between conditions, $F(2, 75) = 0.74$, $p = .481$, $\eta_p^2 = .019$.

To test if self-assessment and task-selection accuracy on the math posttest differed between conditions (Hypothesis 3a,b), we performed an independent-samples Kruskal–Wallis test on the self-assessment accuracy data and a one-way ANOVA on the task-selection accuracy data. These analyses showed no significant differences between conditions in self-assessment, $\chi^2(2) = 1.757$, $p = .415$, $r = .15$, or task selection $F(2, 75) = 2.06$, $p = .135$, $\eta_p^2 = .052$.

4 | DISCUSSION

Prior research has shown that training self-assessment and task-selection skills with video modeling examples improved learning outcomes after a self-regulated learning phase in which students worked on the same kind of problems as demonstrated in the examples (Kostons et al., 2012; Raaijmakers et al., 2017). Moreover, some evidence was found that these skills might transfer; students who had received training and engaged in self-regulated learning in biology made more accurate task-selection choices for fictitious peers (based on accurate information on the peer's performance and invested effort on math problems that had a different number of steps and came from a task database with a different structure; Raaijmakers et al., 2017). This was a rather limited form of transfer, though, and we would want learners to be able to apply the self-regulated learning skills that they learned in one domain also when studying in a different domain.

Therefore, the aim of the present experiment was to establish whether self-assessment and task-selection skills trained in one domain (biology) would transfer, that is, would be applied during self-regulated learning, and therefore lead to better posttest performance in a different domain (mathematics). Secondary education students first engaged in self-assessment and task-selection training with video modeling examples (or not, in the control condition) followed by a manipulation check. Then they engaged in self-regulated learning in math followed by a math posttest.

The manipulation check on whether participants acquired task-selection skills from the training with video modeling examples indeed confirmed that training improved task-selection accuracy, with participants in the algorithmic condition being most accurate in selecting new biology tasks, followed by the heuristic condition that, in turn, was more accurate than the control condition (Hypothesis 1). However, there was no evidence of transfer: There were no differences among conditions in math posttest performance (or self-assessment and task-selection accuracy at posttest; Hypotheses 2a/b and 3a/b), suggesting that students failed to apply the skills trained with biology tasks during the self-regulated learning phase with mathematics tasks.

One possible explanation for why task-selection skills did not transfer is that students might have been unable to map what they had learned with five-step biology problems and a task database with five complexity levels and three levels of support within each complexity level, onto three-step math problems and a task database with five complexity levels and two support levels. Using different metrics for the assessment of performance creates difficulties for transfer by increasing the distance between the source (here: biology self-assessment/task selection) and target (here: mathematics self-assessment/task selection) of transfer (Kimball & Holyoak, 2000; Salomon & Perkins, 1989). We had expected the heuristic to increase the similarity between source and target by transforming the five problem-solving steps into three reference categories (i.e., high/medium/low performance) making it less dependent on the specific task (number of problem-solving steps) or database characteristics (number of support/complexity levels) than the algorithmic training condition but found no indications that this was the case.

However, mapping problems would not explain why prior research (Raaijmakers et al., 2017) did find evidence of transfer. In a prior study, students were found to make more accurate task-selection choices for fictitious peers in scenarios about the peer's performance and effort investment on math problems that had a different number of steps and came from a task database with a different structure. A potential explanation for this discrepancy is that the cognitive load experienced during the self-regulated learning phase in math could have left students with insufficient cognitive resources to simultaneously think about what they had learned during the training and how that would translate to these new

tasks (cf. Van Gog et al., 2011). That is, selecting a task for a fictitious student, based on information that is already given, is presumably much less cognitively demanding than having to keep in mind and adapt self-assessment and task-selection rules while also working on novel and difficult problems (which can be seen as a secondary task).

In conclusion, although prior findings showed that example-based learning of self-assessment and task-selection skills can be an effective and relatively easy to implement method for improving students' self-regulated learning outcomes, secondary school students might not be able to apply these skills when they are engaging in self-regulated learning in a different domain. Because it is rare for spontaneous transfer to occur, especially under conditions of high cognitive load, more explicit instruction might be necessary for task-selection skills to transfer from domain to domain (Salomon & Perkins, 1989). For instance, maybe students would need explicit prompts during self-regulated learning in math, instructing them to think back about what they learned about self-assessment and task selection in the context of biology. How to scaffold the transfer of self-regulated learning skills remains an important question for future research.

ORCID

Steven F. Raaijmakers http://orcid.org/0000-0002-8392-8933

REFERENCES

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.

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.

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.

Bransford, J. D., Brown, A. L., & Cocking, R. (2000). *How people learn. Brain, mind experience and school*. Washington: Academic Press.

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

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.

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, United Kingdom: 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. (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.

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., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, *28*, 803–830. https://doi.org/10.1007/s10648-015-9350-2.

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.

Polman, E. (2012). Self–other decision making and loss aversion. *Organizational Behavior and Human Decision Processes*, *119*, 141–150. https://doi.org/10.1016/j.obhdp.2012.06.005.

Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., Van Merriënboer, J. J. G., & Van Gog, T. (2017). Training self-regulated learning skills with video modeling examples: Do task-selection skills transfer? *Instructional Science*, 1–18. https://doi:10.1007/s11251-017-9434-0.

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: Volume 2* (pp. 169–182). Orlando, FL: US Army Research.

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.

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.

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of concurrent monitoring on cognitive load and performance as a function of task complexity. *Applied Cognitive Psychology*, *25*, 584–587. https://doi.org/10.1002/acp.1726.

Van Merriënboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*, 147–177. https://doi.org/10.1007/s10648-005-3951-0.

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

APPENDIX 1: EXAMPLE OF PROBLEM-SOLVING TASK (THIRD LEVEL OF COMPLEXITY WITH HIGH SUPPORT)

Level 3 – High support

Find the equation for the given red line.

A formula for a linear equation is written as: $y = ax + b$

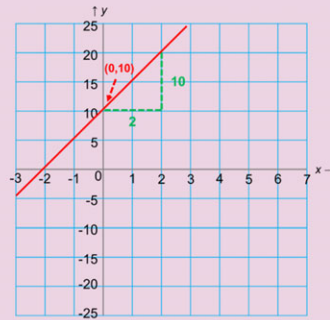where a is the slope of the line which is how much the line rises or falls when you go one unit to the right

and where b is the intercept which is the value of y when x is 0.

1. Find a by checking how much the line rises or falls when you go one unit to the the right on the x-axis.
By choosing two points that are easy to read off the graph, you see that the red line rises **10** units when you go **2** units to the right on the x-axis. With each unit to the right the line rises 5 units. Thus, the slope (a) is 5.

2. Determine b by reading from the graph where the line cuts the y-axis.
The red line cuts the y-axis at the point **(0,10)**.
Thus, the intercept (b) is 10.

3. Finally, fill in the formula with a and b.

Answer

$y = $ . . . .