Received: 22 December 2023 | Revised: 6 March 2024 | Accepted: 15 April 2024
DOI: 10.1002/acp.4200

SHORT PAPER

WILEY

Training task-selection skills: The effect of prompts and explicit instruction on transfer

Jane Pieplenbosch | Gesa van den Broek | Vincent Hoogerheide | Tamara van Gog

Department of Education, Utrecht University, Utrecht, The Netherlands

Correspondence
Jane Pieplenbosch, Department of Education, Utrecht University, P.O. Box 80140, 3508TC Utrecht, The Netherlands.
Email: j.m.pieplenbosch@uu.nl

Funding information
Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: #VI.C.201.079

Abstract

For effective self-regulated learning with problem-solving tasks, students must accurately assess their performance and select a suitable next learning task. However, most students struggle with this. Recent research shows that self-assessment and task-selection skills can be trained through video modeling examples (SATS-training). However, the limited research available suggests that students struggle to transfer trained task-selection skills to other problem-solving contexts. We investigated whether guidance in the form of prompts (stating that the task-selection procedure can be adapted and used) or explicit instruction (on how the procedure can be adapted) would improve task-selection accuracy on transfer tasks with this guidance available and on later, unguided transfer tasks. Explicit instruction significantly enhanced task-selection accuracy compared to prompts and a no-guidance control condition on guided transfer tasks, but not on unguided transfer tasks. Thus, it remains a question how to lastingly improve transfer of task-selection skills also in the absence of guidance.

KEYWORDS
example-based learning, self-regulated learning, STEM, task-selection, transfer

Problem-solving tasks form the backbone of STEM curricula (Science-Technology-Engineering-Math). Students' ability to self-regulate their learning with problem-solving tasks is important as learners generally practice during self-study sessions at school or at home in the absence of teachers. During these (online) self-study sessions, learners have to decide which tasks to practice, when they need more instructional support, and when they master tasks well enough to quit practicing or move on to more difficult tasks. This requires learners to accurately monitor (i.e., self-assess) their performance after completing a task, and to use this information to control (or regulate) subsequent learning activities by selecting a fitting new task (i.e., *task-selection*). For effective task-selection, learners not only have to consider their current performance level but also the task characteristics: Selecting tasks at an appropriate level of complexity and deciding how much (built-in) support they need (e.g., [partially] worked examples/hints/feedback, depending on the learning environment) given their current performance (van Gog et al., 2020). For example, a student who has performed well on a task can move on to a task that is somewhat more complex or has less built-in support (within their zone of proximal development; Vygotsky & Cole, 1978). However, learners often select tasks that are not adapted to their current level of performance, which leads them to practice with tasks that are either too easy or too complex, and ultimately hampers their learning (van Gog et al., 2020). Therefore, it is necessary to investigate how task-selection skills can be improved.

1 | TRAINING TASK-SELECTION SKILLS

Earlier studies have shown that task-selection skills can be trained through video modeling examples (Kostons et al., 2012; Raaijmakers, Baars, Schaap, et al., 2018). Moreover, such training prior to a self-regulated learning phase also improved the domain-specific learning outcomes. Participants in the training condition received a Self-Assessment and Task-Selection (*SATS*) training, in which they observed four video examples of models solving a five-step hereditary biology problem; rating how much mental effort they invested in solving the problem; self-assessing performance (assigning points for correctly performed steps); and using the self-assessed performance and invested mental effort rating to select a new task from a database, using a simple algorithm that specified the column from which to select a new task. The database included tasks with varying levels of complexity and instructional support.

Participants who received the SATS-training before engaging in a self-regulated learning phase, in which they had to choose eight biology tasks from the database, made more accurate task selections during this phase and showed better problem-solving performance on the post-test than participants who did not receive training (Kostons et al., 2012). In a subsequent study, these benefits of the training on problem-solving post-test performance were replicated and extended (Raaijmakers, Baars, Schaap, et al., 2018): Next to the original training with the simple algorithm (algorithm condition), a training condition was included in which participants were taught a more general heuristic (i.e., a rule-of-thumb; e.g., When performance is high and effort is low, you can move to a more complex task or a task with less built-in support in the task database). Results on the post-test showed that participants in both conditions outperformed the control condition that did not receive any SATS-training.

The same applied to transfer: compared to untrained students, students who received the SATS training were somewhat better at selecting tasks based on the performance score and mental effort rating of a fictitious person described in transfer vignettes, even though these had a different task context.[1] Raaijmakers, Baars, Schaap, et al. (2018) expected that the task-selection procedure in the heuristic condition would be easier to transfer because it was described as a rule-of-thumb, but there was no statistically significant difference between the two training conditions in transfer task-selection accuracy. While these initial effects were promising, there was room for further improvement in task-selection accuracy on the transfer vignettes. Moreover, in a later study, Raaijmakers, Baars, Paas, et al. (2018) found that students who received SATS training in the biology context and then engaged in a self-regulated learning phase with math problems that had a different number of steps and a different task database, did not outperform untrained students on the math post-test, suggesting they could not transfer the learned task-selection skills to this new context.

It is important to investigate how transfer of task-selection skills can be improved, because it would not be feasible to train specific task-selection skills for all types of problem-solving tasks learners encounter. Transfer occurs when experience and knowledge learned in one problem context is used to solve a problem in a new context (Mayer & Wittrock, 1996).

To enhance the effectiveness of the training for the transfer of task-selection skills, we first need to understand why learners struggle to transfer the trained task-selection skills. For transfer to occur, learners have to be able to (1) *recall* the trained procedure, (2) *recognize* that they can apply the principles from this procedure to the new task context, and (3) if needed, *adapt* the trained procedure to apply it to the new task context (Barnett & Ceci, 2002; Mayer & Wittrock, 1996). In the present study, we examine how and how well students recall the trained procedure, and whether offering students additional support with recognizing that the procedure applies or explicit instruction on how to adapt it, would help them to transfer task-selection skills to new problem-solving contexts.

2 | THE PRESENT STUDY

The present study tested two strategies for supporting transfer of problem-solving task-selection skills as trained via the SATS training (cf. Kostons et al., 2012; Raaijmakers, Baars, Schaap, et al., 2018; Raaijmakers, Baars, Paas, et al., 2018): *Prompting*, which can support students' recognition of whether the learned procedure can be applied and should be adapted, and *explicit instruction*, which can support students in recognition plus adaptation of the learned procedure, by explaining how to adapt. Prompts do not contain information that is necessary to complete the task but aim to remind the learner to use a certain strategy or procedure (Berthold et al., 2007). While the effect of prompting has not yet been investigated in the context of transfer of task-selection skills, some studies focused on prompts to enhance the transfer of self-regulated learning strategies with some finding positive results (Bannert et al., 2015; Müller & Seufert, 2018), while others showed no effect (Engelmann & Bannert, 2021). Likewise, explicit instruction, in which a strategy or procedure is fully explained to the learner (Clark et al., 2012), has not yet been studied specifically in the context of the transfer of task-selection skills, but other research suggests it can enhance the transfer of self-regulated learning strategies (Salomon & Perkins, 1989; Schuster et al., 2020).

We first explored how and how well participants recall the trained task-selection procedure, which is a necessary first step for transfer to occur. To rule out that the problem might lie in how well participants remembered the task-selection procedure (although the manipulation check conducted by Raaijmakers, Baars, Paas, et al., 2018, suggests this is not the issue), we administered a recall task. This also allowed us to check whether the conditions differed on how well (i.e., proportion of correctly recalled idea units) and in what manner participants recalled the trained task-selection procedure (i.e., with all the details of the algorithm, as a more general heuristic, or both), despite the random assignment. The latter is also interesting because Raaijmakers, Baars, Schaap, et al. (2018) hypothesized (but did not test) that a potential explanation for the lack of difference in task-selection transfer accuracy between their two training conditions might have been that participants in the algorithm condition inferred the underlying heuristic and remembered it as such; the present study sheds light on how participants remember the procedure.

Importantly, the present study focuses on the effect of prompts and explicit instruction on task-selection accuracy on both the transfer tasks for which prompts / explicit instruction were provided (*guided transfer tasks*) and on subsequent transfer tasks when learners no longer received prompts or explicit instruction (*unguided transfer tasks*). Establishing these effects both in the presence and absence of the guidance, allows for determining whether prompts and explicit instruction can successfully *support* transfer of task-selection skills (higher task-selection accuracy on guided tasks only) and also help students *learn to* recognize and adapt the procedure when guidance is removed (higher accuracy also on unguided tasks).

We hypothesized that participants who received explicit instruction (i.e., support for recognizing and adapting, by explaining how to adapt the learned procedure) would make more accurate task-selections than participants who received prompts (i.e., support for recognizing, informing participants that the procedure can be applied in adapted form, but not explaining how to adapt), and that both would make more accurate task-selections than participants who received no guidance (control condition) during guided transfer tasks and that this benefit would remain during subsequent unguided transfer tasks. Additionally, to gain insight into the efficiency of the conditions, we explored how much mental effort participants invested in selecting tasks (van Gog & Paas, 2008).

3 | METHOD

3.1 | Participants and design

A total of 179 Bachelor students (18–21 years old) were recruited via Prolific (https://www.prolific.com) to take part in a 45-min online experiment for a payment of £6.75. Only students who were fluent in English and had English as their first language could participate. The study had a mixed 3×2 factorial design with type of guidance for transfer (no guidance, prompts, explicit instruction) as between-subjects factor and the presence of guidance (with/without) as within-subjects factor. A power analysis in MorePower (Campbell & Thompson, 2012), for a mixed model ANOVA (within-between interaction) with an alpha of .05, a medium effect size ($\eta^2_p = .06$), and a power of .80, resulted in a required sample size of 162; we oversampled by 10% to account for possible dropout. Participants were randomly assigned to one of the three conditions. We had to exclude the data of eight participants because they did not pass the attention checks and of 11 participants because they reported that they had used strategies during the experiment that were not allowed (e.g., take notes, look up information online, take a photo or screenshot, or make a video/audio recording). Of the remaining 160 participants (age: $M = 20.13$, $SD = 0.90$; gender: 73 female, 80 male, 5 non-binary, 1 genderfluid, 1 "I'd rather not say"), 57 were in the control condition, 55 in the prompts, and 48 in the explicit instruction condition.

3.2 | Materials

SATS training. The SATS training had the same content as in Raaijmakers, Baars, Schaap, et al. (2018); Raaijmakers, Baars, Paas, et al. (2018), translated into English. All materials were presented in Qualtrics survey software (http://www.qualtrics.com). The training presented four video modeling examples in which the models (male or female; see Table 1) explained and demonstrated a procedure for self-assessment and task-selection in the context of biology (heredity) problem-solving tasks. The problems could be solved in five steps and were organized in a task database with five complexity levels, three support levels within each complexity level (high support, low support, no support), and five isomorphic problems (structurally similar, but different cover stories) at each support level (i.e., 75 tasks in total, see Figure 1). In each video example, the model (partially) solved a problem (with no support, at the first or second level of complexity; see Table 1), rated their invested mental effort on a 9-point scale (cf. Paas, 1992), self-assessed performance (assigning 1 point for every correct step out of five steps in total: 0–5), and selected a new task from the task database. To select a task at the appropriate level of complexity and support, the models used a table that combined their self-assessed performance and mental effort according to a simple algorithm, which led to an advised "step-size" (i.e., number of columns to the left [negative numbers] or right; see Figure 2). The model first gave a general explanation of the simple algorithm (italicized here) before using the scores on the present problem to select a task: "*In this table, I can see what a suitable next task would be based on my performance and effort. Vertical is performance, horizontal is effort. The table gives me advice for the next task. If I had a high score for performance and a low score for effort, then I am ready for a more difficult task, as you can see in the table. With a high performance and low effort I end up at +2. That means two steps to the right in the task database. Then I get a more difficult task or a task with less support. With a low performance and high effort you end up −2. So, I would have to do an easier task or a task in which more steps have been worked out already. I gave myself an 8 for effort and a 3 for performance. Then I end up at −1 in the table.*"

Cued Recall Test. After the SATStraining, participants were asked three cued recall questions about the trained task-selection procedure. The first two were open questions: (1) "You just watched videos in which students solved a biology problem and then selected a suitable next task to work on from a task database. To be able to select a new task to work on from the task database, the students had to rate two things. Which things did they rate?"; (2) "Explain how the students used these ratings (from the previous question) to select a new task from the task database." The third question asked them to fill in an empty version of the step-size table, including the labels and scales for performance and effort, and the numbers indicating the step sizes (i.e., 17 labels/cells in total).

Isomorphic Vignettes. To check whether participants could apply the trained task-selection procedure, they received seven vignettes, describing a short scenario about their fictitious performance and mental effort on problem-solving tasks identical to the tasks used in

TABLE 1 Features of the four video modeling examples from the SATS training.

Example	Model	Complexity level/Support level	Number of correct steps	Effort	Task-selection step-size
1	Female	Level 1/no support	5	2	+2
2	Male	Level 1/no support	5	5	+1
3	Female	Level 2/no support	4	7	0
4	Male	Level 2/no support	3	8	−1

Complexity level	*Complexity 1* - 2 generations - 1 unknown - 1 solution - deductive			*Complexity 2* - 2 generations - 1 unknown - multiple solutions - deductive			*Complexity 3* - 2 generations - 1 unknown - multiple solutions - inductive			*Complexity 4* - 3 generations - 1 unknown - multiple solutions - inductive and deductive			*Complexity 5* - 2 generations - 1 unknown - multiple solutions - deductive		
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	High support	Low support	No support
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	Hunting-ton	Wolman syndrome	Apert syndrome	Flower color	Hunting-ton	Fruit flies	Tail length	Tongue rolling	Albinism	Fruit flies	Cleft lip	Milk allergy	Tail length	Apple tree
	Apple tree	Chicken beak	Tail length	Fruit flies	Apple tree	Albinism	Chicken beak	P.R.A. disease	Depression	Cat fur	Flower color	P.R.A. disease	Apert syndrome	Wolman syndrome	Fruit flies
	Depression	Cleft lip	Milk allergy	P.R.A. disease	Tongue rolling	Mammary carcinoma	Wolman syndrome	Milk allergy	Apert syndrome	Depression	Tongue rolling	Mammary carcinoma	Hunting-ton	Flower color	Chicken beak

FIGURE 1 The task database from the SATS training and isomorphic vignettes. It has five complexity levels, three support levels within each complexity (high support, low support, no support), and five isomorphic problems (structurally similar, but different cover stories) at each support level (i.e., 75 tasks in total).

the training, and asking them to select a new task from the database identical to the one used in the training. An example of a vignette is: "Suppose you completed a 5-step biology problem at complexity level 1 with no support. You got 1 step correct. You rated your effort as 8 on a scale of 1 to 9. Based on this description and the procedure from the videos, what is a suitable task to work on next?" For each vignette, participants selected a task by clicking on that task in the depicted task database. They did not actually get to work on that task or get a corresponding vignette.

Transfer Vignettes. To measure if participants could apply the trained task-selection procedure in novel contexts (i.e., transfer), participants worked on vignettes that described a problem-solving task that differed from the training (i.e., a different number of solving steps and/or complexity and support levels in the task database). Participants received two blocks of five[2] transfer vignettes. In the guided block, the students in the respective conditions received prompting and explicit instruction (see Appendix A for details). In the unguided block students no longer received any support. We used two sets of transfer vignettes (A and B; see Table 2), of which students received one in the guided and one in the unguided block; the order was

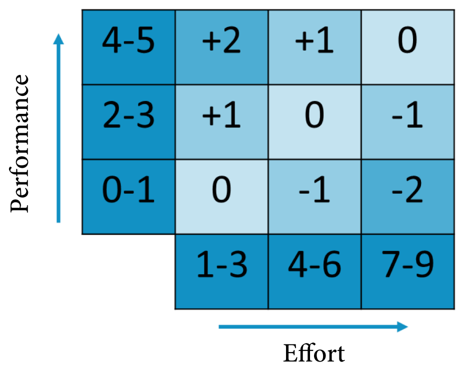


FIGURE 2 The step-size table with the simple algorithm that was used in the video modeling examples to determine the advised next task by combining the self-assessed performance and mental effort (i.e., the step-size shows the number of columns to move to the left [negative numbers] or right in the task database): with high performance and low effort, a task with less support or more complexity is advised; with high performance and high effort, it is advised to practice a similar task again; and with low performance and high effort, to select a task that offers more support or is less complex.

TABLE 2 The number of problem-solving steps, complexity levels, and levels of support in the task database in the isomorphic and transfer vignettes (Sets 1 and 2).

	Number of problem-solving steps	Number of complexity levels	Number of support levels
Isomorphic vignettes	5	5	3
Transfer vignettes Set 1			
Vignette 1	5	4	2
Vignette 2	7	5	3
Vignette 3	8	5	3
Vignette 4	7	4	2
Vignette 5	8	4	2
Transfer vignettes Set 2			
Vignette 1	5	3	2
Vignette 2	11	5	3
Vignette 3	9	5	3
Vignette 4	11	3	2
Vignette 5	9	3	2

counterbalanced between participants. The vignettes in the two sets differed in terms of the problem-solving steps used (see Table 2), but were constructed in such a way that the manner in which the task-selection procedure had to be adapted and the consequences for the to-be-selected tasks were the same (i.e., on the transfer vignettes 3 and 5 in both transfer task sets, multiple adjustments to the algorithm were possible and could be considered correct, but the performance and effort scores in the vignettes were chosen such that there was no ambiguity regarding step-size; see Appendix A; example "Explicit instruction").

Prompts and Explicit Instruction. In the guided transfer block, students in the prompts condition received a prompt with each vignette. The prompts explicated the difference between the transfer task described in the vignette and the tasks from the training videos and stated that the step-size table from the videos for choosing a new task could still be used and, in some cases, had to be adapted (see Appendix A; example "Prompt"). Students in the explicit instruction condition received the same prompt plus an explanation of how the procedure could be adapted to the scenario described in the vignette (see Appendix A; example "Explicit instruction"). Students in the control condition did not receive any guidance in the guided block. In the unguided block, none of the participants received guidance.

Mental Effort Ratings. Each vignette in the isomorphic block and transfer blocks, was followed by the question: "How much effort did you invest in selecting a suitable next task?" on a 9-point scale ranging from 1 (very, very low effort) to 9 (very, very high effort) (Paas, 1992).

3.3 | Procedure

Participants were recruited on Prolific (https://www.prolific.com), through which they accessed the Qualtrics survey (https://www.qualtrics.com), where they would first find an information letter and an informed consent form. After providing consent, participants were asked demographic questions: their age and gender, if they had attention deficit disorder or dyslexia diagnoses, and student status (as an extra check on the bachelor student filter we had set in Prolific). Then, they continued to the experiment, which took ~45 min. First, they received general instruction and were asked to not make any notes during the videos. Then, participants watched the SATS training videos, completed the cued recall test, and worked on the block of seven isomorphic task-selection vignettes. Subsequently, in the guided transfer block, participants were randomly assigned to one of the three conditions and worked on five task-selection transfer vignettes during which they received no support, received prompts, or received explicit instruction, depending on their assigned condition. This was followed by the unguided task-selection transfer block. After each isomorphic and transfer vignette they rated their invested mental effort. We also inserted an attention check in each block, which superficially looked like a vignette but actually instructed participants to select a specific task from the task database (so that we could tell if participants were selecting tasks without reading the vignettes properly). At the end of the experiment, participants were asked "Did you use any of the following strategies during this experiment? Take notes, look up information online, take photo or screenshot, make video/audio recording. Please answer honestly; this is not a reason to reject your submission, but it is important for us to know to check whether participants complied with our instruction."

3.4 | Data analysis

To determine how well participants remembered the task-selection procedure, the first open-answer cued recall question was scored by assigning one point per correct answer. For the second open-answer cued recall question we first divided the explanation of the task-selection procedure into eight idea units, and then scored the number of idea units in participants' answers. The scores for the two open-answer cued recall questions were added into one score with a maximum

TABLE 3 Descriptive statistics of performance scores on cued recall questions for participants who recalled the task-selection procedure as algorithm, heuristic, both, or from whose answer it was unclear how they recalled it.

	Algorithm ($n = 10$)		Heuristic ($n = 68$)		Both ($n = 52$)		Unclear ($n = 30$)	
	M (SD)	Mdn	M (SD)	Mdn	M (SD)	Mdn	M (SD)	Mdn
Open cued recall (range: 0–10)	4.20 (1.48)	4.0	5.06 (1.44)	5.0	5.62 (1.55)	5.0	2.40 (1.40)	2.5
Fully correct	$n = 0$		$n = 0$		$n = 0$		$n = 0$	
Stepsize table cued recall (range: 0–17)	10.00 (6.94)	13.0	9.72 (6.48)	11.0	11.00 (5.61)	13.0	3.83 (5.41)	0.0
Fully correct	$n = 2$		$n = 16$		$n = 9$		$n = 1$	

score of 10 (see coding scheme in Appendix B). The data was independently coded by two coders and Cohen's k was used to determine inter-reliability (Cohen, 1960). There was a good agreement between the two coders, $\kappa = .80$. To determine how (in what manner) participants remembered the procedure, the answers were further classified as falling into one of four categories: with details of the algorithm, as a more general heuristic, both, or neither. There was also a good agreement between coders for the coding of these categories, $\kappa = .84$. To determine how many specifics of the algorithm participants recalled, we scored how they filled in the empty version of the step-size table, assigning one point per correctly filled in label/cell, resulting in a maximum score of 17.

To establish participants' task-selection accuracy on the isomorphic vignettes, we computed the absolute (unsigned) difference between the chosen and the recommended "step-size" (i.e., number of columns to the left/right in task database) according to the trained task-selection algorithm (cf. Raaijmakers, Baars, Paas, et al., 2018). The task-selection accuracy on the transfer vignettes was calculated in the same manner, but for the transfer vignettes describing a different number of problem-solving steps, the algorithm was adjusted (cf. Raaijmakers, Baars, Schaap, et al., 2018; see Appendix A example "Explicit instruction"). The data are available on the Open Science Framework: osf.io/x92fu.

4 | RESULTS

For all measures, Shapiro-Wilk's tests showed that the assumption of normality was violated. Therefore, the data were analyzed with non-parametric tests. Kruskal-Wallis tests were used to test for differences among conditions and the Wilcoxon rank sum test to test for differences in task-selection accuracy on the guided and unguided transfer blocks within conditions. Dunn's test was used as post hoc test, which applies a Bonferroni correction. Below we report the adjusted *p*-values. The Pearson correlation (*r*) is reported as effect size with values .10, .30, and .50, respectively interpreted as small, medium, and large effect sizes (Cohen, 1988).

4.1 | How and how well do participants recall the task-selection procedure?

Table 3 shows the recall data. On average, participants recalled 4.69 of the 10 idea units, and there were no statistically significant[3] differences among conditions, $\chi^2(2) = .82$, $p = .665$, $r = .09$. As for how they remembered the procedure, out of the 160 participants, only 6.2% ($n = 10$) described the task-selection procedure solely in algorithmic terms, 42.5% ($n = 68$) solely as a heuristic, 32.5% ($n = 52$) as both, and in 18.8% ($n = 30$) it was unclear from their answer in what manner they recalled the task-selection procedure. Of the 52 participants who described the procedure both in algorithmic and heuristic terms, 38.5% described the "input" as algorithm (e.g., "If they rated their performance a five and their mental effort a one…"), 50.0% as heuristic (e.g., "If they had high performance and low mental effort…"), and 90.4% described the "output" as algorithm (e.g., "then they would end up at +2 in the table"), and 78.8% as heuristic (e.g., "then they would move to a more difficult task").

Surprisingly, in the group that described the task-selection procedure solely in algorithmic terms, only 20% ($n = 2$) filled in the step-size table correctly; in the heuristic group this was 23.5% ($n = 16$); in the group that described both 17.3% ($n = 9$); and in the group that described neither 3.3% ($n = 1$). There were no statistically significant differences among conditions in cued recall of the step-size table, $\chi^2(2) = 3.36$, $p = .186$, $r = .09$, nor in task-selection accuracy on the isomorphic vignettes (Table 4), $\chi^2(2) = 0.98$, $p = .614$, $r = .09$. On average, participants were quite accurate with an overall mean of 0.81 on the isomorphic tasks. Given that none of the analyses of recall measures showed statistically significant differences among conditions, any differences in scores on the transfer vignettes are unlikely to result from a priori differences among conditions in recall of the trained procedure.

4.2 | Do prompts and explicit instruction Foster task-selection accuracy?

Table 4 shows the task-selection accuracy on the transfer vignettes per condition. On average, participants were quite accurate (overall mean: 0.64 on guided and 0.77 on unguided transfer vignettes). There was a statistically significant difference in the task-selection accuracy scores on the guided transfer vignettes among conditions, $\chi^2(2) = 23.83$, $p < .001$, $r = .37$. In line with our hypothesis, the post hoc tests revealed that participants in the explicit instruction condition showed statistically significant higher task-selection accuracy than those in the prompts condition, $z = 4.87$, $p < .001$ (Bonferroni adjusted *p*-value), $r = .39$, and in the control condition, $z = 2.87$, $p = .006$, $r = .23$ (i.e., lower scores indicate higher task-selection accuracy). However, contrary to our hypothesis, participants in the prompts condition did not show statistically significantly higher task-selection accuracy than those in the control condition, $z = -2.11$, $p = .052$, $r = -.17$.

TABLE 4 Descriptive statistics of performance on cued recall questions, and task-selection accuracy on the isomorphic vignettes, guided transfer vignettes, and unguided transfer vignettes, per condition.

	Control (n = 57)		Prompts (n = 55)		Explicit instruction (n = 48)	
	M (SD)	Mdn	M (SD)	Mdn	M (SD)	Mdn
Open cued recall (range: 0–10)	4.46 (2.04)	4.0	4.84 (1.75)	4.0	4.79 (1.79)	5.0
Stepsize table cued recall (range: 0–17)	8.68 (7.07)	11.0	8.29 (6.26)	9.0	10.35 (6.05)	12.0
TS accuracy[a]						
Isomorphic vignettes (range: 0–9)	0.77 (1.30)	0.3	0.80 (1.09)	0.4	0.87 (1.09)	0.4
Fully correct TS	n = 18		n = 11		n = 15	
Guided transfer	0.51 (0.36)	0.4	0.89 (0.82)	0.8	0.50 (0.90)	0.0
Fully correct TS	n = 6		n = 6		n = 28	
Unguided transfer	0.65 (0.46)	0.6	0.96 (0.82)	0.6	0.71 (0.71)	0.4
Fully correct TS	n = 5		n = 6		n = 8	
Mental effort						
Isomorphic vignettes (range: 0–9)	4.42 (1.74)	4.4	4.61 (1.60)	5.0	5.05 (1.58)	5.0
Guided transfer (range: 0–9)	4.59 (1.65)	4.6	5.13 (1.59)	5.2	5.50 (1.74)	5.7
Unguided transfer (range: 0–9)	4.28 (1.75)	4.2	4.67 (1.68)	4.8	5.09 (1.63)	4.9

[a]Lower = better; 0 = fully accurate; TS = task-selection.

On the unguided transfer vignettes, there was no longer a statistically significant difference in task-selection accuracy among conditions, $\chi^2(2) = 4.38$, $p = .112$, $r = .12$. The within-subjects analyses showed no statistically significant differences in task-selection accuracy on the guided and unguided transfer tasks in the control group, $z = 1351.50$, $p = .115$, $r = .15$, and the prompts group, $z = 1426.50$, $p = .607$, $r = .05$. Participants in the explicit instruction condition showed lower accuracy on the unguided than on the guided transfer tasks, $z = 701.50$, $p < .001$, $r = .35$.

4.3 | Do mental effort ratings differ among conditions?

There was no statistically significant difference among conditions in invested mental effort on the isomorphic vignettes, $\chi^2(2) = 3.03$, $p = .220$, $r = .08$. Effort investment on the guided transfer vignettes did differ among conditions, $\chi^2(2) = 7.47$, $p = .024$, $r = .19$. Post hoc tests indicated that participants in the explicit instruction condition reported having invested more effort than participants in the control condition, $z = -2.64$, $p = .013$, $r = -.21$, whereas effort investment did not differ between participants in the prompts condition and those in the explicit instruction, $z = -0.81$, $p = .624$, $r = .06$, or control condition, $z = -1.88$, $p = .090$, $r = -.15$. On the unguided transfer vignettes, there were no statistically significant differences among conditions, $\chi^2(2) = 5.67$, $p = .059$, $r = .15$.

5 | DISCUSSION

Given the increasing emphasis on effective self-regulated learning in education, a substantial number of studies have investigated how to improve students' self-monitoring and self-regulation skills. However, the important question of whether those skills transfer to other tasks and domains is seldom addressed. Recent research in the context of problem-solving tasks that addressed whether trained task-selection skills would transfer, found mixed results (Raaijmakers, Baars, Paas, et al., 2018; Raaijmakers, Baars, Schaap, et al., 2018). Therefore, the present study tested two strategies for supporting transfer of trained task-selection skills: Prompting, which can support students' recognition of whether the learned procedure can be applied and whether adaptations are needed, and explicit instruction, which can support students' recognition and explains what adaptations are needed.

5.1 | How (well) do participants recall the task-selection procedure?

The open question cued recall data showed that few participants described the task-selection procedure only algorithmically; the majority recalled it as a heuristic or as both. Note that these results should be interpreted with caution, as the answers varied in quality and length. The results regarding the step-size table cued recall question indicated that a relatively small number of participants could recall all the elements of the step-size table correctly. Yet, the descriptive statistics revealed that overall, participants showed quite accurate task-selection on the isomorphic vignettes (deviating on average only 0.81 steps, indicating that participants could apply the trained task-selection procedure relatively well). These findings suggest that participants, as Raaijmakers, Baars, Schaap, et al. (2018) suggested, inferred the heuristic behind the simple algorithm, and that this was sufficient to achieve relatively high task-selection accuracy on the isomorphic vignettes.

5.2 | Do prompts and explicit instruction Foster task-selection accuracy?

Contrary to our hypothesis, participants in the prompts condition did not show more accurate task-selection than those in the control condition. Possibly, learners did not recall the procedure well enough, or struggled to adapt the trained task-selection procedure to the new problem-solving context and therefore did not benefit from prompts only. A study by van Peppen et al. (2022) on the transfer of other higher-order skills found similar results regarding prompting.

As predicted, participants who received explicit instruction on how they could adapt the procedure, showed higher task-selection accuracy on the guided transfer vignettes than those in the other conditions. Interestingly, however, they were still not fully accurate; only 28 out of 48 participants in the explicit instruction condition consistently made correct task-selections (Table 4). Moreover, this higher accuracy was not retained when guidance was removed: On the unguided transfer tasks, accuracy of the explicit condition dropped (with the number of participants consistently making correct task-selections decreasing from 28 to 8), and there were no longer any significant differences among the conditions. The higher accuracy of participants in the explicit instruction condition was accompanied by higher levels of invested mental effort than the control condition on the guided transfer tasks (but not on the unguided tasks). This could potentially be explained by the extra reading needed for the explicit instruction.

In sum, our results suggest that explicit instruction foster task-selection accuracy, but only as long as they are present. That learners seem to struggle to adapt the procedure themselves on new (unguided) transfer vignettes, demonstrates how persistent the problem of how to foster transfer is. However, it should be noted that on average, task-selection on the transfer vignettes was already quite accurate (also compared to the studies by Raaijmakers, Baars, Schaap, et al., 2018; Raaijmakers, Baars, Paas, et al., 2018).

5.3 | Limitations and suggestions for future research

A limitation of this study is that participants did not work on the tasks themselves, but had to apply the trained task-selection procedure learned from the training on vignettes. It is unclear whether the same results would occur if participants worked on the tasks themselves. However, the use of vignettes also had benefits: It allowed us to systematically vary characteristics of the problem-solving context (number of steps, layout of the task database) without confronting learners with many different types of problems to solve and without the high cognitive load involved in first attempting to solve those problems (i.e., they could fully focus on taskselection). Moreover, the vignettes ensured that the values for performance and effort were the same for all participants, making it easier to compare the accuracy among conditions.

An important avenue for future research is to uncover strategies that lead to sustained transfer benefits. In our study, learners in the explicit instruction condition went from full recognition and adaptation support, to no support at all in unguided transfer vignettes. They might benefit more from fading support, as this has been proven to foster learning (Reisslein et al., 2007; Renkl et al., 2002). Specifically, combining fading and prompts to encourage learners to find the underlying principle of the instruction has been proven to improve performance on near and far transfer of problem-solving skills (Atkinson et al., 2003); the question is, however, if the same would apply to task-selection skills.

6 | CONCLUSION

To conclude, although explicit instruction improved transfer of the trained task-selection procedure, leading to more accurate task-selection on the vignettes where it was provided, it did not help learners to select the right tasks on subsequent, unguided transfer vignettes. These results illustrate the persistent problem of transfer. Given that it would be impossible to train learners' task-selection skills for each type of problem-solving task they encounter in their studies, it is important to continue to investigate whether and how transfer of task-selection skills can be fostered, especially in the absence of guidance, in future studies.

ACKNOWLEDGMENTS

We would like to thank Steven Raaijmakers for sharing his Dutch SATS training materials; Florence Lucas for co-creating the training materials and assistance with data analysis; Tonja Choupenko, Aafke Gruintjes, and Sacha McPherson, who wrote their bachelor thesis on this dataset; Barbara Flunger for statistical advice; and the Department of Education pubgroup for feedback and suggestions on a draft of this manuscript.

FUNDING INFORMATION

This research was funded by a NWO Vici grant (#VI.C.201.079) awarded to Tamara van Gog.

CONFLICT OF INTEREST STATEMENT

We have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

The dataset and analyses scripts are stored on the website for Open Science Framework website (osf.io) under a CC BY-NC license for this project: osf.io/x92fu.

ORCID

Jane Pieplenbosch https://orcid.org/0009-0003-7349-315X
Gesa van den Broek https://orcid.org/0000-0001-7624-0957
Vincent Hoogerheide https://orcid.org/0000-0003-4176-0973
Tamara van Gog https://orcid.org/0000-0003-3766-6255

ENDNOTES

[1] That is, the tasks described in these vignettes were in a different domain (math instead of the trained biology tasks), differed in the number of problem-solving steps (eight instead of five) and/or in the task database layout (with four instead of five complexity levels, two instead of three support levels, and four instead of five isomorphic tasks per support level).

[2] Two additional vignettes were added to the unguided transfer tasks for pilot purposes outside the scope of this study; these data will not be reported here.

[3] With statistically significant we mean it met the cut-off of statistical significance ($p < .5$). Note, however, that there is some debate about using the term "statistically significant" and the .05 threshold (Wasserstein et al., 2019).

[4] The italics and underlining were not shown to participants; they received the information according to their assigned condition as plain text.

REFERENCES

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology*, *95*(4), 774–783. https://doi.org/10.1037/0022-0663.95.4.774

Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior*, *52*, 293–306. https://doi.org/10.1016/j.chb.2015.05.038

Barnett, S., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*, *17*(5), 564–577.

Campbell, J. I., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*, 1255–1265. https://doi.org/10.3758/s13428-012-0186-0

Clark, R., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning: The case for fully guided instruction. *American Educator*, *36*(1), 5–11.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/00131644600200

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

Engelmann, K., & Bannert, M. (2021). Analyzing temporal data for understanding the learning process induced by metacognitive prompts. *Learning and Instruction*, *72*, 1–11. https://doi.org/10.1016/j.learninstruc.2019.05.002

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*(2), 121–132. https://doi.org/10.1016/j.learninstruc.2011.08.004

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In *Handbook of educational psychology* (pp. 47–62). Routledge.

Müller, N. M., & Seufert, T. (2018). Effects of self-regulation prompts in hypermedia learning on learning performance and self-efficacy. *Learning and Instruction*, *58*, 1–11. https://doi.org/10.1016/j.learninstruc.2018.04.011

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Raaijmakers, S. F., Baars, M., Paas, F., van Merriënboer, J. J., & Van Gog, T. (2018). Training self-assessment and task-selection skills to foster self-regulated learning: Do trained skills transfer across domains? *Applied Cognitive Psychology*, *32*(2), 270–277. https://doi.org/10.1002/acp.3392

Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., van Merriënboer, J., & van Gog, T. (2018). Training self-regulated learning skills with video modeling examples: Do task-selection skills transfer? *Instructional Science*, *46*(2), 273–290. https://doi.org/10.1007/s11251-017-9434-0

Reisslein, J., Sullivan, H., & Reisslein, M. (2007). Learner achievement and attitudes under different paces of transitioning to independent problem solving. *Journal of Engineering Education*, *96*(1), 45–56. https://doi.org/10.1002/j.2168-9830.2007.tb00914.x

Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *The Journal of Experimental Education*, *70*(4), 293–315. https://doi.org/10.1080/00220970209599510

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanism of a neglected phenomenon. *Educational Psychologist*, *24*(2), 113–142. https://doi.org/10.1207/s15326985ep2402_1

Schuster, C., Stebner, F., Leutner, D., & Wirth, J. (2020). Transfer of metacognitive skills in self-regulated learning: An experimental training study. *Metacognition and Learning*, *15*(3), 455–477. https://doi.org/10.1007/s11409-020-09237-5

van Gog, T., Hoogerheide, V., & van Harsel, M. (2020). The role of mental effort in fostering self-regulated learning with problem-solving tasks. *Educational Psychology Review*, *32*(4), 1055–1072. https://doi.org/10.1007/s10648-020-09544-y

van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, *43*(1), 16–26. https://doi.org/10.1080/00461520701756248

van Peppen, L. M., van Gog, T., Verkoeijen, P. P., & Alexander, P. A. (2022). Identifying obstacles to transfer of critical thinking skills. *Journal of Cognitive Psychology*, *34*(2), 261–288. https://doi.org/10.1080/20445911.2021.1990302

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p< 0.05". *The American Statistician*, *73*(2), 1–19. https://doi.org/10.1080/00031305.2016.1154108

How to cite this article: Pieplenbosch, J., G. van den Broek, Hoogerheide, V., & T. van Gog (2024). Training task-selection skills: The effect of prompts and explicit instruction on transfer. *Applied Cognitive Psychology*, *38*(3), e4200. https://doi.org/10.1002/acp.4200

APPENDIX A

Example of a Transfer Vignette in the Control (plain text part), Prompting (plain text + *italicized parts*), and Explicit Instruction Condition (plain text + *italicized* + underlined parts).[4]

Imagine you are practicing tasks from this task database:

Complexity level	***Complexity 1*** - 2 generations - 1 unknown - 1 solution - deductive			***Complexity 2*** - 2 generations - 1 unknown - multiple solutions - deductive			***Complexity 3*** - 2 generations - 1 unknown - multiple solutions - inductive			***Complexity 4*** - 3 generations - 1 unknown - multiple solutions - inductive and deductive			***Complexity 5*** - 2 generations - 1 unknown - multiple solutions - deductive		
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**	**High support**	**Low support**	**No support**
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	Hunting-ton	Wolman syndrome	Apert syndrome	Flower color	Hunting-ton	Fruit flies	Tail length	Tongue rolling	Albinism	Fruit flies	Cleft lip	Milk allergy	Tail length	Apple tree
	Apple tree	Chicken beak	Tail length	Fruit flies	Apple tree	Albinism	Chicken beak	P.R.A. disease	Depression	Cat fur	Flower color	P.R.A. disease	Apert syndrome	Wolman syndrome	Fruit flies
	Depression	Cleft lip	Milk allergy	P.R.A. disease	Tongue rolling	Mammary carcinoma	Wolman syndrome	Milk allergy	Apert syndrome	Depression	Tongue rolling	Mammary carcinoma	Hunting-ton	Flower color	Chicken beak

You have just completed a 7-step biology problem at complexity level 3 with no support.

1. You got 4 steps correct.
2. You rated your effort a 1 on a scale of 1 to 9.

Based on this description and the procedure from the videos, what is a suitable task to work on next?

Here is a hint for determining what a suitable next task would be.

Hint:

In this description, the problem-solving task has a different number of steps compared to the tasks from the videos (7 instead of 5), which means that you cannot use the exact same stepsize table as used in the videos to select a new task. However, you can adapt the stepsize table from the videos and distribute the performance scores as evenly as possible; there are several different options to do so and you can choose any one of them.

In the stepsize table from the videos, the performance scores were evenly distributed over three rows.

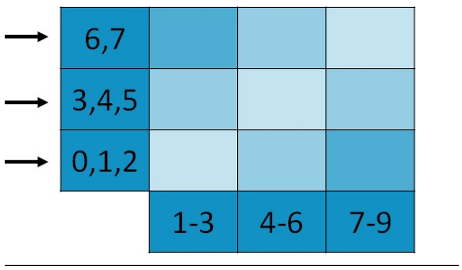
4,5			
2,3			
0,1			
	1-3	4-6	7-9

With 7 steps instead of 5, that is not possible. This is because with 7 steps you have 8 possible performance scores (0, 1, 2, 3, 4, 5, 6, 7), which cannot be evenly distributed over three rows.

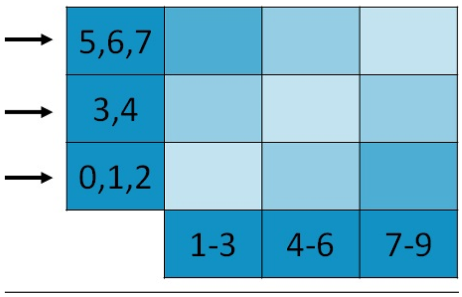
With 8 performance scores there will always be one row with 2 instead of 3 performance scores. The row with two performance scores can be the top, middle, or bottom row.

This leaves you with three different options to distribute the performance scores as evenly as possible over the three rows:

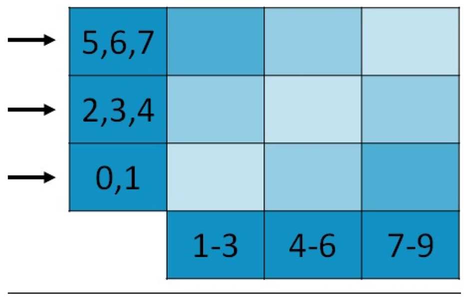
Option 1.



Option 2.



Option 3.



Luckily, with a performance of 4 and effort of 1, the answer is +1, regardless of which stepsize table you choose.

Based on this description and the procedure for selecting a new task shown in the videos, what is a suitable task to work on next? Select the task you would choose in the task database below.

APPENDIX B

Coding Scheme for the Open Cued Recall Questions.

Performance (1).
Mental effort (1).
Combine/cross reference performance and effort. They specifically write down that they used performance and effort or they imply they did this (1).
Use this information to see whether they should increase or decrease complexity (1) or support (1), or move in the task database (1).
High performance + little effort (or specific in numbers) = More complex/difficult task or less support and x steps to the right in the task database (1).
Low performance + high effort (or specific in numbers) = Less more complex/difficult task or more support and x steps to the left in the task database (1).
Medium/average performance + medium/average effort (or specific in numbers) = Same complexity/type of task or same support and 0 steps (1).
The stepsize/increase in complexity depends on how well/poor their performance was and how high/low they rated their performance (1).
Score (max. 10 points)