



The Role of Mental Effort in Fostering Self-Regulated Learning with Problem-Solving Tasks

Tamara van Gog¹ · Vincent Hoogerheide¹ · Milou van Harsel^{1,2}

Published online: 9 July 2020
© The Author(s) 2020

Abstract

Problem-solving tasks form the backbone of STEM (science, technology, engineering, and mathematics) curricula. Yet, how to improve self-monitoring and self-regulation when learning to solve problems has received relatively little attention in the self-regulated learning literature (as compared with, for instance, learning lists of items or learning from expository texts). Here, we review research on fostering self-regulated learning of problem-solving tasks, in which mental effort plays an important role. First, we review research showing that having students engage in effortful, generative learning activities while learning to solve problems can provide them with cues that help them improve self-monitoring and self-regulation at an item level (i.e., determining whether or not a certain type of problem needs further study/practice). Second, we turn to self-monitoring and self-regulation at the task sequence level (i.e., determining what an appropriate next problem-solving task would be given the current level of understanding/performance). We review research showing that teaching students to regulate their learning process by taking into account not only their performance but also their invested mental effort on a prior task when selecting a new task improves self-regulated learning outcomes (i.e., performance on a knowledge test in the domain of the study). Important directions for future research on the role of mental effort in (improving) self-monitoring and self-regulation at the item and task selection levels are discussed after the respective sections.

Keywords Self-regulated learning, · Problem-solving, · Example-based learning, · Mental effort

Learning to solve problems constitutes an important part of the curriculum in many school subjects and particularly in STEM domains (science, technology, engineering, and mathematics). Most of the problems students encounter are well-structured problems (for a typology of

✉ Tamara van Gog
t.vangog@uu.nl

¹ Department of Education, Utrecht University, P.O. Box 80140, 3508 Utrecht, TC, Netherlands

² Learning and Innovation Centre, Avans University of Applied Sciences, Breda, Netherlands

problem variations, see Jonassen 2000), and students have to learn what series of actions they should perform (possibly bounded by rules regarding what actions are or are not allowed) to get from A (given information on an initial state) to B (a described goal state) (Newell and Simon 1972). Learning to solve problems requires the acquisition of both the procedural knowledge of what actions to perform and how to perform them and the conceptual knowledge of why to perform those actions. Decades of research inspired by cognitive load theory (Sweller et al. 2011) have shown that this knowledge is not efficiently acquired by having students mainly solve practice problems. Rather, in the initial stages of skill acquisition, students benefit (i.e., attain higher test performance with less investment of time/effort) from studying examples of fully worked-out solution procedures and (subsequently) completing partially worked-out solution procedures, before solving conventional practice problems on their own (for a recent review, see Van Gog et al. 2019). In this review, we use the term “problem-solving tasks” to encompass all of those formats (from worked examples to completion problems and practice problems).

As students engage with problem-solving tasks, it is important that they can adequately self-regulate their learning process, especially during self-study sessions in which teachers are not (directly) available to guide their learning. In self-regulated learning, students have to invest effort in both the tasks (i.e., the to-be-learned content or object-level processing) and in self-regulated learning processes (i.e., meta-level processing; De Bruin et al., introduction to this special issue). Different models of self-regulated learning exist, most of which describe self-regulated learning as a cyclical process that includes preparatory (planning), performance, and evaluation phases (i.e., deciding which task to perform; performing it, monitoring progress, and possibly adjusting strategies *while* working on the task; judging performance after the task is completed; using this as input for deciding on the next task; et cetera) and involves cognitive, behavioral, motivational, and affective processes in each phase (for a review, see Panadero 2017). These models of self-regulated learning emphasize that accurate self-monitoring (i.e., judging one’s level of performance) and self-regulation (i.e., deciding on how to proceed) are essential for the *effectiveness* of self-regulated learning (i.e., for learning outcomes on the to-be-learned content).

It is well known, however, that students’ monitoring judgments are often inaccurate, and self-monitoring of problem-solving tasks is no exception (e.g., Baars et al. 2014a, 2017; De Bruin et al. 2005, 2007). This is problematic, because accurate self-monitoring is considered a necessary (though not sufficient) condition for accurate self-regulation and, thereby, for the effectiveness of self-regulated learning (e.g., Dunlosky and Rawson 2012). Therefore, researchers have been trying to identify ways of improving students’ monitoring accuracy. However, the majority of research on how to improve self-monitoring (and, thereby, self-regulation and learning outcomes) was conducted in the context of learning lists of items (e.g., word pairs, concept definitions; e.g., Dunlosky and Rawson 2012) or learning from expository texts (see Griffin et al. 2019, for a review). How to improve self-monitoring and self-regulation of problem-solving tasks has received much less attention, and given that the to-be-monitored cognitive processes involved in problem-solving are very different, findings from research on learning lists of items or learning from texts might not generalize (De Bruin and Van Gog 2012; Ackerman and Thompson 2017).

Here, we review some of the available research on fostering self-regulated learning of problem-solving tasks, in which mental effort (i.e., the “cognitive capacity that is actually allocated to accommodate the demands imposed by the task”; Paas et al. 2003, p. 64) plays an important role in two different ways. First, we review research showing that having students

engage in effortful, generative learning activities while learning to solve problems can provide them with cues that help them improve self-monitoring and self-regulation at an item level (i.e., determining whether or not a certain type of problem needs further study or practice). Then, we turn to self-monitoring and self-regulation at the task sequence level (i.e., determining what an appropriate next problem-solving task would be given the current level of performance). We review research showing that teaching students to regulate their learning process by taking into account a combination of their (self-assessed) performance and their (self-reported) invested mental effort on a prior task when selecting a new task improves self-regulated learning outcomes (i.e., performance on a knowledge test in the domain of study). After each section, we discuss important directions for future research on the role of effort in (improving) the accuracy of self-monitoring and self-regulation.

Improving Self-Monitoring and Self-Regulation at Item Level

Generative learning activities are activities that help learners to actively construct meaning from the learning materials (Fiorella and Mayer 2016). These activities, which are usually quite effortful, have mainly been investigated as a means to improve learning outcomes: they foster deeper processing of the learning materials and thereby improve learning and transfer performance (i.e., in cognitive load theory terms, generative activities impose *germane* cognitive load: the effort invested in them contributes to learning; Sweller et al. 2011). However, they can also be used to improve students' monitoring accuracy. Research on text comprehension has shown that having students generate keywords (De Bruin et al. 2011; Thiede et al. 2005), summaries (Anderson and Thiede 2008; Thiede and Anderson 2003), diagrams (Van de Pol et al. 2019; Van Loon et al. 2014), drawings (Kostons and De Koning 2017; Schleinschok et al. 2017), or concept maps (Redford et al. 2012; Thiede et al. 2010), after studying a text and before they are asked to make a prospective judgment of learning (predicting their performance on a future test) regarding that text, improves their monitoring accuracy (i.e., their predicted performance is closer to their actual performance; see also Van de Pol et al., this special issue).

The finding that generative learning activities improve monitoring accuracy can be explained by the cue-utilization framework (Koriat 1997). Learners base their monitoring judgments on different sources of information (i.e., cues) originating from the (presentation of the) study material and their experience while studying. The extent to which their judgment is accurate depends on the cue diagnosticity, that is, on how predictive the cues are for their actual test performance. For performing well on a text comprehension test, learners need to understand the gist of a text (as opposed to simply memorizing facts). Because generative activities provide learners with access to cues regarding their understanding of the gist of the text, monitoring accuracy improves.

While most of the research on the use of generative activities to improve monitoring accuracy has focused on learning lists of items (e.g., word pairs, concept definitions) or learning from (expository) texts, several studies have investigated whether and how they affect monitoring accuracy when learning to solve problems by studying worked examples. Generative activities that were investigated include solving an isomorphic practice problem after studying a worked example (i.e., generating the entire solution procedure oneself), completing steps in a partially worked-out example (i.e., generating parts of the solution procedure oneself), self-explaining solution steps, and comparing one's answers to a standard (i.e., correct answer/solution procedure).

In a study in primary education, Baars et al. (2014a) found that providing students with an isomorphic practice problem after they had studied a worked example on how to solve a water jug problem improved their prospective monitoring accuracy (i.e., it reduced their overestimations of their performance on a future test). However, it did not improve their regulation accuracy (i.e., their judgment of which examples they should restudy). For secondary education students (Baars et al. 2017), the opportunity to solve a practice problem did seem to affect their self-regulation: Adolescents who studied examples on how to solve heredity problems in biology and then received a practice problem to solve made more accurate monitoring judgments and were more accurate at indicating which examples they should restudy.

Interestingly, completing solution steps was found to make secondary education students *underestimate* their performance. Baars et al. (2013) presented students either with fully worked-out examples of heredity problems in biology or with partially worked-out completion problems, in which students had to generate some steps themselves. In the completion condition, students' predictions of their own test performance were lower than their actual performance. They also reported investing significantly more mental effort than students in the fully worked-out example condition. Given that the self-reported mental effort negatively correlated with their judgments of learning, these findings may suggest that students used the high effort as a cue indicating a lack of understanding (see also Baars et al., this special issue). When students fail to take into account that they (presumably) learned from engaging in the completion activity, this would lead to underestimation of their test performance. Possibly, findings would have been different if students would first have been given the opportunity to study a fully worked example, before moving to completion problems (this would have allowed them to acquire knowledge of the steps and then test the quality of that knowledge during step completion).

In contrast to generating full or partial solutions, self-explaining did not seem to improve monitoring. While self-explaining had been found to be effective for improving monitoring accuracy of text comprehension (Griffin et al. 2008), it did not affect secondary education students' retrospective monitoring accuracy (i.e., judging how well they performed a task just completed, also known as self-assessment) when applied after studying examples or solving problems (Baars et al. 2018a).

Baars et al. (2018a) did find that problem complexity affected monitoring accuracy: students made more accurate retrospective monitoring judgments on less complex problems than on more complex problems. There may be several (not mutually exclusive) reasons for this. First, the more complex a problem is, the higher the cognitive load it imposes (i.e., the more effort it requires), and the less resources are available for monitoring (keeping track of) what one is doing *during* task performance, which reduces the quality of the cues available for judging performance afterwards (Van Gog et al. 2011a). Second, as performance is also better on less complex problems, lower monitoring accuracy on more complex problems may reflect the “unskilled and unaware” effect (Kruger and Dunning 1999): Students who lack the knowledge to perform well on a task also lack knowledge of the standards against which their performance should be assessed (i.e., of what would constitute good performance). Third, and related to the issue of (lacking) performance standards, on more complex problems, it may be more difficult for students to know which parts of the problem they solved (in)correctly than on less complex problems. Indeed, providing students with a standard (i.e., the correct answer/solution procedure) against which to compare their own performance while making monitoring judgments was found to improve the accuracy of students' retrospective monitoring judgments (Baars et al. 2014b; Oudman et al. submitted for publication; for similar findings in learning concept definitions from texts: Rawson and Dunlosky 2007).

Directions for Future Research on the Role of Mental Effort in (Improving) the Accuracy of Self-Monitoring and Self-Regulation at the Item Level

The research reviewed in the previous section shows that having students invest effort in generative activities during or after example study seems to improve their monitoring accuracy, presumably because these activities provide learners with good cues about their understanding of the solution procedure. An important direction for future research in this area would be to gain a better understanding of what cues students derive from the generative activity and of the role of mental effort in this process. The available research suggests that students may use the effort involved in completing the task or generative activity as a cue for making monitoring judgments: Several studies have found students' effort ratings to correlate negatively with their monitoring judgments (e.g., Baars et al. 2013; Baars et al. 2018b; Baars et al. 2014b; see also Baars et al., this special issue). This suggests that students tend to see high effort as a sign of poor performance and low effort as a sign of good performance (cf. a data-driven interpretation of effort; Koriat et al. 2014).

Indeed, the amount of effort students need to invest to complete a generative activity like completing steps in a partially worked-out example, or solving a practice problem, might be a good indicator of the quality of their mental model of the solution procedure, because the higher the students' level of skill is, the higher their performance efficiency is (i.e., higher accuracy achieved in less time and with less effort; see Kalyuga and Sweller 2005; Van Gog and Paas 2008). However, whether invested effort is a good cue to use when judging (future test) performance would depend on several other factors: (1) the accuracy of students' performance on the generative activity; (2) the design of the generative activity; and (3) when making prospective monitoring judgments, the possibility that engaging in the activity contributes to learning (i.e., performance on future tasks).

First, ratings of invested mental effort cannot be meaningfully interpreted without looking at the quality of the associated performance (Van Gog and Paas 2008). For instance, when a student completes a step rapidly and with little effort, but wrongly, the invested effort is not very informative of the quality of the mental model of the problem-solving procedure, and using it as a cue for a monitoring judgment will lead to poor monitoring accuracy. However, students will not always be able to infer with certainty whether their performance was correct or not.

Second, whether invested mental effort is a good cue for making monitoring judgments depends on the *origin* of the effort (i.e., the processes effort is invested in), which can be affected by other (extraneous) variables than the quality of students' mental model of the solution procedure. For instance, when the design of the generative activity requires students to invest a lot of effort in extraneous processing (i.e., simply originating from the way in which the material is presented, e.g., because split-attention is induced or because redundant information is presented; Sweller et al. 2011), then the amount of effort invested is no longer very informative about the quality of their mental model and, thus, not predictive of their (future test) performance. Vice versa, if the design of a generative activity includes high levels of instructional support (e.g., hints on how to complete steps), then it might be completed with little effort, but again, this might not be very informative of the quality of the mental model and, hence, of (future) performance in the absence of that support. In such cases, using effort as a cue will not improve (and more likely reduce) monitoring accuracy. It is unlikely, however, that students will be able to distinguish to what extent the effort they had to invest to complete the activity originated from the way in which the activity was designed.

Third, when making prospective monitoring judgments (i.e., predicting future test performance), the usefulness of effort invested in the generative activity as a cue also depends on whether students learn from engaging in the generative activity (for which there is evidence, see Fiorella and Mayer 2016). If the activity requires high effort, but simultaneously contributes to improving their mental model of the solution procedure (cf. the concept of germane cognitive load; Sweller et al. 2011), then students might underestimate their future test performance when using effort invested in the generative activity as a cue (unless they recognize that the high effort they invest is contributing to their learning; cf., a goal-driven interpretation of effort; Koriat et al. 2014).

Thus, future research should aim to establish whether students indeed use effort as a cue in making monitoring judgments (the evidence so far is correlational) and under which circumstances using effort as a cue does and does not improve monitoring accuracy. Such insights could then perhaps be used to design interventions to help students establish when effort is or is not a good cue to use for making monitoring judgments. Moreover, future research should continue to investigate what kind of generative activities can foster students' monitoring accuracy. In developing and testing generative activities, it should be kept in mind that the design of these activities is crucial: Ideally the (effort) cues they generate should be associated with the quality of the mental model and not originate from other, extraneous factors.

Another important direction for future research that we currently know very little about is to what extent students consider mental effort when making regulation decisions. That is, in deciding whether or not an example would have to be restudied or a problem should be practiced further, it is likely that students would make an estimation of the amount of effort that a task requires and would consider whether they are willing to invest that effort. Indeed, their willingness to (continue to) invest effort in (re)studying might be an important variable in explaining why increasing monitoring accuracy is not always sufficient for improving regulation accuracy (as some of the findings reviewed above also showed): students may know they would have to restudy a task, but may not be willing to do so (e.g., if they feel it requires too much effort and/or feel that restudying would not help them to get a better understanding of the task).

Last but not least, an interesting avenue for future research might be to investigate whether students' effort investment is also a useful cue for teachers in monitoring their students' learning. Teachers' monitoring accuracy of students' learning is not only a requirement for adaptive teaching but also necessary if teachers are to help their students to improve their self-monitoring and self-regulation skills. Recent findings suggest that generative activities on problem-solving tasks do not only provide students' with cues that help them monitor their own learning but may also aid teachers in monitoring their students' learning. For instance, having access to students' performance on practice problems improves teachers' judgments of students' test performance (Oudman et al. 2018; for similar findings in text comprehension, see Thiede et al. 2019; Van de Pol et al. 2019, this special issue). Perhaps, giving teachers' insight into the amount of mental effort students invested might further improve their accuracy when monitoring and regulating their students' learning (e.g., if two students performed a task very well, and one of them had to invest little effort while the other had to invest a lot of effort, the first student might not have to practice the task further, while the other might benefit from further practice; teachers could not glean this information from the accuracy of students' performance only). Moreover, this might enable teachers to better assist their students in self-regulating their learning (e.g., discussing with them under which conditions invested or required effort might or might not be useful to consider when making monitoring judgments and regulation decisions).

Improving Self-Monitoring and Self-Regulation at Task Sequence Level

While regulation in terms of deciding whether or not to study or practice a similar task again does play a role in self-study, students typically have to make more complex regulation decisions during self-regulated learning with problem-solving tasks. They will usually be confronted with a whole array of potential problem-solving tasks in STEM textbooks/workbooks or in online learning environments, spanning across multiple levels of complexity. Students should ideally be working on tasks that are at an optimal level of complexity and provide an optimal level of support given their current level of knowledge (Van Merriënboer 1997; Van Merriënboer and Kirschner 2013). That is, the tasks should be a little more complex but, with the appropriate support, within reach (cf. zone of proximal development; Vygotsky 1978).

As such, self-regulated learning of problem-solving tasks requires monitoring and regulation at a *task sequence* level: students need to be able to self-assess their performance on a task just completed (i.e., make a retrospective monitoring judgment), and then select a next task with the right level of complexity and/or support. When students show poor self-assessment accuracy and underestimate or (more commonly) overestimate their own performance (Kruger and Dunning 1999; Hacker and Bol 2019), their task selection choices will probably not be adaptive to their current level of performance, which has detrimental effects on their learning outcomes (i.e., performance on a posttest on the to-be-learned content). That is, students would end up wasting valuable study time and effort on tasks they can already do (in case of underestimation) or might move on to tasks that are still too difficult (in case of overestimation). Both scenarios might well have detrimental effects on their motivation (i.e., may leave them underwhelmed or overwhelmed, respectively) and, thus, on their willingness to continue to invest effort in self-study. Students might also quit studying prematurely in case of overestimation, thinking they are already well prepared for their next class or an exam, when in fact they are not.

In sum, and as mentioned earlier, both accurate self-assessment (monitoring) and accurate task selection (regulation) are key for the effectiveness of self-regulated learning (i.e., learning outcomes on the to-be-learned content), and accurate self-assessment is a necessary but not a sufficient condition for accurate task selection. The reason it is not sufficient is that regulation at the task sequence level also requires learners to take into account the characteristics of the prior as well as the novel (i.e., not yet studied) tasks that are available, which makes it substantially more complex than regulation at the item level (i.e., deciding whether or not to restudy). For instance, students have to pay attention to the complexity level of the just-completed and to-be-selected task and to the amount of instructional support the task provides (e.g., a worked example: highest possible support; completion problem: some support, which can be higher or lower depending on the ratio of steps worked out vs. to be completed; conventional practice problem: no support). In other words, in deciding on a new task, students are either implicitly or explicitly also regulating their effort investment, because the complexity and format of the new task (in combination with students' current level of performance) will determine the amount of effort they will need to invest.

We still know very little about what task selection choices students make (and in doing so, how students regulate their effort) when they engage in self-regulated learning with examples and practice problems. Below, we will first review some recent empirical studies on this question. Subsequently, we will discuss research showing that students' task selection choices and, importantly, the effectiveness of self-regulated learning (i.e., performance on a posttest on

the to-be-learned content) improve from teaching them to consider the combination of their (self-assessed) performance and invested mental effort on a prior task when selecting a next one.

Task Selection Choices During Self-Regulated Learning with Problem-Solving Tasks

Instructional design research has shown that studying examples is typically less effortful for novice learners than practice problem solving (Van Gog et al. 2019). An interesting question regarding students' regulation of effort (De Bruin et al., introduction to this special issue) then is whether students realize this and would select tasks in line with what we know to be effective task sequences, for instance, starting (at each new complexity level) with studying an example before moving to a practice problem (cf. Renkl and Atkinson 2003; Van Gog et al. 2011b; Van Harsel et al. 2019, *in press*; Van Merriënboer 1997).

A recent study with university students (Foster et al. 2018) found that when students were given a choice over what tasks to work on, the ratio of students' selection of worked examples versus practice problems was on average about 40% versus 60%. In line with instructional design principles (e.g., Van Merriënboer 1997), the percentage of examples selected decreased and the percentage of practice problems increased over time and there was a higher probability of an example being selected after a failed than after a correct problem-solving attempt. In contrast to instructional design principles, however, students rarely started the training phase with example study.

Yet this was different in Van Harsel et al. (manuscript in preparation). Higher education students (first year students at a Dutch university of applied sciences) learned to solve math problems on the trapezoid rule. After a pretest, a self-regulated learning phase followed, in which students were instructed to select six tasks of their own choice, from a database with 45 tasks, with the aim of preparing for the posttest that would follow. The tasks in the database were organized at three levels of complexity, with three task formats per complexity level: video modeling examples (in which the solution procedure was being worked out by an instructor, appearing on the screen step by step in handwriting, and explained in a voice-over), worked examples (text based, the worked-out solution procedure was visible all at once), or practice problems. At each level, students had 5 isomorphic tasks (i.e., different cover stories) for each format to choose from (i.e., $3 \times 3 \times 5 = 45$ tasks in total; they could select each task only once). Almost 77% of students started with observing a video modeling example, 19% with a worked example, and only 4% with a practice problem, and the vast majority of students started at the lowest complexity level (almost 89%). While the selection of examples gradually decreased, and the selection of practice problems gradually increased over the course of the self-regulated learning phase, the ratio of examples (i.e., video and worked examples combined) remained quite high (see Table 1).

Interestingly, the findings also revealed a difference in task sequences chosen between students who scored the highest and those who scored the lowest on the posttest tasks that were isomorphic to the self-regulated learning phase tasks. In the second half of the learning phase, low-scoring students continued to select tasks at the lowest complexity level and continued to select examples more often than high-scoring students. Unfortunately, performance data on the practice problems was not available, so it is not entirely clear whether the study behavior of low-scoring students simply reflects that they did not yet grasp the lowest complexity problems (in which case their study behavior is quite adaptive), reflects poor regulation (i.e., if they would have been able to move on to more complex problems, but did not study/practice those, this explains why they would perform lower on the posttest), or reflects a lack of motivation.

Table 1 Percentages of selected video modeling examples, worked examples, and practice problems in the self-regulated training phase (Van Harsel et al. manuscript in preparation)

	Video modeling example	Worked example	Practice problem
Task 1	76.9%	19.0%	4.1%
Task 2	24.5%	31.3%	44.2%
Task 3	37.4%	22.4%	40.1%
Task 4	35.4%	28.6%	36.1%
Task 5	30.1%	34.9%	34.9%
Task 6	17.0%	28.6%	54.5%

Recent large-scale observational studies (Rienties et al. 2019; Tempelaar et al. *in press*) with university students, during an 8-week introductory mathematics course in a blended learning environment also showed that students made substantial use of worked examples in the online study environment (in up to 43% of the exercises; Rienties et al.). However, large differences among students were observed: several clusters of students were identified that differed widely in the amount of examples studied and practice problems completed and in when they worked on those tasks (i.e., in preparation for which activity in the course: tutorial sessions, quizzes, or the final exam). While some of these patterns of self-regulated study behaviors are likely more effective and efficient than others, these studies unfortunately did not allow for drawing strong conclusions about that (e.g., because of large differences in background and prior knowledge in mathematics and because only part of the learning activities in the course took place in the only study environment).

In sum, we know relatively little about task selection choices students make and even less about *why* they make those choices. The available data suggest that students do seem to be (implicitly or explicitly) aware of the fact that examples are efficient for learning (i.e., by providing high instructional support and requiring relatively little effort compared to practice problems) as they make substantial use of them, although there are differences among studies and among students in when students opt for example study. Yet, it remains unclear from these studies why students switch from examples to practice problems and vice versa and whether they (implicitly or explicitly) use effort as a cue in making these decisions to alternate between formats. However, as we will show in the next section, we do know that training students to consider invested effort during task selection can improve the effectiveness of self-regulated learning.

Training Students to Make Task Selection Choices Based on a Combination of Their (Self-Assessed) Performance and Invested Mental Effort

In this section, we review two studies (Kostons et al. 2012; Raaijmakers et al. 2018b) that have shown that training students' self-assessment and task selection skills by means of video modeling examples that teach them to consider the combination of their (self-assessed) performance and (self-reported) invested mental effort on a prior problem-solving task when selecting a new one can improve the effectiveness of self-regulated learning.¹ This training

¹ Note that there are other effective approaches to improving self-assessment and task selection skills in the context of problem-solving tasks as well (e.g., in intelligent tutoring systems: Long and Alevan 2016, 2017; Roll et al. 2011). We did not review these studies here as they fall outside the scope of this article (i.e., they did not focus on the role of mental effort, and task selection was not always fully self-regulated but often implemented as "shared control" in which the system first makes a pre-selection of tasks at an appropriate level, from which the learner then gets to choose; see also Corbalan et al. 2008).

approach was inspired by adaptive task selection systems (Anderson et al. 1995; Koedinger and Alevan 2007). Adaptive systems monitor individual students' performance on a problem on several parameters, which are then used in an algorithm to select a suitable next problem for the student from a task database. In the studies reviewed here, secondary education students were taught a simplified task selection algorithm, combining their (self-assessed) performance and invested mental effort on the current problem.

The studies were conducted in the domain of biology; more specifically, students had to learn to solve monohybrid cross problems (Mendel's law of heredity). The procedure for solving these problems consisted of five steps: (1) translating the information given in the cover story into genotypes, (2) putting this information in a family tree, (3) determining the number of required Punnett squares, (4) filling in the Punnett square(s), and (5) finding the answer(s) in the Punnett square(s). An online learning environment was designed, with a task database presenting the problem-solving tasks that students could choose from. The task database was designed according to two key instructional design principles from the Four-Component Instructional Design model (4C/ID Model; Van Merriënboer 1997; Van Merriënboer and Kirschner 2013) stating that: (1) learning tasks should be ordered from simple to complex, and (2) within each level of complexity, learning tasks should be offered that initially provide a high level of instructional support, which is slowly reduced (faded out) in later tasks as the learners' knowledge increases (i.e., the "completion" or "fading" principle; Paas 1992; Renkl and Atkinson 2003; Van Merriënboer 1990). Accordingly, the problem-solving tasks in the database (see Fig. 1) were ordered in five increasing levels of complexity, and within each level of complexity, three levels of instructional support were provided: high (completion problems with most solution steps worked out, some for the learners to complete), low (completion problems with some steps worked out, most for the learners to complete), or no support (conventional problem: no steps worked out, learners solve the entire problem themselves). At each level of support within a complexity level, students could choose from five isomorphic tasks that had the same solution procedure (i.e., the same structural features) but different cover stories (i.e., different superficial features; e.g., a task on eye color vs. hair structure).

The self-assessment and task selection training was provided by means of video modeling examples in which another student (i.e., the model) demonstrated and explained the self-assessment and task selection procedure. The training consisted of a general introduction followed by four video modeling examples, two by male and two by female models (to prevent potential effects of model-observer similarity on training outcomes; Schunk 1987). Students first observed the model solving a problem (low complexity, no support) from the task database (note that two of the models did not correctly solve the problem to create variability in performance/self-assessment). After solving the problem, the model rated invested mental effort (scale of 1-9; Paas 1992) and self-assessed performance, assigning 1 point for every problem-solving step correctly completed (0-5). Then, the model selected a new task from the database by combining the (self-assessed) performance and mental effort rating, according to a table (based on task selection algorithms used in prior research, Corbalan et al. 2008; Salden et al. 2006; see Fig. 2): When high performance was attained with low effort, the table advised selecting a task with much less support/more complexity; with high performance and high effort to practice a similar task again (until some efficiency is reached), with low performance and high effort to choose a less complex task or more support.

Kostons et al. (2012) first conducted an experiment to test the effectiveness of the training. Students in the experimental group completed a pretest (to check that they indeed had little if

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

Fig. 1 Task database (Raaijmakers et al. 2018b): 5 complexity levels, 3 levels of instructional support within each complexity level, and 5 different learning tasks within each complexity-support level

any prior knowledge of the biology problems) and then observed the four video modeling examples of the entire performance – assessment – selection cycle. Students in the “no training” control group received only the performance part of the video modeling examples (this was done to ensure that any effects on the posttest were not due to the knowledge on how to solve the biology problems conveyed in the examples). The training was shown to enhance self-assessment accuracy (i.e., the absolute deviation between the judged and actual performance) and task selection accuracy (i.e., the deviation [number of columns in Fig. 1] between the chosen task and the recommended task).

In a second experiment, Kostons et al. (2012) investigated the effects on learning outcomes. Students first completed a pretest, then entered the training phase (same experimental/control

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

Fig. 2 Table used by the models to select a new task from the database. Suppose a no support problem at complexity level 1 was completed, with self-assessed performance of 4 and mental effort rating of 2, the task selection advice would be +2, meaning 2 columns to the right in the database in Fig. 1: a low support problem at complexity level 2

conditions), after which all students engaged in a self-regulated learning phase, working on eight tasks of their choice from the database, and finally, completed a posttest. Students in the training condition outperformed students in the control condition on the posttest problems. Thus, the *effectiveness* of the self-regulated learning phase (in terms of how much knowledge students acquired on how to solve the biology problems) was higher when students had received self-assessment and task selection training. In other words, students who had been trained made better choices and gained more from self-regulated study.

Importantly, this finding was replicated and extended in a study by Raaijmakers et al. (2018b), who showed that the effectiveness of self-regulated learning not only improved when the training was based on the specific algorithm (shown in Figure 2) but also improved when a more general task selection heuristic was used (e.g., the model saying “I attained a high score on performance with a medium amount of effort, so I am ready for a more difficult task/one with less support.”).² Moreover, they found some evidence that the task selection skills might transfer: Trained students made better task selection choices when they were presented with vignettes, in which the performance and invested mental effort of another (fictitious) student in a new problem context were given (e.g., math problems with 8 steps and a database with 4 complexity and 2 support levels) and were asked to select a suitable next task for their (fictitious) peer. However, it is as yet unclear whether the effects of the training can indeed be (made to) transfer to self-regulated learning in other problem contexts. In a follow-up study (Raaijmakers et al. 2018a), students who first received the self-assessment and task selection training in the context of the biology problems and then had to engage in a self-regulated learning with math problems (linear equation problems consisting of 3 steps; organized in a database with 5 complexity and 2 support levels) did not show better math posttest performance than students who did not receive the training. As it would be unfeasible to train task selection skills for each type of problem students might encounter, it is important for future research to investigate obstacles and conditions for successful transfer of self-assessment and task selection skills.

Nevertheless, the fact that self-assessment and task selection training were found to improve the effectiveness of subsequent self-regulated learning—and that this finding was replicated—is important. It shows that teaching students a relatively simple task selection rule based on a combination of their self-assessed performance and invested mental effort improves their self-regulated learning outcomes. Despite a large body of research on interventions to improve monitoring accuracy across a variety of learning tasks, there are relatively few interventions that were actually shown to affect learning outcomes (Hacker and Bol 2019).

Directions for Future Research on the Role of Mental Effort in Self-Assessment and Task Selection at the Task Sequence Level

The previous section not only shows that there is a paucity of research into the role of mental effort in self-regulated learning at the task sequence level but also shows that research in this direction is very promising for fostering self-regulated learning with problem-solving tasks. Yet a lot of work remains to be done, and we will discuss several directions for future research

² Note, though, that initial experiments (included as supplementary material in Raaijmakers et al. 2018b) revealed that particular design features of the video examples are key to the effectiveness of the training (see supplementary materials of Raaijmakers et al. 2018b; see also Baars et al. 2014b, who found no effects with written training examples).

here, along the lines outlined by De Bruin et al. (introduction to this special issue). As reviewed in the previous section, we still know very little about what task selection choices students make when they engage in self-regulated learning with examples and practice problems. But we know even less about *why* they make those choices and what role mental effort plays in this, both in terms of effort invested in the prior task (i.e., effort monitoring) and in terms of effort they expect to have to invest on a subsequent task (i.e., effort regulation).

First, as for effort (in) monitoring, it is imperative to gain more insight into the cues students use for monitoring their performance and for monitoring invested effort. On the one hand, as mentioned earlier, students may use effort as a cue in monitoring their performance, and the research on training them to do so (reviewed in the previous section) shows that this can be a fruitful strategy. On the other hand, the negative association between effort ratings and monitoring judgments found in research at the item level might also indicate that there are other cues that students use both in making performance judgments and in making effort ratings. Indeed, a conceptual as well as a measurement issue regarding self-reported mental effort is that it is not entirely clear what effort ratings are based on (see also Scheiter et al., this issue). Some studies have shown that effort ratings can be influenced, for instance, by *when* effort is reported (i.e., immediately after a task vs. after a series of tasks; Schmeck et al. 2015; Van Gog et al. 2012) and by whether performance feedback is given prior to making an effort rating. Raaijmakers et al. (2017) manipulated feedback valence (regardless of actual performance, i.e., students were led to believe they performed correctly/incorrectly, or better/worse than they thought), and negative feedback was found to result in higher and positive feedback in lower effort ratings. This suggests that students' affective response to the feedback and/or their altered perceptions of the task demands influenced their effort ratings. Therefore, next to clarifying the cues on which effort monitoring is based, future research should also continue to address when and how students can best be asked to report their invested mental effort. Thus far, when using rating scales, it seems advisable to assess effort immediately upon task completion and before any feedback on the task is provided.

Second, with regard to effort regulation, we need to gain more insight into the cues students use to regulate their effort in the task selection phase (i.e., the preparatory phase of self-regulated learning cycle). How do students perceive the demands imposed by the different task formats that they can choose from (i.e., how much effort do they expect they would have to invest) and how does this affect their choice of complexity level or task format? Could we help them to more effectively regulate their (effort during) learning by instructing students on the merits of different task formats for different phases in their learning process? And how much effort are students *willing* to invest to continue to improve their performance (see also Paas et al. 2005)? What role do affective and motivational variables (e.g., perceived self-efficacy) play, both in how they perceive task demands and in how much effort they are willing to invest? For instance, a lack of motivation might be a potential explanation for the task selection pattern (i.e., often staying at the lowest complexity level) observed in students who performed poorly on the posttest in the study by Van Harsel et al. (manuscript in preparation). Also, in the training studies reviewed above, there was a substantial variance in the extent to which students benefitted from self-assessment and task selection training. Kostons et al. (2012) suggested that this might have been due to differences in motivation, and indeed, a study by Baars and Wijnia (2018) showed that students' task-specific motivation was associated with their monitoring accuracy after training (they used the training approach from Kostons et al. 2012; Raaijmakers et al. 2018a, b).

One problem in addressing the basis for students' monitoring judgments, effort ratings, and regulation decisions is that it is unclear to what extent students are aware of these judgment and decision-making processes or whether they occur rather automatically. This affects what methods and techniques could be used to study these questions. For instance, process-tracing methods such as verbal protocols or eye tracking (or a combination of both; Van Gog et al. 2005) have been used for studying self-assessment and task selection processes (cf. Nugteren et al. 2018) and might be one way to reveal, for instance, why students switch from examples to practice problems and vice versa and whether they use effort as a cue in making these decisions to alternate between formats. Another method that has been used is to ask students to report on what cues they use during self-assessment and task selection either as open question (Van Harsel et al. manuscript in preparation) or with the aid of a list of potential cues (cf. Van de Pol et al. submitted for publication). However, these methods assume that students are aware of and can introspect on their monitoring and regulation processes. As this might not always be the case, these methods could/should ideally be used in conjunction with experimental approaches that, for instance, systematically vary which cues are available (see e.g., studies by Van de Pol et al. submitted for publication, and Oudman et al. 2018, with teachers) or try to establish which factors differentially affect effort judgments, monitoring judgments, and objective measures of effort and performance (cf. Ackerman 2019; Scheiter et al., this special issue).

Third, regarding the question of how we can optimize cognitive load during self-regulated learning, as students have to invest effort both in studying the to-be-learned content (primary task) and in self-regulating their learning process (secondary task), the training approach we reviewed above seems very promising. However, the caveats mentioned earlier (in the prior "future directions" section) also apply here; if effort ratings are affected by extraneous variables, the trained task selection approach will not work optimally. Future research should also look into further optimizing the training. Even though our simple approach to teaching students to combine (self-assessed) performance and effort for the selection of new task was effective for improving learning outcomes, it is likely that the algorithm that was used can be further improved upon (note, though, that any algorithm would have to be relatively simple, and much simpler than the advanced algorithms used in adaptive systems, or students will not be able to understand, remember, and apply it during self-regulated learning). Another important avenue for future research would be to investigate obstacles and conditions for successful transfer and the role of cognitive load therein. That students were not able to transfer the task selection skills trained in one type of problem context to another problem context could have been due to high cognitive load imposed by the primary task (i.e., solving the new problems), in which case students might have too little resources left to simultaneously remember, adapt, and apply the learned task selection procedures.

Last but not least, future research should move beyond controlled contexts, and when doing so, investigating students' motivation to invest effort (Paas et al. 2005) becomes even more important. While most of the research discussed here was already conducted in a classroom context, it was still a controlled context. Students could choose which tasks to work on during the self-regulated learning phase, but they had to choose a certain number of tasks and knew they would be tested immediately after the learning phase (e.g., Foster et al. 2018; Van Harsel et al. manuscript in preparation; Kostons et al. 2012; Raaijmakers et al. 2018b). The findings from the studies by Rienties et al. (2019) and Tempelaar et al. (in press) suggest that (individual differences in) motivation have a huge influence on what tasks students select and when in naturalistic course contexts. Given that effective self-regulated learning keeps students within their zone of proximal development, it will be effortful to maintain, as students

are always working on tasks that are just above their current level of performance. So, even if we can train students to become more effective self-regulators, an important open question is whether they would persist in this kind of effortful self-study when it is up to them.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ackerman, R. (2019). Heuristic cues for meta-reasoning judgments: review and methodology. *Psychological Topics*, 28(1), 1–20. <https://doi.org/10.31820/pt.28.1.1>.
- Ackerman, R., & Thompson, V. A. (2017). Meta-Reasoning: monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, 21(8), 607–617. <https://doi.org/10.1016/j.tics.2017.05.004>.
- Anderson, M. C. M., & Thiede, K. W. (2008). Why do delayed summaries improve metacomprehension accuracy? *Acta Psychologica*, 128(1), 110–118. <https://doi.org/10.1016/j.actpsy.2007.10.006>.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: lessons learned. *Journal of the Learning Sciences*, 4(2), 167–207. https://doi.org/10.1207/s15327809jls0402_2.
- Baars et al., this issue (n.d.)
- Baars, M., & Wijnia, L. (2018). The relation between task-specific motivational profiles and training of self-regulated learning skills. *Learning and Individual Differences*, 64, 125–137. <https://doi.org/10.1016/j.lindif.2018.05.007>.
- Baars, M., Visser, S., Van Gog, T., De Bruin, A., & Paas, F. (2013). Completion of partially worked-out examples as a generation strategy for improving monitoring accuracy. *Contemporary Educational Psychology*, 38(4), 395–406. <https://doi.org/10.1016/j.cedpsych.2013.09.001>.
- Baars, M., Van Gog, T., De Bruin, A., & Paas, F. (2014a). Effects of problem solving after worked example study on primary school children's monitoring accuracy. *Applied Cognitive Psychology*, 28(3), 382–391. <https://doi.org/10.1002/acp.3008>.
- Baars, M., Vink, S., Van Gog, T., De Bruin, A. B. H., & Paas, F. (2014b). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. *Learning and Instruction*, 33, 92–107. <https://doi.org/10.1016/j.learninstruc.2014.04.004>.
- Baars, M., Van Gog, T., De Bruin, A., & Paas, F. (2017). Effects of problem solving after worked example study on secondary school children's monitoring accuracy. *Educational Psychology*, 37(7), 810–834. <https://doi.org/10.1080/01443410.2016.1150419>.
- Baars, M., Leopold, C., & Paas, F. (2018a). Self-explaining steps in problem-solving tasks to improve self-regulation in secondary education. *Journal of Educational Psychology*, 110(4), 578–595. <https://doi.org/10.1037/edu0000223>.
- Baars, M., Van Gog, T., de Bruin, A., & Paas, F. (2018b). Accuracy of primary school children's immediate and delayed judgments of learning about problem-solving tasks. *Studies in Educational Evaluation*, 58, 51–59. <https://doi.org/10.1016/j.stueduc.2018.05.010>.
- Corbalan, G., Kester, L., & Van Merriënboer, J. J. (2008). Selecting learning tasks: effects of adaptation and shared control on learning efficiency and task involvement. *Contemporary Educational Psychology*, 33(4), 733–756. <https://doi.org/10.1016/j.cedpsych.2008.02.003>.
- De Bruin, A. B. H., & Van Gog, T. (2012). Improving self-monitoring and self-regulation: from cognitive psychology to the classroom. *Learning and Instruction*, 22(4), 245–252. <https://doi.org/10.1016/j.learninstruc.2012.01.003>.
- De Bruin, A. B. H., Rikers, R. M. J. P., & Schmidt, H. G. (2005). Monitoring accuracy and self-regulation when learning to play a chess endgame. *Applied Cognitive Psychology*, 19(2), 167–181. <https://doi.org/10.1002/acp.1109>.
- De Bruin, A. B. H., Rikers, R. M. J. P., & Schmidt, H. G. (2007). Improving meta-comprehension accuracy and self-regulation when learning to play a chess endgame: the effect of learner expertise. *European Journal of Cognitive Psychology*, 19(4-5), 671–688. <https://doi.org/10.1080/09541440701326204>.

- De Bruin, A. B. H., Thiede, K. W., Camp, G., & Redford, J. (2011). Generating keywords improves metacomprehension and self-regulation in elementary and middle school children. *Journal of Experimental Child Psychology*, 109(3), 294–310. <https://doi.org/10.1016/j.jecp.2011.02.005>.
- De Bruin et al., (n.d.) this issue
- Dunlosky, J., & Rawson, K. A. (2012). Underconfidence produces underachievement: inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, 22(4), 271–280. <https://doi.org/10.1016/j.learninstruc.2011.08.003>.
- Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, 28, 717–741. <https://doi.org/10.1007/s10648-015-9348-9>.
- Foster, N. L., Rawson, K. A., & Dunlosky, J. (2018). Self-regulated learning of principle-based concepts: do students prefer worked examples, faded examples, or problem solving? *Learning and Instruction*, 55, 124–138. <https://doi.org/10.1016/j.learninstruc.2017.10.002>.
- Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, 36(1), 93–103. <https://doi.org/10.3758/MC.36.1.93>.
- Griffin, T., Mielicki, M., & Wiley, J. (2019). Improving students' metacomprehension accuracy. In J. Dunlosky & K. Rawson (Eds.), *The Cambridge Handbook of Cognition and Education* (pp. 619–646). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108235631.025>.
- Hacker, D. J., & Bol, L. (2019). Calibration and self-regulated learning. In J. Dunlosky & K. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 647–677). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108235631.026>.
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research & Development*, 48(4), 63–85. <https://doi.org/10.1007/BF02300500>.
- Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development*, 53(3), 83–93. <https://doi.org/10.1007/BF02504800>.
- Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), 239–264. <https://doi.org/10.1007/s10648-007-9049-0>.
- Koriat, A. (1997). Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126(4), 349–370. <https://doi.org/10.1037/0096-3445.126.4.349>.
- Koriat, A., Nussinson, R., & Ackerman, R. (2014). Judgments of learning depend on how learners interpret study effort. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 1624–1637. <https://doi.org/10.1037/xlm000000>.
- Kostons, D., & De Koning, B. B. (2017). Does visualization affect monitoring accuracy, restudy choice, and comprehension scores of students in primary education? *Contemporary Educational Psychology*, 51, 1–10. <https://doi.org/10.1016/j.cedpsych.2017.05.001>.
- 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>.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>.
- Long, Y., & Aleven, V. (2016). Mastery-oriented shared student/system control over problem selection in a linear equation tutor. In A. Micarelli, J. Stamper, & K. Panourgia (Eds.), *Intelligent Tutoring Systems. ITS 2016. Lecture Notes in Computer Science* (Vol. 9684, pp. 90–100). Cham: Springer. https://doi.org/10.1007/978-3-319-39583-8_9.
- Long, Y., & Aleven, V. (2017). Enhancing learning outcomes through self-regulated learning support with an open learner model. *User Modeling and User-Adapted Interaction*, 27(1), 55–88. <https://doi.org/10.1007/s11257-016-9186-6>.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.
- Nugteren, M. L., Jarodzka, H., Kester, L., & Van Merriënboer, J. J. G. (2018). Self-regulation of secondary school students: self-assessments are inaccurate and insufficiently used for learning-task selection. *Instructional Science*, 46(3), 357–381. <https://doi.org/10.1007/s11251-018-9448-2>.
- Oudman, S., Van de Pol, J., Bakker, A., Moerbeek, M., & Van Gog, T. (2018). Effects of different cue types on the accuracy of primary school teachers' judgments of students' mathematical understanding. *Teaching and Teacher Education*, 76, 214–226. <https://doi.org/10.1016/j.tate.2018.02.007>.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive load approach. *Journal of Educational Psychology*, 84(4), 429–434. <https://doi.org/10.1037/0022-0663.84.4.429>.

- Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8.
- Paas, F., Tuovinen, J. E., van Merriënboer, J. J. G., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: optimizing learner involvement in instruction. *Educational Technology Research and Development*, *53*(3), 25–34. <https://doi.org/10.1007/BF02504795>.
- Panadero, E. (2017). A review of self-regulated learning: six models and four directions for research. *Frontiers in Psychology*, *8*, 422. <https://doi.org/10.3389/fpsyg.2017.00422>.
- Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., & Van Gog, T. (2017). Effects of performance feedback valence on perceptions of invested mental effort. *Learning and Instruction*, *51*, 36–46. <https://doi.org/10.1016/j.learninstruc.2016.12.002>.
- Raaijmakers, S. F., Baars, M., Paas, F., Van Merriënboer, J. J. G., & Van Gog, T. (2018a). 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.3392124>.
- Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., Van Merriënboer, J. J. G., & Van Gog, T. (2018b). 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>.
- Rawson, K., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology*, *19*(4-5), 559e579–559e579. <https://doi.org/10.1080/09541440701326022>.
- Redford, J. S., Thiede, K. W., Wiley, J., & Griffin, T. D. (2012). Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction*, *22*(4), 262–270. <https://doi.org/10.1016/j.learninstruc.2011.10.007>.
- Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skills acquisition: a cognitive load perspective. *Educational Psychologist*, *38*(1), 15–22. https://doi.org/10.1207/S15326985EP3801_3.
- Rienties, B., Tempelaar, D., Nguyen, Q., & Littlejohn, A. (2019). Unpacking the intertemporal impact of self-regulation in a blended mathematics environment. *Computers in Human Behavior*, *100*, 345–357. <https://doi.org/10.1016/j.chb.2019.07.007>.
- Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Metacognitive practice makes perfect: improving students' self-assessment skills with an intelligent tutoring system. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Artificial intelligence in education. AIED 2011. Lecture Notes in Computer Science, vol 6738* (pp. 288–295). Berlin: Springer. https://doi.org/10.1007/978-3-642-21869-9_38.
- Salden, R. J., Paas, F., & Van Merriënboer, J. J. (2006). Personalised adaptive task selection in air traffic control: effects on training efficiency and transfer. *Learning and Instruction*, *16*(4-5), 350–362. <https://doi.org/10.1080/09541440701326022>.
- Scheiter et al.,(n.d.) this issue
- Schleinschok, K., Eitel, A., & Scheiter, K. (2017). Do drawing tasks improve monitoring and control during learning from text? *Learning and Instruction*, *51*, 10–25. <https://doi.org/10.1016/j.learninstruc.2017.02.002>.
- Schmeck, A., Opfermann, M., Van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. *Instructional Science*, *43*(1), 93–114. <https://doi.org/10.1007/s11251-014-9328-3>.
- Schunk, D. (1987). Peer models and children's behavioral change. *Review of Educational Research*, *57*(2), 149–174. <https://doi.org/10.3102/00346543057002149>.
- Sweller, J., Ayres, P. L., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer. <https://doi.org/10.1007/978-1-4419-8126-4>.
- Tempelaar, D., Rienties, B., & Nguyen, Q. (in press). Individual differences in the preference for worked examples: lessons from an application of dispositional learning analytics. *Applied Cognitive Psychology*. <https://doi.org/10.1002/acp.3652>.
- Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, *28*(2), 129–160. [https://doi.org/10.1016/S0361-476X\(02\)00011-5](https://doi.org/10.1016/S0361-476X(02)00011-5).
- Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed- keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1267–1280. <https://doi.org/10.1037/0278-7393.31.6.1267>.
- Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*, *47*, 331–362. <https://doi.org/10.1080/01638530902959927>.
- Thiede, K., Oswald, S., Brendefur, J., Carney, M., & Osguthorpe, R. (2019). Teachers' judgments of student learning of mathematics. In J. Dunlosky & K. Rawson (Eds.), *The Cambridge Handbook of Cognition and Education* (pp. 678–695). Cambridge: Cambridge University Press.

- Van de Pol, J., de Bruin, A. B., Van Loon, M. H., & Van Gog, T. (2019). Students' and teachers' monitoring and regulation of students' text comprehension: effects of comprehension cue availability. *Contemporary Educational Psychology*, 56, 236–249. <https://doi.org/10.1016/j.cedpsych.2019.02.001>.
- Van de Pol et al., (n.d.) this issue
- Van Gog, T., & Paas, F. (2008). Instructional efficiency: revisiting the original construct in educational research. *Educational Psychologist*, 43(1), 16–26. <https://doi.org/10.1080/00461520701756248>.
- Van Gog, T., Paas, F., Van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, 11(4), 237–244. <https://doi.org/10.1037/1076-898X.11.4.237>.
- Van Gog, T., Kester, L., & Paas, F. (2011a). 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.172634>.
- Van Gog, T., Kester, L., & Paas, F. (2011b). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, 36(3), 212–218. <https://doi.org/10.1016/j.cedpsych.2010.10.004>.
- Van Gog, T., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and frequency of mental effort measurement: evidence in favor of repeated measures. *Applied Cognitive Psychology*, 26(6), 833–839. <https://doi.org/10.1002/acp.2883>.
- Van Gog, T., Rummel, N., & Renkl, A. (2019). Learning how to solve problems by studying examples. In J. Dunlosky & K. Rawson (Eds.), *The Cambridge Handbook of Cognition and Education* (pp. 183–208). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108235631.009>.
- Van Harsel, M., Hoogerheide, V., Verkoeijen, P. P. J. L., & Van Gog, T. (2019). Effects of different sequences of examples and problems on motivation and learning. *Contemporary Educational Psychology*, 58, 260–275. <https://doi.org/10.1016/j.cedpsych.2019.03.005>.
- Van Harsel, M., Hoogerheide, V., Verkoeijen, P. P. J. L., & Van Gog, T. (in press). Examples, practice problems, or both? Effects on motivation and learning in shorter and longer sequences. *Applied Cognitive Psychology*.
- Van Loon, M. H., de Bruin, A. B., Van Gog, T., Van Merriënboer, J. J., & Dunlosky, J. (2014). Can students evaluate their understanding of cause-and-effect relations? The effects of diagram completion on monitoring accuracy. *Acta Psychologica*, 151, 143–154. <https://doi.org/10.1016/j.actpsy.2014.06.007>.
- Van Merriënboer, J. J. G. (1990). Strategies for programming instruction in high school: program completion vs. program generation. *Journal of Educational Computing Research*, 6(3), 265–285. <https://doi.org/10.2190/4NK5-17L7-TWQV-1EHL>.
- Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs: Educational Technology Publications.
- Van Merriënboer, J. J. G., & Kirschner, P. A. (2013). *Ten steps to complex learning: A systematic approach to four-component instructional design* (2nd ed.). New York: Taylor & Francis.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.