Hide Cover

Journal of Educational Psychology
Issue: Volume 110(4), May 2018, p 578-595

Publication Type: [Problem Solving]
DOI: 10.1037/edu0000223
ISSN: 0022-0663
Accession: 00004760-201805000-00007
Keywords: self-explaining, self-monitoring, self-regulation, problem solving, worked examples

[Problem Solving]

Self-Explaining Steps in Problem-Solving Tasks to Improve Self-Regulation in Secondary Education

Baars, Martine; Leopold, Claudia; Paas, Fred
Section Editor(s): Graham, Steve Editor

Author Information

Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam
Centre for Teacher Education, University of Freiburg
Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, and Early Start Research Institute, University of Wollongong
We thank Tamara van Gog and Joachim Wirth for their suggestions about the design of the study. We thank the schools, teachers, and students for their participation in this research. Furthermore, we thank Faizah Moallemzadeh, Selma Meinderts, and Kevin Kamermans for their help with the data collection.
Martine Baars, Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands baars@fsw.eur.nl
Received November 25, 2016
revised June 23, 2017
Accepted June 26, 2017

Abstract

The ability to learn in a self-regulated way is important for adolescents' academic achievements. Monitoring one's own learning is a prerequisite skill for successful self-regulated learning. However, accurate monitoring has been found to be difficult for adolescents, especially for learning problem-solving tasks such as can be found in math and biology. This study investigated whether a self-explaining strategy, which has been found effective for improving monitoring accuracy in learning from text, can improve monitoring and regulation-choice effectiveness, and problem-solving performance in secondary biology education. In 2 experiments, one half of the participants learned to solve biology problems by studying video-modeling examples, and the other one half learned by giving step-by-step self-explanations following the video-modeling examples (Experiment 1) or by following the posttest problem-solving tasks (Experiment 2). Results showed that in contrast to earlier studies, self-explaining did not improve monitoring and regulation-choice effectiveness. However, the quality of self-explanations was found to be related to monitoring accuracy and performance. Interestingly, the complexity of the problem-solving tasks affected monitoring and regulation-choice effectiveness, and problem-solving performance. These results are discussed in relation to the cognitive demands that monitoring and regulating learning to solve problems combined with self-explaining pose on learners.

It is important to be able to learn in a self-regulated way during school years and beyond (cf. lifelong learning; OECD, 2015; Pellegrino & Hilton, 2013). Self-regulated learning (SRL) is viewed as proactive processes that students use to acquire academic skills, such as setting goals, selecting and deploying strategies, and self-monitoring one's effectiveness (Zimmerman, 2008). Keeping track of one's own learning process (i.e., monitoring) and using this information to regulate the learning process (i.e., control) are essential components of the feedback loop of self-regulated learning and thus prerequisites for successful self-regulation (Dent & Koenka, 2016; Winne & Hadwin, 1998). Studies on monitoring learning word pairs (for a review see, Rhodes & Tauber, 2011) and expository text (for a review see, Thiede, Griffin, Wiley, & Redford, 2009) have found several ways to support and improve monitoring judgments (e.g., judgments of learning, JOLs) and study control (e.g., restudy choices). For example, generative strategies such as making keywords (Thiede, Anderson, & Therriault, 2003) or self-explaining (Griffin, Wiley, & Thiede, 2008) were found to improve monitoring judgments when learning from expository texts.

Despite the fact that the ability to accurately monitor and regulate ones learning process might be even more important when learning to solve problem-solving tasks than learning word pairs or expository text, it has received less research attention. Studies on monitoring judgments (e.g., JOLs or self-assessments) and regulating learning to solve problems by making restudy choices have shown that even with instructional support strategies, monitoring judgments were not perfect and regulation choices were not always related to monitoring (Baars, van Gog, de Bruin, & Paas, 2014, 2016; Baars, Visser, van Gog, de Bruin, & Paas, 2013). Therefore, it is considered important to investigate the cognitive processes during the making of monitoring judgments and regulation choices when learning to solve problems and how students can best be supported in these processes. An important potentially successful strategy for supporting students in monitoring and regulating their learning to solve problems is by asking them to give self-explanations. Self-explaining when learning to solve problems has been found to improve problem-solving performance (i.e., the self-explanation effect, Bielaczyc, Pirolli, & Brown, 1995; Chi, Bassok, Lewis, Reimann, & Glaser, 1989). The current study investigates the effect of self-explaining on monitoring and regulation of learning to solve problems.

Monitoring and Control Processes When Learning to Solve Problems

Many studies have investigated monitoring and control processes to explore to what extent students are able to monitor and control their learning processes when learning word pairs or when learning from expository text (e.g., Rhodes & Tauber, 2011; Thiede et al., 2009). In these studies, participants studied word pairs or texts and were asked to give monitoring judgments for each word pair or text. Although it has received less attention, monitoring and regulating one's learning is also important when learning to solve problems. In problem-solving intensive domains, such as arithmetic, science, biology, economics, and math, well-structured problems are commonly used. These problems consist of a well-defined initial state, a known goal state, and can be solved using a constrained set of logical operators (Jonassen, 2011). An effective and efficient way of learning to solve problems when students have little or no prior knowledge, is by studying worked-out examples, which provide a step-by-step worked-out solution procedure to a problem (for reviews see, Atkinson, Derry, Renkl, & Wortham, 2000; Sweller, Van Merrienboer, & Paas, 1998; van Gog & Rummel, 2010).

To investigate monitoring when learning to solve problems, students are often asked to make monitoring judgments about their learning process. A monitoring judgment, such as a judgment of learning (JOL), can be prompted either prospectively or retrospectively (Baars, Vink, van Gog, de Bruin, & Paas, 2014). An example of a prospective JOL after having solved a problem-solving task is "How many steps of a similar problem do you expect to solve on a future test?" (Baars et al., 2013). An example of a retrospective JOL (i.e., self-assessment) after a problem-solving task is "How many steps do you think you have performed correctly?" (Kostons, van Gog, & Paas, 2012). The accuracy of these monitoring judgments is analyzed by comparing monitoring judgments to actual future test performance (i.e., prospectively) or to actual performance on the problem that was judged (i.e., retrospectively).

If students make accurate monitoring judgments, they can use this information to make accurate choices for the remainder of their learning process (i.e., control). Together, accurate monitoring and control processes can lead to better learning outcomes (e.g., Thiede et al., 2003). Control processes are typically investigated by considering regulation choices one makes during learning, for example by measuring which items are chosen for restudy (e.g., de Bruin, Thiede, Camp, & Redford, 2011; Thiede et al., 2003) or how study time is allocated to different tasks (Ariel, Dunlosky, & Bailey, 2009; Son & Metcalfe, 2000). Models of SRL (e.g., Winne & Hadwin, 1998; Zimmerman, 2008) propose that students base their regulation choices on information obtained from monitoring judgments. If monitoring is accurate, using information from monitoring judgments can lead to effective regulation choices. For example, it would be effective to spend some more time on an item (i.e., a regulation choice) if monitoring shows that it was not well understood or learned (i.e., discrepancy-reduction model of regulation, de Bruin et al., 2011; Nelson, Dunlosky, Graf, & Narens, 1994).

According to a review study by Schneider (2008), children's monitoring judgments and regulation skills improve during school years (for a review see, Schneider, 2008). Even primary schoolchildren were found to be able to monitor their learning (e.g., Krebs & Roebers, 2010; Roebers, Schmid, & Roderer, 2009). Also, in line with the findings for adults, primary schoolchildren made more accurate JOLs after a delay (e.g., Schneider, Vise, Lockl, & Nelson, 2000).

Yet, research on more complex tasks such as language tasks with children (de Bruin et al., 2011) and adult learners (Dunlosky & Lipko, 2007; Thiede et al., 2009) and problem-solving tasks with children (Baars et al., 2014) and adult learners (Metcalfe, 1986; Metcalfe & Wiebe, 1987) has shown that learners have difficulties to make accurate monitoring judgments and regulation choices. Fortunately, so-called generative strategies have been found to substantially improve monitoring accuracy when college students were learning from expository texts (Thiede, Dunlosky, Griffin, & Wiley, 2005; Thiede et al., 2009). Generative strategies are learning activities that induce learners to elaborate on the learning materials and generate new information (Fiorella & Mayer, 2016; Wittrock, 1992). Examples of generative strategies are generating keywords, making summaries or concept maps, giving self-explanations, practicing problems, or completing partially worked-out examples. Strategies that stimulate students to elaborate on, find meaningful patterns in, or rehearse materials can foster learning (Ormrod, 2016). Moreover, generative strategies can provide students with predictive cues on their comprehension.

According to the cue utilization framework proposed by Koriat (1997), students can use different cues to base their monitoring judgments on, and generative strategies can provide students with valid cues to monitor their learning. There are three types of cues: intrinsic, extrinsic, and mnemonic cues. The first type is based on the features of the task itself, which can show the learner the ease or difficulty of the task. The second type of cue consists of the conditions of the learning process (e.g., massed or spaced) or the encoding operations (e.g., level of processing). The third type of cue is based on an internal indication about how well a task is learned (e.g., cue familiarity or previous recall attempts). Both intrinsic and extrinsic cues can affect monitoring judgments directly but mnemonic cues can affect them indirectly (Koriat, 1997). Because generative strategies require students to process the learning materials on a deeper level, the cues students get from generative strategies are likely to be extrinsic according to the cue utilization framework.

Generative strategies that have been found to improve monitoring accuracy when learning from expository texts are delayed keyword generation (de Bruin, et al., 2011; Thiede et al., 2003), delayed summary writing (Anderson & Thiede, 2008), immediate self-explaining (Griffin et al., 2008), and immediate concept mapping (Redford et al., 2012). Further, de Bruin et al. (2011) and Thiede et al. (2003) showed that both for adults and children generating keywords after a delay also improved effectiveness of regulation choices. Keywords or summaries improved monitoring accuracy when they were made after all texts of a set of texts had been read but not when they were made immediately after reading each text. The authors argued that generating information after a delay is more diagnostic because it allows students to access their situation model representation of the text which is related to deep understanding (Anderson & Thiede, 2008). After a delay, the information will be accessed from long-term memory, which will provide cues for monitoring that are more predictive for future performance. In contrast, making a concept map or providing self-explanations immediately after reading a text did also improve monitoring accuracy, which suggests that these strategies gave students the opportunity to access their situation model immediately after reading a text (Thiede et al., 2009).

In line with the effect of generative strategies on monitoring accuracy for expository texts, solving practice problems after worked-out example study (Baars et al., 2014, 2016) and completion of partially worked-out examples (Baars et al., 2013) were found to improve monitoring accuracy when learning to solve problems. Also, when using practice problems after worked-example study in secondary education, regulation choices showed a stronger relation to JOLs for the conditions in which students were provided with practice problems. Therefore, it seems that generating (part of) the problem-solving solution during or after worked-out example study provided students with more diagnostic information (i.e., better cues) about their performance, which helped them to improve their monitoring judgments when learning to solve problems.

Although monitoring accuracy when learning to solve problems was affected by generative strategies such as solving practice problems (Baars et al., 2014, 2016) or completion problems (Baars et al., 2013), it was not perfect. With practice problems after worked-out examples as a generative strategy, primary schoolchildren still overestimated their future test performance (Baars et al., 2014). In secondary education relative accuracy of JOLs after practice problems was still only moderate (mean gamma correlation: 0.30, Baars et al., 2016). In addition, completing partially worked-out examples as an immediate strategy to help secondary education students monitor their learning led to underestimation of future test performance (Baars et al., 2013).

Possibly the effect of generative strategies on monitoring and regulation of learning is affected by the complexity of the learning materials. Monitoring (Griffin et al., 2008; van Gog, Kester, & Paas, 2011) and regulating (Dunlosky & Thiede, 2004) one's learning processes requires working memory resources, which are limited (Baddeley, 1986; Cowan, 2001; Sweller et al., 1998). Therefore, it is likely that the more complex a task becomes, the more working memory is challenged when attempting to understand the problem-solving procedure and to monitor it simultaneously. One defining element of task complexity is the number of interacting information elements in a task (Sweller et al., 1998). For example, in a well-defined hereditary problem with only one unknown generation (the child), there are less possible interacting elements compared with hereditary problems with two unknown generations (one of the parents and the child). Because of the particular strain on working memory when learning to solve new and complex problem-solving tasks, the cues for making monitoring judgments may be affected (Kostons, van Gog, & Paas, 2009).

The relation between task complexity and monitoring accuracy was demonstrated in previous studies on word pairs, that is, a negative relation between the difficulty of word pairs and monitoring accuracy (e.g., Griffin & Tversky, 1992; Koriat, Lichtenstein, & Fischhoff, 1980; Lichtenstein & Fischhoff, 1977). Similar effects of task complexity on monitoring accuracy were found when learning to solve problems by completing partially worked-out examples (Baars et al., 2013) or solving practice problems after worked-example study (Baars et al., 2016) in secondary education. Results of both studies showed that even though students learned from worked examples and monitored their learning by using a generative strategy, monitoring accuracy was better for less complex tasks.

In conclusion, even though generative strategies seem to improve monitoring accuracy when learning to solve problems, it remains unclear what information students use as a basis for their JOL after applying a generative strategy and how their JOL is related to their regulation choices. To know what information (i.e., cues) students use when monitoring and controlling their learning process when learning to solve new and complex problems, it is necessary to investigate the online processes when monitoring and controlling learning to solve problems. The generative strategy of self-explaining when learning to solve problems from worked examples (e.g., Chi et al., 1989; Renkl, 1999, 2002; Renkl, Stark, Gruber, & Mandl, 1998) does not only provide more insight into the online processes of monitoring and control, it is also expected to improve monitoring accuracy similarly to the results found with learning from text (Griffin et al., 2008). Therefore, we expect self-explaining to be a suitable generative strategy to investigate how monitoring accuracy can be improved and what mechanisms are responsible for this improvement when learning to solve problems.

Self-Explaining as a Generative Strategy

Chi et al. (1989) showed that college students learned most from worked examples about Newton's law when they explained the solutions presented in the worked examples. This was called the self-explanation effect. Because in the study by Chi et al. (1989), time on task differed between students who self-explained and students who did not, Renkl (1997, 1999) investigated how self-explanations affect learning to solve math problems from worked examples using fixed learning time for all students. It was found that the quality of self-explanations was significantly related to learning outcomes. The successful learners in the study could be described by the features of their self-explanations. That is, successful learners gave more principle-based explanations (i.e., identifying underlying domain principles and the meaning of operations), frequently explicated the goal-operator combinations (i.e., the goals that can be reached by performing a certain operation/action) and engaged in anticipative reasoning (i.e., anticipating on the next step).

However, not all students engage in self-explaining activities, and the quality of self-explanations differs among students (Chi et al., 1989; Renkl, 1997, 1999). Chi et al. (1989) found that "good" learners generated explanations about actions and their relations to principles in the materials, whereas "poor" students did not generate sufficient explanations. Similar to the results found by Chi et al., four clusters of self-explainers were identified in a study conducted by Renkl (1997): Two of these four clusters consisted of successful learners and two clusters consisted of unsuccessful learners. One of the successful clusters was labeled as *principle-based explainers* and the other cluster was labeled as the *anticipative explainers*. The principle-based explainers used mostly principle-based and goal-operator explanations, but they did not frequently use anticipative explanations. Yet, the anticipative explainers used mostly anticipative explanations and less principle-based or goal-operator explanations. The two unsuccessful clusters were called *passive explainers* and *superficial explainers*. The passive explainers showed very little self-explaining activity. The superficial explainers spent little time on each worked example and only used few explanations.

Chi, De Leeuw, Chiu, and LaVancher (1994) investigated whether self-explanations can also be prompted in a text learning context and whether self-explaining enhances learning outcomes. Self-explanations were prompted by asking the students to explain the meaning of each sentence they read. Chi et al. (1994) replicated earlier results and showed that high explainers learned and understood more of the learning materials compared with low explainers. Moreover, it was found that students who were prompted, gave more self-explanations and thereby learned and understood more. Self-explaining was assumed to aid understanding because of the following three characteristics of self-explaining. First, self-explaining is a constructive activity by which new declarative or procedural knowledge is created. Second, self-explaining supports the integration of newly learned materials with existing prior knowledge. If one does not integrate new knowledge but just adds new knowledge to prior knowledge, this could lead to incorrect knowledge or an incorrect mental model of the knowledge one needs to learn. The third characteristic of self-explaining counteracts this problem. That is, the third characteristic of self-explaining is that it concerns a continuous process of opportunities to detect and resolve conflicts between the mental model one is constructing and the knowledge one should gain from the learning material (Chi et al., 1994).

The benefits of self-explaining were found in various domains such as learning chess (de Bruin, Rikers, & Schmidt, 2007), solving science problems (Chi et al., 1989), calculating interest (Renkl et al., 1998), probability calculation (Renkl, 1997), and computer programming (Bielaczyc et al., 1995). Besides studies showing the self-explanation effect for adult learners (e.g., Chi et al., 1989; de Bruin et al., 2007; Renkl, 1997, 1999), this effect was also found for younger students in secondary education (e.g., Chi et al., 1994; Hilbert & Renkl, 2009) and even in kindergarten (Calin-Jageman & Horn Ratner, 2005). Chi et al. (1994) investigated the self-explanation effect with eight graders learning the circulatory system from text passages. Students were asked to explain what each sentence meant to them by giving them a general instruction about how to self-explain at the beginning of the procedure. It was found that self-explaining facilitated learning about the circulatory system. In the study by Hilbert and Renkl (2009), the authors asked 11th grade children to self-explain while learning to create concept maps from examples. The students were asked to self-explain by asking them to elaborate on how and why (parts of) their concept map would be applicable to the example. It was found that both concept mapping skills and concept knowledge about concept mapping were improved by self-explaining.

Next to facilitating learning, self-explaining can activate metacognitive processes to help students reflect on their understanding of the learning materials (e.g., Ainsworth & Loizou, 2003; Griffin et al., 2008). Griffin et al. (2008) investigated whether self-explaining does also improve monitoring accuracy. Because self-explaining is a constructive activity, during which new knowledge is generated and integrated into the students' mental model about the learning material (Chi et al., 1994), it was expected that self-explaining would increase access to valid cues about one's own learning and thereby improve monitoring accuracy. This is in line with the cue utilization model that states that extrinsic cues from deep processing of the materials can help students monitor their learning process (Koriat, 1997). Students were instructed to explain each sentence or paragraph in a text. Specifically, students were asked to explain and ask themselves the following questions: "What new information does this paragraph add? How does it relate to previous paragraphs?" "Does it provide important insights into the major theme of the text?" and "Does the paragraph raise new questions in your mind?" (Griffin et al., 2008, p. 97). It was found that self-explaining significantly improved monitoring accuracy. Furthermore, it was shown that readers do not necessarily have to construct good explanations; that is, just by creating explanations, students get access to cues about their understanding of the text. Thus, in line with the cue utilization framework, self-explaining seems to provide students with valid cues to base their monitoring judgments on.

The Current Study

To investigate whether and how monitoring judgments and regulation choices when learning to solve problems can be improved by having participants self-explain their learning process, we focus on the following research questions. Does self-explaining during learning from video-modeling examples improve the accuracy of monitoring judgments and regulation choices? (Question 1). On the basis of findings from learning with expository text (Griffin et al., 2008) and from problem-solving tasks (Chi et al., 1989; Renkl, 1997, 1999), it is expected that self-explaining when studying worked examples will improve monitoring judgments accuracy (Hypothesis 1a), subsequent regulation-choice effectiveness (Hypothesis 1b), and posttest performance (Hypothesis 1c). In addition, we aimed to investigate the quality of the self-explanations. What types of explanations do student give and how are they related to performance? (Question 2). On the basis of previous work (Chi et al., 1994; Renkl, 1997, 1999), we expected the principle-based, goal-operator, and anticipative explanations to be related to more accurate monitoring (Hypothesis 2a) and better performance (Hypothesis 2b). In addition, because working memory resources are limited (Baddeley, 1986; Cowan, 2001; Sweller et al., 1998) and monitoring one's learning processes requires working memory resources (Griffin et al., 2008; van Gog et al., 2011), the effect of complexity in problem-solving tasks on monitoring accuracy, regulation choice effectiveness, and performance for the different conditions will be explored (Question 3).

Experiment 1

Method

Participants and design

Participants were 82 Dutch secondary education students between 12 and 15 years old (M_{age} = 13.94, *SD* = 0.40, 48 females) who were recruited from schools in the North and South-west of the Netherlands. They were enrolled in their second year of preuniversity or higher education track (Voortgezet Wetenschappelijk Onderwijs [VWO] and Hoger Algemeen Voortgezet Onderwijs [HAVO] in the Dutch educational system). Parents of the students in the second year and the students themselves received a letter with information about the purpose of the study, an invitation to participate, and that asked for their consent. One student and her parents did not give consent, and therefore this student did not participate. Participants completed the experiment in their classrooms and were randomly allocated to either the self-explanation condition (*n* = 43) or the control condition (*n* = 39).

Materials

All materials were programmed and presented in Qualtrics Survey Software to students who were seated in the computer room in their school.

Pre- and posttest tasks

The pre- and posttest consisted of three problem-solving tasks about hereditary problems based on Mendel's laws (cf., Kostons et al., 2012). These tasks consisted of five steps: (1) determining genotypes from phenotypes, (2) constructing a family tree, 3) determining whether the reasoning should be deductive or inductive, (4) filling out the crosstabs, and (5) distracting the answer from the cross tabs (see Appendix A for an example). In Problem-Solving Tasks 1 and 2, students had to find the genotype of the child based on information about the parents (i.e., deductive). In terms of complexity, Problem-Solving Task 1 was easier than Problem-Solving Task 2 because in Task 1, the genotypes of the parents were homozygote, whereas in Task 2 they were heterozygote. Problem-Solving Task 3 was more complex than Problem-Solving Tasks 1 and 2 because the genotype of one of the parents had to be found based on information about the other parent and the child (i.e., inductive). These problem-solving tasks were cumulative in the sense that the correct answer to previous steps was necessary to find the correct answer to the next step.

Instructional video

In an instructional video, all the new biology concepts needed to learn to solve the problem (e.g., homozygote or heterozygote) were explained while the five steps of solving a hereditary problem were shown.

Video-modeling examples

Four videos showed how to solve a hereditary problem in a worked-out, step-by-step manner based on theories about example-based learning (e.g., van Gog & Rummel, 2010). In the videos, a model was thinking aloud about how to solve the problem and wrote down the solution step by step (the writing was shown as visual information; see Appendix B for an example). Two videos had a human female model, and two videos had a human male model, neither of whom was visible but who provided an auditory explanation about how to solve a problem. The average duration of the videos ranged between 1.34 and 2.18 min. In addition, in each video after solving the problem, the model made a written self-assessment and verbally explained it (cf. Raaijmakers et al., 2017). The hereditary problems explained in the videos had a similar solution procedure because in all four problems the goal was to find the genotype of the child on the basis of information about the parents (i.e., deductive). The surface features were different between the problems explained in the videos (e.g., length of eyelashes, hair color, freckles).

Self-explaining instruction

Students in the self-explaining condition had to explain the steps in solving the hereditary problem as was shown in the four videos directly after watching each video-modeling example. Before the students started self-explaining, they received a general instruction on how to self-explain the steps in the problem-solving task in the video. They were instructed to explain how and why one should solve a step in the problem-solving task from the example. Also, they were instructed to use their own words. Directly after each video-modeling example, students were instructed to give a written explanation of what one should do when performing each step by asking them: "Explain in your own words what you should do at Step 1."

Self-monitoring

After each posttest problem-solving task, students were asked to make a self-assessment as a measure of self-monitoring (cf. Baars et al., 2014). Self-assessments were provided on a 6-point rating scale, which asked participants to rate "How many steps did you perform correctly?", with the answer scale ranging from 0 to 5 (0 = *every step was wrong*, 5 = *every step was correct*). The internal consistency (Cronbach's alpha) over the three self-assessment was good with [alpha] = .76.

Regulation

Regulation was measured by asking participants to indicate whether they would need to practice the problem-solving task they just tried to solve again.

Measurement

Self-assessment accuracy was measured as bias and absolute accuracy (Schraw, 2009). Bias was measured as the difference between performance and self-assessment. The closer to zero the more accurate self-assessment was. A negative bias means that a student underestimated performance and a positive bias means a student overestimated performance. Absolute accuracy is measured as the square root of the squared bias. The lower absolute accuracy is, the more accurate self-monitoring (i.e., self-assessment) was. Bias and absolute accuracy per problem-solving task and the mean bias and absolute accuracy across the posttest were calculated for 40 students from the self-explaining condition and 38 students from the control condition. This was the case because 3 students from the self-explaining and 1 student from the control condition did not fill out all the monitoring judgments.

To calculate regulation-choice effectiveness, we used a gradual measure, which varies between 0 and 1, based on each possible combination of self-assessment (0 through 5) and regulation choice for a problem-solving task (0 or 1; cf. Baars et al., 2013). As can be inferred from Table 1, lower self-assessments combined with the choice not to restudy the task resulted in lower regulation choice effectiveness, whereas lower self-assessments combined with the choice to restudy the task resulted in higher regulation-choice effectiveness; similarly, higher self-assessments combined with the choice not to restudy the task resulted in higher regulation-choice effectiveness, whereas higher self-assessments combined with the choice to restudy the task resulted in lower regulation-choice effectiveness. This measure of regulation choices is based on the discrepancy-reduction model of regulation of study (de Bruin et al., 2011; Nelson et al., 1994). Mean regulation-choice effectiveness over the three problem-solving tasks was calculated. The higher this regulation-choice effectiveness was the better self-assessment and restudy choices corresponded.

Table 1
Scoring of Regulation Choice (Self-Assessment) Effectiveness per Problem

Self-assessment	No restudy (0)	Restudy (1)
0	0	1
1	.20	.80
2	.40	.60
3	.60	.40
4	.80	.20
5	1	0

1 Scoring of Regulation Choice (Self-Assessment) Effectiveness per Problem

Type of self-explanations

The self-explanations were coded using the types principle-based, goal-operator, and anticipative explanations from previous research (Chi et al., 1994; Renkl, 1997, 1999). Principle-based explanations show if learners assign meaning to a problem-solving step by identifying underlying principles of that step. Goal-operator explanations show if learners assign meaning to a step by explicating the goals they can achieve in that step. Anticipative explanations show if learners assign meaning to a step in problem-solving task by explaining how the step is connected to future steps in the problem-solving process (cf. Renkl, 1999).

To be able to distinguish between self-explanations of high quality (concerning the content and/or procedure of the problem-solving task) and those of low quality, we added two extra categories: *less helpful* and *not relevant*. The less helpful self-explanations were only loosely related to the problem-solving tasks. The not relevant self-explanations had nothing to do with the problem-solving task. Examples of each of these self-explanation categories are shown in Table 2. We included high- and low-quality self-explanations because research on generative strategies indicates close relations between quality-based differences in strategy use and learning performance (Glogger, Schwonke, Holzapfel, Nuckles, & Renkl, 2012; Leopold & Leutner, 2015).

Table 2
Examples of Self-Explanation Categories

Type of self-explanation	Example
Principle-based	In this step you draw a cross table with the father on the left side and the mother in the cells on top. Each gen of the genotype of the father and mother should be in a separate cell. You fill out all the cells by combining genes. If father has a T and mother a t, then the combination in the cell would be Tt.
	You can fill out the cross table using the genotypes from the parents.
Goal-operator	In this step you fill out the family tree.
	You decide how many cross tables you need and fill them in.
Anticipative	Fill out the family tree to be able to decide upon the direction.
	You fill out the cross table because you need to find out what the genotype of the unknown family member could be.
Less helpful	Determine
	Naming
Not relevant	That is something that makes us tired.
	Tok tok I am a rabbit.

2 Examples of Self-Explanation Categories

Two raters received instruction on the five types of explanations and rated 10% of the explanations to calculate interrater reliability. The interrater analysis revealed a kappa of 0.888 ($p < .001$), which means that the agreement among the raters were outstanding. One rater rated the remainder of the self-explanations. Because we expected the types principle-based, goal-operator, and anticipative explanations to help students monitor their learning process, we counted how often students used these types of explanations compared with less helpful and not relevant and calculated the correlation of that number with monitoring accuracy.

Procedure

After a pretest of three problem-solving tasks on genetics, students studied how to solve hereditary problems by watching an instructional video and four video-modeling examples of solving the task. Half of the students gave self-explanations per step of the problem-solving task after watching the four video-modeling examples (self-explanation condition) and the other half did not give self-explanations, but watched the video again (control condition). Finally, a posttest consisting of three problem-solving tasks about hereditary problems was administered. Then students made a self-assessment and a restudy choice after each problem-solving task.

Data analysis

Test performance

Performance on the pre- and posttest problems was rated per response to each step as either incorrect (0) or correct (1) according to a predefined answer model. No half credits or credits for procedure were granted. Thus pretest or posttest score could range between 0 and 5 credits per problem-solving task.

A principal component analysis (PCA) was conducted on the three items of the pretest and posttest separately to check whether these items measured the same construct (performance on the hereditary problems). For both PCAs, oblique rotation (promax) was used. On the pretest, the first item correlated significantly with the second ($r = .508$) but not with the third item ($r = .177$), which was the most complex one. The second and third item were significantly related to each other ($r = .395$). For the pretest, the Kaiser-Meyer-Olkin (KMO) measure verified the sampling adequacy for the analysis (KMO = .550), and all KMO values for individual items were >.53, which is above the acceptable limit of .5 (Field, 2009). Bartlett's test of sphericity, $[chi]^2(3) = 37.180$, $p < .001$, indicated that correlations between items were sufficiently large for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. One component had an eigenvalue over Kaiser's criterion of 1 and explained 57.89% of the variance. All items in the pretest cluster on one component which suggests that this component represents performance on the hereditary problems in the pretest.

On the posttest, the first item correlated significantly with the second ($r = .569$) and the third item ($r = .243$). The second and third item were significantly related to each other ($r = .532$). For the posttest, the KMO measure verified the sampling adequacy for the analysis (KMO = .553), and all KMO values for individual items were >.53, which is above the acceptable limit of .5 (Field, 2009). Bartlett's test of sphericity, $[chi]^2(3) = 57.868$, $p < .001$, indicated that correlations between items were sufficiently large for PCA. Again, one component had an eigenvalue over Kaiser's criterion of 1 and explained 63.66% of the variance. All items in the posttest cluster on one component which suggest that this component represents performance on the hereditary problems in the posttest. Table 3 shows the component score coefficient matrix for both the pretest and posttest.

Table 3
Summary of Exploratory Factor Analysis for the Pretest and Posttest in Experiment 1 and 2

Item	Experiment 1		Experiment 2	
	Pretest 1: Hereditary problems	Posttest 1: Hereditary problems	Pretest 1: Hereditary problems	Posttest 1: Hereditary problems
Problem 1	.434	.396	.450	.418
Problem 2	.499	.471	.487	.428
Problem 3	.372	.381	.465	.434
Eigenvalue	1.737	1.910	1.525	1.831
% of variance	57.89%	63.66%	50.85%	61.04%
α	.63	.70	.50	.68

3 Summary of Exploratory Factor Analysis for the Pretest and Posttest in Experiment 1 and 2

Results

Descriptive statistics

Table 4 shows means and standard deviations of pretest performance, posttest performance, self-assessment for both conditions, the number of restudy choices, bias, absolute accuracy, and regulation choices effectiveness per problem-solving task. As a check on successful randomization, a *t* test showed that both conditions did not differ on the pretest, $t(82) < 1$, $p = .400$.

Table 4
Means and Standard Deviations of Pretest Performance, Posttest Performance, Self-Assessment, Posttest Mental Effort per Problem-Solving Task, the Number of Restudy Choices, Bias and Absolute Accuracy per Problem-Solving Task per Condition for Experiments 1 and 2

Variable	Experiment 1		Experiment 2	
	Self-explaining	Control	Self-explaining	Control
Pretest performance	(n = 43)	(n = 39)	(n = 33)	(n = 27)
1	2.63 (1.53)	2.31 (1.67)	1.33 (1.11)	1.11 (1.19)
2	2.33 (1.69)	2.51 (1.50)	1.03 (.98)	1.00 (.73)
3	1.56 (1.05)	1.08 (1.01)	.82 (.77)	.67 (.73)
Posttest performance	(n = 43)	(n = 39)	(n = 33)	(n = 27)
1	4.19 (1.53)	3.92 (1.51)	4.19 (1.53)	3.92 (1.51)
2	4.30 (1.21)	4.15 (1.35)	4.30 (1.21)	4.15 (1.35)
3	3.30 (1.17)	2.95 (1.26)	3.30 (1.17)	2.95 (1.26)
Self-assessment	(n = 42)	(n = 38)	(n = 25)	(n = 27)
1	4.48 (1.13)	4.10 (1.52)	4.36 (.85)	4.67 (.73)
2	4.48 (.97)	4.13 (1.60)	4.50 (.91)	4.44 (1.16)
3	3.50 (1.40)	3.59 (1.39)	3.26 (1.13)	3.44 (1.19)
Number of restudy choices	(n = 42)	(n = 38)	(n = 25)	(n = 27)
1	7	6	3	5
2	8	5	2	2
3	19	13	7	14
Posttest mental effort			(n = 25)	(n = 27)
1	—	—	3.32 (2.50)	3.00 (2.11)
2	—	—	3.65 (2.38)	3.59 (2.08)
3	—	—	5.44 (1.67)	5.67 (2.00)
Bias	(n = 40)	(n = 38)	(n = 18)	(n = 27)
1	.20 (1.52)	.09 (1.84)	.50 (.99)	.81 (1.59)
2	2.23 (1.76)	1.58 (1.93)	.22 (1.52)	.33 (1.30)
3	2.10 (1.28)	2.50 (1.64)	.06 (1.83)	.52 (1.76)
Absolute accuracy	(n = 40)	(n = 38)	(n = 18)	(n = 27)
1	.80 (1.30)	1.13 (1.44)	.61 (.92)	1.11 (1.40)
2	2.38 (1.55)	2.05 (1.39)	.89 (1.23)	.85 (1.03)
3	2.15 (1.19)	2.66 (1.36)	1.50 (.99)	1.26 (1.32)
Regulation-choice effectiveness	(n = 40)	(n = 38)	(n = 18)	(n = 27)
1	.82 (.33)	.78 (.35)	.88 (.20)	.84 (.31)
2	.81 (.32)	.75 (.39)	.89 (.18)	.81 (.33)
3	.66 (.30)	.61 (.34)	.69 (.24)	.63 (.28)

4 Means and Standard Deviations of Pretest Performance, Posttest Performance, Self-Assessment, Posttest Mental Effort per Problem-Solving Task, the Number of Restudy Choices, Bias and Absolute Accuracy per Problem-Solving Task per Condition for Experiments 1 and 2

Monitoring accuracy

To analyze monitoring accuracy, bias in self-assessments during the posttest was investigated. A repeated-measures analysis of variance (ANOVA) with bias per complexity level (Task 1: easy, Task 2: medium, Task 3: complex) as the within-subject factor and condition (self-explaining vs. control) as the between-subjects factor showed that bias differed across the three complexity levels, $F(2, 152) = 53.50$, $p < .001$, $[\text{eta}]_p^2 = .413$. Post hoc analyses showed that bias in self-assessment on the problem-solving task of complexity Level 1 differed significantly from Level 2 ($p < .001$) and from Level 3 ($p < .001$). Whereas the bias in self-assessment between the second and third problem-solving task did not differ ($p = .212$). This means self-assessments were more accurate at complexity Level 1. In contrast to Hypothesis 1a, no effect of condition was found, $F(1, 76) < 1$, $p = .661$, $[\text{eta}]_p^2 = .003$. The interaction between bias across complexity levels and condition approached significance, $F(2, 152) = 2.77$, $p = .066$, $[\text{eta}]_p^2 = .035$.

Next to bias, absolute accuracy of self-assessment during the posttest was measured to investigate monitoring accuracy. A repeated measures ANOVA with absolute accuracy per complexity level (Task 1: easy, Task 2: medium, Task 3: complex) as the within-subject factor and condition (self-explaining vs. control) as the between-subjects factor showed that absolute accuracy differed as a function of complexity level, $F(2, 152) = 27.70$, $p < .001$, $[eta]_p^2 = .267$. Post hoc analysis showed that absolute accuracy in self-assessment on problem-solving task complexity Level 1 differs significantly from Level 2 ($p < .001$) and from Level 3 ($p < .001$), whereas the absolute accuracy in self-assessment between the second and third problem-solving task did not differ ($p = .927$). Again, this means self-assessments were more accurate at complexity Level 1. In contrast to hypothesis 1a, no effect of the between-subject factor condition was found, $F(1, 76) < 1$, $p = .382$, $[eta]_p^2 = .010$. There was no interaction between absolute accuracy across complexity levels and condition, $F(2, 152) = 2.17$, $p = .118$, $[eta]_p^2 = .028$.

Regulation-choice effectiveness

To analyze regulation effectiveness, the relation between self-assessments and restudy choices made after each problem-solving task in the posttest was calculated (cf. discrepancy-reduction model, de Bruin et al., 2011; Nelson et al., 1994). A repeated-measures ANOVA with regulation-choice effectiveness per complexity level (Task 1: easy, Task 2: medium, Task 3: complex) as within-subject factor and condition (self-explaining vs. control) as between-subjects factor was performed. The results showed that regulation-choice effectiveness differed between the three complexity levels, $F(2, 152) = 11.32$, $p < .001$, $[eta]_p^2 = .130$. Post hoc analysis showed that regulation-choice effectiveness on problem-solving task complexity Level 3 differed significantly from Level 1 ($p < .001$) and from Level 2 ($p = .001$), whereas the regulation-choice effectiveness between the first and second problem-solving task did not differ ($p = 1.00$). As can been seen in Table 3, regulation choice effectiveness was lower for the third complexity level, which means that the regulation choices showed less correspondence with the self-assessments. In contrast to Hypothesis 1b, no effect of the between-subject factor condition was found, $F(1, 76) < 1$, $p = .429$, $[eta]_p^2 = .008$. In addition, no interaction between regulation-choice effectiveness over complexity levels and condition was found, $F(2, 152) < 1$, $p = .942$, $[eta]_p^2 = .001$.

Types of explanations

In Table 5 the number of occurrences of the different types of self-explanations and the correlation between the five types, absolute accuracy of self-assessment, regulation-choice effectiveness, and posttest performance are shown. Because anticipative self-explanations only occurred once, this type of explanation was left out of the correlation analysis. Partly in line with Hypothesis 2a, only the self-explanations that were not relevant showed a significant correlation ($r = .486$) with absolute accuracy. The correlation is positive which means that the more "not relevant" self-explanations were given, the higher the absolute deviation between self-assessment and actual performance was (i.e., less accurate self-assessment). No significant correlations between the types of self-explanations and regulation choice effectiveness were found. Partly in line with Hypothesis 2b, posttest performance did correlate with both the goal-operator ($r = .581$) and the not relevant ($r = -.719$) type of self-explanations. The correlation with goal-operator was positive which means that with more goal-operator self-explanations, posttest performance was higher. In contrast, the correlation between not relevan' and posttest performance was negative. This means that with more "not relevant" self-explanations, posttest performance was lower.

Table 5
Pearson's Correlations Between the Five Types of Self-Explanations, Absolute Accuracy of Self-Assessment, Regulation-Choice Effectiveness and Posttest Performance for Experiment 1 and Experiment 2

	Experiment 1					Experiment 2			
Variable	Principle-based (n = 75)	Goal-operator (n = 708)	Less helpful (n = 6)	Not relevant (n = 70)	Variable	Principle-based (n = 16)	Goal-operator (n = 331)	Less helpful (n = 3)	Not relevant (n = 145)
Absolute accuracy (n = 43)	−.230	−.206	.032	.486**	Absolute accuracy (n = 18)	.348	−.083	−.106	.040
Regulation-choice effectiveness (n = 43)	.080	.001	.016	−.083	Regulation-choice effectiveness (n = 18)	−.191	.056	−.030	−.020
Posttest performance (n = 43)	.187	.581**	−.021	−.719**	Posttest performance (n = 33)	.068	.380*	.022	−.432*

* $p < .05$. ** $p < .001$.

5 Pearson's Correlations Between the Five Types of Self-Explanations, Absolute Accuracy of Self-Assessment, Regulation-Choice Effectiveness and Posttest Performance for Experiment 1 and Experiment 2

Performance

A repeated-measures ANOVA with test moment (pretest vs. posttest) as within-subject factor and condition (self-explaining vs. control) as between-subjects factor showed that all participants improved from the pretest to the posttest, $F(1, 80) = 154.23$, $p < .001$, $[eta]_p^2 = .658$. In contrast to Hypothesis 1c, no effect of condition was found, $F(1, 80) = 1.42$, $p = .237$, $[eta]_p^2 = .017$. Also, there was no interaction effect between test moment and condition, $F(1, 80) < 1$, $p = .858$, $[eta]_p^2 < .001$.

Discussion

The results of Experiment 1 showed consistent relations between problem complexity, monitoring and regulation choices. Performance in lower complexity problems was not only more accurately monitored than performance in higher complexity problems but was also related to more correct regulation choices. These results support the close connection between monitoring accuracy and regulation choices proposed in self-regulation theories (Thiede et al., 2003; Zimmerman, 2008). However self-assessments and restudy choices were not very accurate for more complex problem-solving tasks. Thus, accurate monitoring and regulation seemed to be difficult for high school students when learning to solve complex problems. In contrast to Hypotheses 1a and 1b, self-explaining how to solve biology problems after watching video-modeling examples on this subject did not improve participants monitoring or regulation choice effectiveness. That is, both self-assessment accuracy and regulation-choice effectiveness did not differ between the conditions in which participants had to self-explain the problem-solving procedure compared with watching the video-modeling example again. Hence, self-explaining as a generative activity did not seem to provide learners with more valid cues to judge there understanding of the problem-solving procedure.

Yet, a closer inspection of the type of self-explanations shows that "not relevant" self-explanations were significantly related to self-assessment accuracy. This relation indicates that self-assessment accuracy was lower when students made more not relevant self-explanations. Posttest performance was related to two types of self-explanations (goal-operator and not relevant). The goal-operator self-explanations were helpful for performance, whereas not relevant self-explanations were not. These findings seem to be in line with the cue utilization framework (Koriat, 1997), as students who gave not relevant self-explanations might have focused on invalid cues for monitoring.

Although the correlation coefficients between high-quality self-explanations and monitoring accuracy did not reach statistical significance, they pointed into the expected direction. A possible explanation for our findings that self-explaining did not significantly enhance monitoring, is that providing these self-explanations right after watching the video was not the right moment for the students to engage in self-explaining. That is, students were watching how to solve a problem but did not solve the problem themselves. According to the cue-utilization framework (Koriat, 1997), task experience is important as it provides cues on performance that can be used to monitor the learning process. Indeed, in studies in which students had to solve a practice problem after studying a worked example, monitoring accuracy improved presumably because the experience of solving the practice problem gave students valuable cues about their performance (Baars et al., 2014, 2016). Therefore, in Experiment 2 students were asked to self-explain the problem-solving procedure after trying to solve a posttest problem.

Experiment 2

In Experiment 2 the same research questions as in Experiment 1 were investigated. Yet, there was a slight difference in the moment during which participants had to self-explain their learning process. That is, students engaged in self-explaining during solving the posttest problem-solving tasks. In Experiment 1 the complexity of the problem-solving task affected monitoring and regulation choices. This was presumably the case because working memory resources are limited (Baddeley, 1986; Cowan, 2001) and monitoring one's learning processes also requires working memory resources (Griffin et al., 2008; van Gog et al., 2011). Similarly Winne (1995) pointed out that less experienced learners have not yet automatized monitoring processes and therefore monitoring tasks can even obstruct access to cognitive resources needed for solving the particular problem.

When a student has to solve problems, monitor his or her performance and on top of that self-explain what he or she is trying to learn, the demand on limited cognitive resources might be even higher. To investigate the cognitive load experienced by students in Experiment 2, the mental effort that students invested in solving the problem-solving tasks was also measured during the posttest. This way we could investigate if higher experienced mental effort is related to self-explaining quality, monitoring, and regulation choices accuracy. Studies on monitoring judgments when learning to solve problems found mental effort ratings to be related to monitoring judgments that could indicate mental effort is used as a cue to make monitoring judgments (Baars et al., 2013, 2016).

Like in Experiment 1, it was expected that self-explaining would improve monitoring accuracy (Hypothesis 1a), regulation choice effectiveness (Hypothesis 1b) and performance (Hypothesis 1c). Also, we aimed to investigate the quality of the self-explanations (Question 2). On the basis of previous work (Chi et al., 1994; Renkl, 1997, 1999), we expected the principle-based, goal-operator and anticipative explanations to be related to more accurate monitoring (Hypothesis 2a) and better performance (Hypothesis 2b). Furthermore, we expected that invested mental effort would be positively related to the monitoring accuracy measure (Hypothesis 3a). This means the lower mental effort would be, the lower the deviation between the monitoring judgments and the actual score would be, and thus the better monitoring accuracy would be. Invested mental effort was also expected to be negatively related to regulation choices effectiveness (Hypothesis 3b). The higher mental effort required, the lower the regulation choice effectiveness would be. Finally, invested mental effort was expected to be negatively related to the number of principle-based, goal-operator and anticipative self-explanations (Hypothesis 3c).

Method

Participants and design

Participants were 60 Dutch secondary education students (M_{age} = 14, SD = 0.18, 32 female) who were recruited from schools in the southwest of the Netherlands. They were enrolled in their second year of preuniversity or higher education track (VWO and HAVO). Parents of the students in the second year and the students themselves received a letter with information about the purpose and the content of the study, an invitation to participate, and that asked for their consent. Participants were randomly allocated to either the self-explanation condition (n = 33) or control condition (n = 27).

Materials

The materials used in Experiment 2 were similar to thosed used in Experiment 1, except for the mental effort ratings that were added to Experiment 2.

Mental effort ratings

Self-reported mental effort is a widespread measure within cognitive load theory research (for an overview see Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Whelan, 2006). It has been used in many different studies, both with children and adults (e.g., Baars et al., 2014; Paas, 1992; van Gog, Kirschner, Kester, & Paas, 2012). Mental effort invested during the posttest phase was measured on a 9-point scale, ranging from 1 (*very, very low mental effort*) to 9 (*very, very high mental effort*), after each test task. Because we had multiple ratings for mental effort during the posttest phase, we calculated an average score of mental effort for the posttest. Cronbach's alpha for the mental effort ratings on the posttest was .88. The mental effort ratings were sensitive to variations in complexity as evidenced by differences across the posttest tasks, $F(2, 86) = 51.52$, $p < .001$, $[\text{eta}]_p^2 = .55$. Mental effort ratings after the first posttest task were not significantly different from mental effort ratings after the second posttest task ($p = .066$). Both the first ($p < .001$) and the second ($p < .001$) mental effort ratings were different from the third mental effort ratings.

Self-monitoring

The Cronbach's alpha for the three self-assessment was .63. Bias and absolute accuracy were calculated for 18 students from the self-explaining condition and 27 students from the control condition. This was the case because 15 students from the self-explaining condition did not fill out all the monitoring judgments.

Procedure

The procedure was similar to the procedure in Experiment 1. Only the self-explanations were given at a different moment. Students explained each step of the problem-solving task for each of the three posttest problem-solving tasks after a mental effort rating and self-assessment were provided. Therefore, the control condition did not have to watch the video-examples again. They could continue to the next posttest problem-solving task without self-explaining.

Data analysis

Test performance

Like in Experiment 1, performance on the pre- and posttest problems was rated per step as either incorrect (0) or correct (1). Also for the pre- and posttest in Experiment 2, a PCA was conducted with oblique rotation (promax).

On the pretest the first item correlated significantly with the second ($r = .268$) but not with the third item ($r = .228$), which was the most complex one. The second and third items were significantly related to each other ($r = .291$). For the pretest, the KMO measure verified the sampling adequacy for the analysis (KMO = .610), and all KMO values for individual items were >.59, which is above the acceptable limit of .5 (Field, 2009). Bartlett's test of sphericity, $[\text{chi}]^2(3) = 10.854$, $p = .013$, indicated that correlations between items were sufficiently large for PCA. An initial analysis was run to obtain eigenvalues for each component in the data. One component had an eigenvalue over Kaiser's criterion of 1 and explained 50.85% of the variance. All items in the pretest cluster on one component which suggests that this component represents performance on the hereditary problems in the pretest.

On the posttest the first item correlated significantly with the second ($r = .393$) and the third item ($r = .413$). The second and third item were significantly related to each other ($r = .440$). Also for the posttest, the KMO measure verified the sampling adequacy for the analysis (KMO = .665), and all KMO values for individual items were >.65, which is above the acceptable limit of .5 (Field, 2009). Bartlett's test of sphericity, $[\text{chi}]^2(3) = 26.956$, $p < .001$, indicated that correlations between items were sufficiently large for PCA. Again, one component had an eigenvalue over Kaiser's criterion of 1 and explained 61.04% of the variance. All items in the posttest cluster on one component, which suggests that this component represents performance on the hereditary problems in the posttest. Table 3 shows the component score coefficient matrix for both the pretest and posttest.

Results

Descriptive statistics

Table 4 shows means and standard deviations of pretest performance, posttest performance, self-assessment, number of restudy choices, posttest mental effort, bias and absolute accuracy, and regulation choice effectiveness per problem-solving task for both conditions. Also, the number of restudy choices per problem-solving task is provided in Table 4. Both conditions did not differ on the pretest, $t(58) < 1$, $p = .437$.

Monitoring accuracy

To analyze monitoring accuracy, bias in self-assessments during the posttest was investigated. A repeated-measures ANOVA with bias per complexity level (Task 1: easy, Task 2: medium, Task 3: complex) as a within-subject factor and condition (self-explaining vs. control) as a between-subjects factor showed that bias did not significantly differ across the three complexity levels, $F(2, 86) = 1.42$, $p = .246$, $[eta]_p^2 = .032$. In contrast to Hypothesis 1a, no effect of condition was found, $F(1, 43) < 1$, $p = .461$, $[eta]_p^2 = .015$. There was no significant interaction between bias across complexity levels and condition, $F(2, 86) < 1$, $p = .790$, $[eta]_p^2 = .005$.

Next to bias, absolute accuracy of self-assessment during the posttest was also measured to investigate monitoring accuracy. A repeated measures ANOVA absolute accuracy per complexity level (Task 1: easy, Task 2: medium, Task 3: complex) as with a within-subject factor and condition (self-explaining vs. control) as a between-subjects factor showed that absolute accuracy differed across the three complexity levels, $F(2, 86) = 4.37$, $p = .016$, $[eta]_p^2 = .092$. Post hoc analysis showed that absolute accuracy in self-assessment on problem-solving task complexity Level 1 differs from Level 3 ($p = .059$) but this does not reach the significance level. The absolute accuracy on the second complexity level differed significantly from the third level ($p = .045$). Whereas the first and second problem-solving task did not differ ($p = 1.00$). As shown in Table 4, absolute accuracy is better at complexity Levels 1 and 2 compared with Level 3. In contrast to Hypothesis 1a, no effect of condition was found, $F(1, 43) < 1$, $p = .788$, $[eta]_p^2 = .002$. There was no interaction between absolute accuracy across complexity levels and condition, $F(2, 86) = 1.82$, $p = .169$, $[eta]_p^2 = .041$.

Regulation choices accuracy

To analyze regulation effectiveness, the accuracy of restudy choices made after each problem-solving task in the posttest was calculated. A repeated-measures ANOVA with regulation-choice effectiveness per complexity level (Task 1: easy, Task 2: medium, Task 3: complex) as a within-subject factor and condition (self-explaining vs. control) as a between-subjects factor was performed. The results showed that regulation-choice effectiveness differed across the three complexity levels, $F(2, 86) = 12.09$, $p < .001$, $[eta]_p^2 = .219$. Post hoc analysis showed that regulation-choice effectiveness on problem-solving task complexity Level 3 differs significantly from Level 1 ($p = .001$) and from Level 2 ($p = .002$), whereas the regulation-choice effectiveness between the first and second problem-solving task did not differ ($p = 1.00$). Regulation-choice effectiveness was higher for complexity Levels 1 and 2 compared with Level 3. In contrast to Hypothesis 1b, no effect of condition was found, $F(1, 76) < 1$, $p = .366$, $[eta]_p^2 = .019$. There was no interaction between regulation-choice effectiveness across complexity levels and condition, $F(2, 86) < 1$, $p = .936$, $[eta]_p^2 = .002$.

Types of explanations

In Table 5, the number of occurrences of the five types of self-explanations, absolute accuracy of self-assessment, regulation-choice effectiveness, and posttest performance are shown. The anticipative self-explanations were not used by the students and are therefore not present in this analysis. Partly confirming Hypothesis 2b, posttest performance correlates significantly with two types of self-explanations. The goal-operator correlates positively with posttest performance (r = .380), which means that when more goal-operator self-explanations were used, posttest performance was higher. Also, "not relevant" self-explanations correlated with posttest performance but they did so negatively (r = -.432). This means that when more "not relevant" self-explanations were used, posttest performance was lower.

Mental effort ratings

Both the self-explanations (M = 4.26, SD = 2.11) and the control (M = 4.09, SD = 1.85) condition invested a medium amount of mental effort which did not differ significantly between conditions, $t(43) < 1$, p = .773. Also, for both the self-explanations (r = -.566, p = .014) and the control (r = -.567, p = .002) condition, invested mental effort was significantly related to self-assessment score. Furthermore, in line with Hypothesis 3a and 3b, invested mental effort was significantly related to the absolute accuracy of monitoring judgments (r = .576, $p < .001$), regulation-choice effectiveness (r = -.342, p = .021), and performance (r = -.522, $p < .001$). In contrast to Hypothesis 3c, no significant correlations with one of the types of self-explanations were found.

Performance

A repeated-measures ANOVA with a within-subject factor moment (pretest to posttest) and a between-subjects factor condition (self-explaining vs. control) showed that, all participants improved between the pretest and the posttest, $F(1, 58) = 183.92$, $p < .001$, $[eta]_p^2 = .760$. In contrast to Hypothesis 1c, there was no effect of condition, $F(1, 58) = 1.47$, $p = .231$, $[eta]_p^2 = .025$. Yet, a significant interaction between test moment and condition was found, $F(1, 58) = 5.30$, $p = .025$, $[eta]_p^2 < .084$. A t test with pretest score as dependent variable and condition as the independent variable showed no differences between conditions, $t(58) < 1$, p = .437. A t test with posttest performance as dependent variable and condition as independent variable showed that the difference between the control (M = 3.63, SD = 1.12) and the self-explaining condition (M = 2.98, SD = 1.48) approached significance, $t(58) = -1.88$, p = .065.

General Discussion

The aim of the current study was to investigate self-explaining as a generative strategy to improve monitoring and regulation processes when learning to solve problems. Two experiments showed that the quality of students' self-explanations when either watching video-modeling examples or when solving problems differed between students. In line with findings by Renkl (1997), Chi et al. (1989), and the cue utilization framework (Koriat, 1997), the quality of self-explanations matters for monitoring accuracy and performance. In contrast to the findings by Griffin et al. (2008), monitoring and control processes were not improved for students who self-explained when learning problem-solving procedures or when performing problem-solving tasks. This result calls for new ways of supporting students when monitoring their learning process when learning to solve problems.

Monitoring judgment accuracy was measured with bias and absolute accuracy in both experiments. In Experiment 1 bias and absolute accuracy of the monitoring judgments were more accurate for the least complex problem-solving task. In the second experiment absolute accuracy showed the same pattern in which monitoring judgments seemed more accurate for the least complex problem-solving task, but this did not reach significance. These results are in line with earlier findings (Griffin & Tversky, 1992; Koriat et al., 1980; Lichtenstein & Fischhoff, 1977) and indicate that the complexity of the problem-solving tasks affects monitoring accuracy. Possibly, because the demand on working memory was smaller for the easier problems compared with the more complex problems, there could have been more room for monitoring the learning process when working on an easier problem.

It is worth noting that self-explaining did not aid monitoring as it did in the study by Griffin et al. (2008). Self-explaining might have been an extra task next to learning to solve problems which was too demanding for students to benefit from it. Next to learning how to solve the problems and monitoring this process, there might have been little room for providing self-explanations. Looking at the number of principle-based and anticipative explanations in Experiment 1 and the absence of anticipative self-explanations in Experiment 2 (see Table 5), it seems that it was hard for students to give these types of qualitative explanations during the learning process. From a cue utilization perspective (Koriat, 1997), the high-quality self-explanations provide students with valid cues to make monitoring judgments and base their regulation choices on. Yet, self-explanations of poor quality could provide students with invalid cues which will not help them to monitor their learning and could even harm this process (e.g., Thiede, Griffin, Wiley, & Anderson, 2010). Looking at the negative correlation between the 'not-relevant' self-explanations and monitoring accuracy in Experiment 1, it seems plausible that poor quality self-explanations provided invalid cues.

Additionally, students who participated in the current study never employed self-explanations to guide their learning process in the classroom before. Perhaps if students first receive a training about how to provide self-explanations and which types of self-explanations are useful for their learning process (i.e., principle-based, goal-operator and anticipative; cf. Renkl, 1999), self-explanations would indeed be more beneficial for their learning process. Also, in a study by Cho and Jonassen (2012), having high school students reflect on their self-explanations by comparing them with the instructional explanation improved the effectiveness of self-explaining. This could be an interesting avenue for future research on the effectiveness of self-explaining when learning to solve problems. Especially, by reflecting on self-explanations students could be stimulated to think about the metacognitive processes during learning, which could provide them with diagnostic cues for making monitoring judgments (cf. mnemonic cues; Koriat, 1997).

For regulation choices we found similar results as for monitoring judgments. Regulation-choice effectiveness was not improved in the self-explaining condition. According to models of SRL (Winne & Hadwin, 1998), it was to be expected that regulation would not be improved if monitoring was not affected to begin with. Yet, regulation choices were more accurate for the least complex problems indicating that for making accurate regulation choices the limited cognitive resources could also have played a role.

An important difference between the studies by Chi et al. (1989) and Griffin et al. (2008) and the current study is the type of materials that students had to study. Chi et al. found that "good" students who used explanations about actions and principles also showed more accurate monitoring when learning from worked examples. Griffin et al. found that self-explaining provided students access to more valid cues to monitor learning from explanatory texts. Perhaps, students are better able to give self-explanations about explanatory texts or worked examples (which also contain some text). To learn to solve problems, one needs both declarative knowledge and procedural knowledge. Possibly, the procedural knowledge needed to solve a problem is more difficult to provide self-explanations about which in turn affects monitoring and regulation negatively (cf. Winne, 1995). However, the video-modeling examples in the current study provided students with an example that is similar to the worked examples in the study by Chi et al. (1989). Moreover in the study by Chi et al., the students also had to use procedural knowledge to complete problem-solving tasks during the test. Yet, the sample size used in the study by Chi et al. was very small (n = 10), which could also be a reason why our findings deviate from their findings. In addition, Ainsworth and Loizou (2003) found that using diagrams reduced memory load and elicited more self-explanations compared with using text (without diagrams) when learning about blood circulation. Possibly, using diagrams when learning to solve biology problems encouraged students to access their situation model and by that helped students to give self-explanations. Finally, the monitoring judgment and measurement of monitoring accuracy differed between the current study and the study by Griffin et al. (2008). Griffin et al. used correlations between judgments about understanding/judgments of comprehension of a text and actual performance (i.e., relative measure), whereas in the current study the absolute deviation between the judgments about the amount of correct steps and actual correct steps (i.e., precision measure) was used. According to Schraw (2009), different judgments and measures of accuracy can reveal different results. Future research could investigate the role of the type of judgments (e.g., prospective vs. retrospective) and the measurement of accuracy when using generative strategies to improve monitoring judgment accuracy.

In line with earlier work on self-explanations (Chi et al., 1989; Renkl, 1997, 1999), the type of self-explanations affected both cognitive and metacognitive processes in the current study. Partly in line with Hypothesis 2b, in both experiments goal-operator self-explanations were positively correlated to performance. Moreover, the self-explanations that were not relevant were negatively correlated to performance. Yet, only in Experiment 1, the "not relevant" self-explanations were negatively related to the absolute accuracy of monitoring judgments (Hypothesis 2a).

In Experiment 2, invested mental effort was measured to investigate the relation between mental effort ratings, monitoring, regulation and self-explaining performance. Unlike metacognitive judgments, mental effort ratings are not related to a specific domain or task but can be seen as a basic feeling of workload (cf. physical effort). Gopher and Braune (1984; see also Paas & Sweller, 2012) have shown that persons can introspect on their cognitive processes and assign numerical values to the invested mental effort. In contrast, metacognitive judgments are based on a high level cognitive processing related to the content of a particular task: A person has to monitor whether he or she understands the task or performed it correctly. Interestingly, mental effort was related to metacognitive accuracy (the lower mental effort the better monitoring accuracy), suggesting that students used perceived mental effort as a possible cue for their metacognitive judgments. This finding is consistent with earlier studies on learning to solve problems and monitoring the learning process (Baars et al., 2013, 2016), thus providing additional support for the idea that students used their mental effort ratings as a cue for their monitoring judgments.

The current study has some limitations. Following up on previous work (Chi et al., 1989; Griffin et al., 2008; Renkl, 1997, 1999), this study aimed to investigate the effect of self-explanations on monitoring, regulation, and problem-solving performance in secondary education. Yet, because many learning activities take place in online learning environments in which students communicate by writing, we used written self-explanations in an online learning environment. This could have created a difference to previous studies in which spoken self-explanations were employed and consequently lead to different results. Future research could investigate possible differences between written and spoken self-explanations in the context of learning to solve problems. Furthermore, in the current study both experiments took place in a 50-min session fitting the timetable of the participating secondary schools. This also means that students only had 50 min to learn about the problem-solving tasks, to make monitoring judgments, regulation choices, and to self-explain. Future research could investigate whether a multiple session study in which students are first trained to perform all these actions would help students to benefit from self-explaining during the learning process. Also, the type of problem-solving task should be considered in future research. That is, in the current study a cumulative problem-solving task was used in which steps were either correct or incorrect. This means that students needed the answer to Step 1 to move on to the other steps, and no credits were given for using the right procedure with wrong answers. This might have affected the performance of students negatively. However, posttest performance was quite high in the current study, and therefore the type of task did not seem to have harmed student performance. In addition, a more elaborate way of measuring regulation of study could be used in such a multiple session study. Students could be allowed to actually restudy or practice the problem-solving task they choose again and asked to provide reasons for their regulation choices (i.e., interest or boredom). Finally, in the second Experiment, 15 students did not complete all the monitoring judgments which could have led to a bias in the results. However, these students did not differ in their pretest performance from the other students, $t(58) < 1$, $p = .766$.

To conclude, results from the current study showed that the complexity of the problems affected monitoring and regulation indicating that the demand on cognitive resources plays an important role in monitoring and regulating learning processes. Also the quality of self-explanations affected monitoring and performance. Yet, students in the self-explanation condition did not monitor, regulate, or perform better when learning to solve genetics problems. Our findings nuance existing findings (e.g., Griffin et al., 2008) because our findings show that self-explaining new and complex problem-solving tasks does not improve monitoring accuracy for secondary education students. From a cue utilization perspective (Koriat, 1997), self-explanations of low quality could have led to invalid cues, whereas high-quality self-explanations could have provided students with more valid cues. This means that the effect of self-explaining on SRL could be different for different types of task depending on the quality of self-explanations and complexity of the materials. Therefore, the findings in our study show new avenues for research which can advance the field. For example, training students how to provide high-quality self-explanations is a promising idea for future research. Furthermore, self-explaining could have created an extra demand on the limited resources students have which might have reduced the effectiveness of self-explaining to improve monitoring when learning to solve problems. Reflecting on self-explanations could theoretically help students to think about metacognitive processes after providing self-explanations about a problem-solving task. That way, reflecting on self-explanations could provide students with more diagnostic cues for making monitoring judgments without increasing the load during problem solving or providing self-explanations. Future research could follow up on these questions by using a set of tasks differing in complexity to investigate when self-explaining is beneficial for learning to solve problems and how we can scaffold monitoring processes when learning to solve problems.

References

Ainsworth S., Loizou A. T. (2003). The effects of self-explaining when learning with text or diagrams. Cognitive Science, 27, 669-681. SFX | Bibliographic Links | Library Holdings | [Context Link]

Anderson M. C., Thiede K. W. (2008). Why do delayed summaries improve metacomprehension accuracy? Acta Psychologica, 128, 110-118. SFX | Bibliographic Links | Library Holdings | [Context Link]

Ariel R., Dunlosky J., Bailey H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. Journal of Experimental Psychology: General, 138, 432-447. Ovid Full Text | SFX | Bibliographic Links | Library Holdings | [Context Link]

Atkinson R. K., Derry S. J., Renkl A., Wortham D. (2000). Learning from examples: Instructional principles from the worked examples research. Review of Educational Research, 70, 181-214. SFX | Bibliographic Links | Library Holdings | [Context Link]

Baars M., van Gog T., de Bruin A., Paas F. (2014). Effects of problem solving after worked example study on primary school children's monitoring accuracy. Applied Cognitive Psychology, 28, 382-391. SFX | Full Text | Bibliographic Links | Library Holdings | [Context Link]

Baars M., van Gog T., de Bruin A., Paas F. (2016). Effects of problem solving after worked example study on secondary school children's monitoring accuracy. Educational Psychology, 37, 810-834. [Context Link]

Baars M., Vink S., van Gog T., de Bruin A., Paas F. (2014). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. Learning and Instruction, 33, 92-107. SFX | Bibliographic Links | Library Holdings | [Context Link]

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, 395-406. SFX | Bibliographic Links | Library Holdings | [Context Link]

Baddeley A. D. (1986). Working memory. New York, NY: Oxford University Press. [Context Link]

Bielaczyc K., Pirolli P. L., Brown A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. Cognition and Instruction, 13, 221-252. SFX | Bibliographic Links | Library Holdings | [Context Link]

Calin-Jageman R. J., Horn Ratner H. (2005). The role of encoding in the self-explanation effect. Cognition and Instruction, 23, 523-543. SFX | Bibliographic Links | Library Holdings | [Context Link]

Chi M. T., Bassok M., Lewis M. W., Reimann P., Glaser R. (1989). Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13, 145-182. SFX | Bibliographic Links | Library Holdings | [Context Link]

Chi M. T., De Leeuw N., Chiu M. H., LaVancher C. (1994). Eliciting self-explanations improves understanding. Cognitive Science, 18, 439-477. SFX | Bibliographic Links | Library Holdings | [Context Link]

Cho Y. H., Jonassen D. H. (2012). Learning by self-explaining causal diagrams in high-school biology. Asia Pacific Education Review, 13, 171-184. SFX | Ovid Free Journal Service | Bibliographic Links | Library Holdings | [Context Link]

Cowan N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences, 24, 87-114. SFX | Ovid Free Journal Service | Bibliographic Links | Library Holdings | [Context Link]

de Bruin A. B., Rikers R. M., Schmidt H. G. (2007). The effect of self-explanation and prediction on the development of principled understanding of chess in novices. Contemporary Educational Psychology, 32, 188-205. SFX | Bibliographic Links | Library Holdings | [Context Link]

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, 294-310. SFX | Bibliographic Links | Library Holdings | [Context Link]

Dent A. L., Koenka A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. Educational Psychology Review, 28, 425-474. [Context Link]

Dunlosky J., Lipko A. R. (2007). Metacomprehension a brief history and how to improve its accuracy. Current Directions in Psychological Science, 16, 228-232. SFX | Bibliographic Links | Library Holdings | [Context Link]

Dunlosky J., Thiede K. W. (2004). Causes and constraints of the shift-to-easier-materials effect in the control of study. Memory & Cognition, 32, 779-788. [Context Link]

Field A. (2009). Discovering statistics using SPSS. London, UK: SAGE. [Context Link]

Fiorella L., Mayer R. E. (2016). Eight ways to promote generative learning. Educational Psychology Review, 28, 717-741. SFX | Full Text | Bibliographic Links | Library Holdings | [Context Link]

Glogger I., Schwonke R., Holzapfel L., Nuckles M., Renkl A. (2012). Learning strategies assessed by journal writing: Prediction of learning outcomes by quantity, quality, and combinations of learning strategies. Journal of Educational Psychology, 104, 452-468. Ovid Full Text | SFX | Bibliographic Links | Library Holdings | [Context Link]

Gopher D., Braune R. (1984). On the psychophysics of workload: Why bother with subjective measures? Human Factors, 26, 519-532. SFX | Bibliographic Links | Library Holdings | [Context Link]

Griffin D., Tversky A. (1992). The weighing of evidence and the determinants of confidence. Cognitive Psychology, 24, 411-435. [Context Link]

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, 93-103. [Context Link]

Hilbert T. S., Renkl A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. Computers in Human Behavior, 25, 267-274. [Context Link]

Jonassen D. H. (2011). Learning to solve problems: A handbook for designing problem-solving learning environments. New York, NY: Routledge. [Context Link]

Koriat A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. Journal of Experimental Psychology: General, 126, 349-370. Ovid Full Text | SFX | Bibliographic Links | Library Holdings | [Context Link]

Koriat A., Lichtenstein S., Fischhoff B. (1980). Reasons for confidence. Journal of Experimental Psychology: Human Learning and Memory, 6, 107-118. Ovid Full Text | SFX | Bibliographic Links | Library Holdings | [Context Link]

Kostons D., van Gog T., Paas F. (2009). How do I do? Investigating effects of expertise and performance-process records on self-assessment. Applied Cognitive Psychology, 23, 1256-1265. SFX | Full Text |

Bibliographic Links | Library Holdings | [Context Link]

Kostons D., van Gog T., Paas F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. Learning and Instruction, 22, 121-132. SFX | Bibliographic Links | Library Holdings | [Context Link]

Krebs S. S., Roebers C. M. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. The British Journal of Educational Psychology, 80, 325-340. SFX | Bibliographic Links | Library Holdings | [Context Link]

Leopold C., Leutner D. (2015). Improving students' science text comprehension through metacognitive self-regulation when applying learning strategies. Metacognition and Learning, 10, 313-346. [Context Link]

Lichtenstein S., Fischhoff B. (1977). Do those who know more also know more about how much they know. Organizational Behavior and Human Performance, 20, 159-183. SFX | Bibliographic Links | Library Holdings | [Context Link]

Metcalfe J. (1986). Feeling of knowing in memory and problem solving. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 288-294. [Context Link]

Metcalfe J., Wiebe D. (1987). Intuition in insight and noninsight problem solving. Memory & Cognition, 15, 238-246. [Context Link]

Nelson T. O., Dunlosky J., Graf A., Narens L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. Psychological Science, 5, 207-213. SFX | Bibliographic Links | Library Holdings | [Context Link]

OECD. (2015). OECD skills outlook 2015: Youth, skills, and employability. Paris, France: OECD Publishing. [Context Link]

Ormrod J. E. (2016). Human learning (7th ed.). Harlow, England: Pearson Education. [Context Link]

Paas F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. Journal of Educational Psychology, 84, 429-434. [Context Link]

Paas F., Sweller J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. Educational Psychology Review, 24, 27-45. SFX | Full Text | Bibliographic Links | Library Holdings | [Context Link]

Paas F., Tuovinen J. E., Tabbers H., Van Gerven P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. Educational Psychologist, 38, 63-71. SFX | Library Holdings | [Context Link]

Pellegrino J. W., Hilton M. L. (). (2013). Education for life and work: Developing transferable knowledge and skills in the 21st century. Washington, DC: National Academies Press. [Context Link]

Raaijmakers S. F., van Gog T., Baars M., Schaap L., Paas F., Van Merrienboer J. (2017). Training self-regulated learning skills with video modeling examples: Do task-selection skills transfer? Manuscript submitted for publication. [Context Link]

Redford J. S., Thiede K. W., Wiley J., Griffin T. D. (2012). Concept mapping improves metacomprehension

accuracy among 7th graders. Learning and Instruction, 22, 262-270. [Context Link]

Renkl A. (1997). Learning from worked-out examples: A study on individual differences. Cognitive Science, 21, 1-29. SFX | Bibliographic Links | Library Holdings | [Context Link]

Renkl A. (1999). Learning mathematics from worked-out examples: Analyzing and fostering self-explanations. European Journal of Psychology of Education, 14, 477-488. SFX | Bibliographic Links | Library Holdings | [Context Link]

Renkl A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. Learning and Instruction, 12, 529-556. SFX | Bibliographic Links | Library Holdings | [Context Link]

Renkl A., Stark R., Gruber H., Mandl H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. Contemporary Educational Psychology, 23, 90-108. SFX | Bibliographic Links | Library Holdings | [Context Link]

Rhodes M. G., Tauber S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. Psychological Bulletin, 137, 131-148. [Context Link]

Roebers C. M., Schmid C., Roderer T. (2009). Metacognitive monitoring and control processes involved in primary school children's test performance. The British Journal of Educational Psychology, 79, 749-767. SFX | Bibliographic Links | Library Holdings | [Context Link]

Schneider W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. Mind, Brain, and Education, 2, 114-121. [Context Link]

Schneider W., Vise M., Lockl K., Nelson T. O. (2000). Developmental trends in children's memory monitoring: Evidence from a judgment-of-learning task. Cognitive Development, 15, 115-134. SFX | Library Holdings | [Context Link]

Schraw G. (2009). A conceptual analysis of five measures of metacognitive monitoring. Metacognition and Learning, 4, 33-45. [Context Link]

Son L. K., Metcalfe J. (2000). Metacognitive and control strategies in study-time allocation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26, 204-221. Ovid Full Text | SFX | Bibliographic Links | Library Holdings | [Context Link]

Sweller J., Van Merrienboer J. J. G., Paas F. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10, 251-296. SFX | Full Text | Bibliographic Links | Library Holdings | [Context Link]

Thiede K. W., Anderson M. C. M., Therriault D. (2003). Accuracy of metacognitive monitoring affects learning of texts. Journal of Educational Psychology, 95, 66-73. Ovid Full Text | SFX | Bibliographic Links | Library Holdings | [Context Link]

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, 1267-1280. Ovid Full Text | SFX | Bibliographic Links | Library Holdings | [Context Link]

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. SFX | Bibliographic Links | Library

Holdings | [Context Link]

Thiede K. W., Griffin T. D., Wiley J., Redford J. S. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, A. C. Graesser (), Handbook of metacognition and self-regulated learning. Mahwah, NJ: Lawrence Erlbaum. [Context Link]

van Gog T., Kester L., Paas F. (2011). Effects of concurrent monitoring on cognitive load and performance as a function of task complexity. Applied Cognitive Psychology, 25, 584-587. SFX | Full Text | Bibliographic Links | Library Holdings | [Context Link]

van Gog T., Kirschner F., Kester L., Paas F. (2012). Timing and frequency of mental effort measurement: Evidence in favour of repeated measures. Applied Cognitive Psychology, 26, 833-839. SFX | Full Text | Bibliographic Links | Library Holdings | [Context Link]

van Gog T., Rummel N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. Educational Psychology Review, 22, 155-174. SFX | Full Text | Bibliographic Links | Library Holdings | [Context Link]

Whelan R. R. (2006). The multimedia mind: Measuring cognitive load in multimedia learning (pp. 1-178). (Doctoral dissertation). New York University. [Context Link]

Winne P. H. (1995). Inherent details in self-regulated learning. Educational Psychologist, 30, 173-187. [Context Link]

Winne P. H., Hadwin A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, A. C. Graesser (), Metacognition in educational theory and practice (pp. 277-304). Mahwah, NJ: Lawrence Erlbaum. [Context Link]

Wittrock M. C. (1992). Generative learning processes of the brain. Educational Psychologist, 27, 531-541. [Context Link]

Zimmerman B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. American Educational Research Journal, 45, 166-183. [Context Link]

[Context Link]

A still from one of the video modeling examples showing how the model is able to write out the problem-solving steps underneath the problem statement. [Context Link]

1. Chin dimple
A chin dimple is caused by a gene which will lead to a smooth chin in its dominant form (K) and a chin dimple in its recessive form (k). Two parents have a child. Father has a smooth chin and is heterozygote for this trait. Mother has a chin dimple. What will be the genotype of their child?

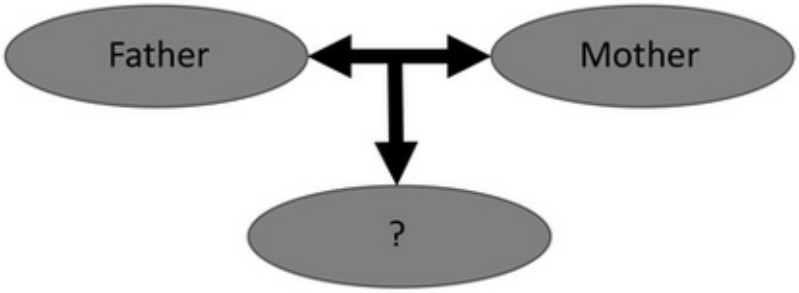
Step 1: Translate to genotypes

Father

Mother

Child

Step 2: Fill out the family tree



Step 3: Choose direction and number of cross tables

Choose the direction:

deductive

inductive

Choose the number of cross tables:

1 cross table

2 cross tables

3 cross tables

Figure. No caption available.

Step 4: Fill out the cross tables

Cross table 1

		mother	mother
Father			
Father			

Cross table 2

		mother	mother
Father			
Father			

Cross table 3

		mother	mother
Father			
Father			

Step 5: The answer

Find you answer in the cross tables and choose the right answer option below.

What is the genotype of the child?

- KK
- Kk
- kk
- KK or Kk
- KK or kk
- Kk or kk
- KK, Kk or kk

Figure. No caption available.

---

self-explaining; self-monitoring; self-regulation; problem solving; worked examples

---

IMAGE GALLERY

Select All

Export Selected to PowerPoint

Table 1
Scoring of Regulation Choice (Self-Assessment) Effectiveness per Problem

Self-assessment	No restudy (0)	Restudy (1)
0	0	1
1	.20	.80
2	.40	.60
3	.60	.40
4	.80	.20
5	1	0

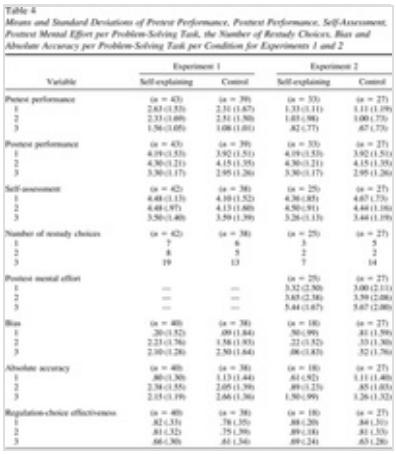
1 Scoring of Regulat...

2 Examples of Self-E...

Table 3
Summary of Exploratory Factor Analysis for the Pretest and Posttest in Experiment 1 and 2

	Experiment 1		Experiment 2	
Item	Pretest 1: Hereditary problems	Posttest 1: Hereditary problems	Pretest 1: Hereditary problems	Posttest 1: Hereditary problems
Problem 1	.434	.396	.450	.418
Problem 2	.499	.471	.487	.428
Problem 3	.372	.381	.465	.434
Eigenvalue	1.737	1.910	1.525	1.831
% of variance	57.89%	63.66%	50.85%	61.04%
α	.63	.70	.50	.68

3 Summary of Explora...



4 Means and Standard...

5 Pearsons Correlati...

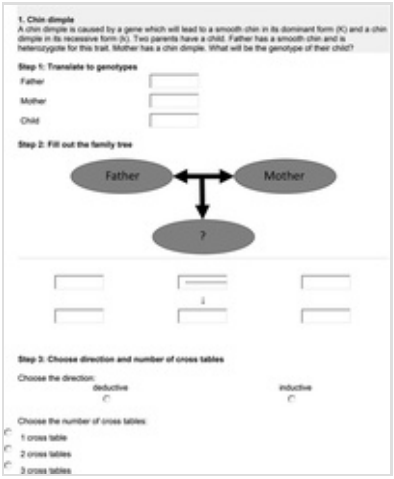


Figure. No caption a...

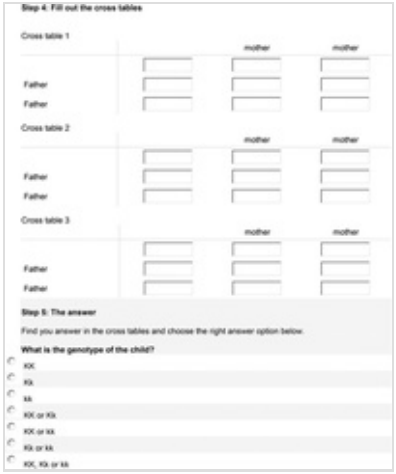


Figure. No caption a...

Back to Top