CrossMark

Accuracy of metacognitive judgments as a moderator of learner control effectiveness in problem-solving tasks

Loredana Mihalca[1] · Christoph Mengelkamp[2] ·
Wolfgang Schnotz[3]

Abstract A possible explanation for why students do not benefit from learner-controlled instruction is that they are not able to accurately monitor their own performance. The purpose of this study was to investigate whether and how the accuracy of metacognitive judgments made during training moderates the effect of learner control on performance when solving genetics tasks. Eighty-six undergraduate students solved self-selected genetics tasks using either a full learner control or a restricted learner control. Results indicated that learner control effectiveness was moderated by the absolute accuracy (i.e., absolute bias) of metacognitive judgments, and this accuracy was a better predictor of learning performance for full learner control than for restricted learner control. Furthermore, students' prior knowledge predicted absolute accuracy of both ease-of-learning judgments (EOLs) and retrospective confidence judgments (RCJs) during training, with higher prior knowledge resulting in a better absolute accuracy. Overall, monitoring guided control, that is, EOLs predicted time-on-task and invested mental effort regardless of the degree of learner control, whereas RCJs predicted the total training time, but not the number of tasks selected during training. These results suggest that monitoring accuracy plays an important role in effective regulation of learning from problem-solving tasks, and provide further evidence that metacognitive judgments affect study time allocation in problem solving context.

Keywords Learner control · Monitoring · Metacognitive judgments · Absolute accuracy · Problem solving

✉ Loredana Mihalca
mihalca@webster.ch

1   Psychology and Counseling Department, Webster University Geneva, Route de Collex 15,
     1293 Bellevue, Switzerland

2   Educational Media Department, University of Wuerzburg, Wuerzburg, Germany

3   Department of General and Educational Psychology, University of Koblenz-Landau, Landau,
     Germany

🖄 Springer

Managing one's own learning has become increasingly important in the context of self-regulated learning in classrooms as well as in informal learning settings (Paas et al. 2011). In classroom settings, the degree of learning regulation can be thought of as a continuum, ranging from teacher-regulated learning at one end, where students follow a predetermined learning path imposed by teachers, to self-regulated learning at the other end, where the student follows his/her own learning path (Loyens et al. 2008). As the level of students' expertise including domain-specific knowledge and metacognitive skills increases, a gradual shift from teacher-regulated learning to self-regulated learning can and should take place (Corbalan et al. 2006).

Self-regulated learning is frequently expected to be more effective than teacher-regulated learning because it allows for the instruction to be better adapted to the individual needs and preferences of students (Corbalan et al. 2008; Niemiec et al. 1996). Positive effects of self-regulated learning on performance are found, however, only if students have high prior knowledge (e.g., Kopcha and Sullivan 2007; Lawless and Brown 1997). Moreover, recent metaanalyses demonstrate that learner-controlled instruction leads to equal or only slightly better performance compared to program-controlled instruction (e.g., Karich et al. 2014; Kraiger and Jerden 2007).

A possible explanation for why students, especially low prior knowledge students, do not benefit from learner-controlled instruction is that they are unable to monitor accurately their own performance (Kostons et al. 2012). Research has shown that accurate monitoring positively influences self-regulation or control processes such as study choices and allocation of study time, which in turn improves performance (e.g., Baars et al. 2014a; Thiede et al. 2003; for an overview, see Dunlosky and Metcalfe 2009). In most studies on monitoring accuracy, self-regulation was limited to the selection of items for restudy and/or control of study time (for more details, see Kimball et al. 2012). To the best of our knowledge, no previous study has manipulated the level of self-regulation by experimentally varying the degree to which students were allowed to control their own learning process (i.e., study time and task selection) when problem-solving tasks are used.

Furthermore, the majority of research on monitoring accuracy has focused on learning from word pairs or expository texts rather than on learning from problem solving (de Bruin and van Gog 2012). Learning from problem-solving tasks with related conceptual structures might be based on other mechanisms than learning from unrelated word pairs or texts. In particular, when learning from multiple texts that are not conceptually related, students have to construct a mental representation of the individual texts that includes making sense of the sentences and their within-text connections (i.e., conceptual processing; see Pieger et al. 2016). Therefore, when making judgments about unrelated texts students have only to judge the difficulties in conceptual processing of the texts and their comprehension levels (Pieger et al. 2016), not to judge how these texts relate (i.e., between-texts connections). In contrast, when students learn from conceptually related problem-solving tasks, they have to judge not only the quality of the problem schemas they construct, but also how to use these schemas to solve problems with a similar conceptual structure (de Bruin and van Gog 2012). For making appropriate task selections students must be able to notice the analogies between problems in terms of their relevant aspects (i.e., the difficulty and support levels) rather than considering the irrelevant task aspects (i.e., the cover stories; see Quilici and Mayer 1996). In addition, students must determine whether the next tasks should be more or less difficult, and contain more or less support than the previous tasks (cf., Baars et al. 2014b).

Moreover, in most of the studies on monitoring accuracy when learning from problem solving (e.g., Baars et al. 2014a, b; for an exception, see Kostons et al. 2012), the task selections or restudy choices were limited and similar to those of text selection (i.e., participants had to indicate for three or six tasks whether they would choose to solve the tasks again or not). Therefore, the current study aimed at answering the question whether accuracy of monitoring moderates the effect of learner control on performance when students learn from a larger set of problem-solving tasks.

Self-regulated learning: prior knowledge and accuracy of metacognitive judgments

Models of self-regulated learning assume that accurate monitoring affects the effectiveness of self-regulation and, as a consequence, leads to better performance (e.g., Dunlosky et al. 2005; Nelson and Narens 1990; Winne and Hadwin 1998). In other words, when students are able to evaluate accurately how well they understood what has been presented, their subsequent choices about what to study next and for how long become more effective (e.g., Koriat 2012; Metcalfe and Kornell 2005; Son and Metcalfe 2000).

Monitoring has been examined prospectively by asking students to predict how easy it will be to learn specific information (i.e., ease-of-learning judgments [EOLs]; e.g., Leonesio and Nelson 1990; Son and Metcalfe 2000), to predict their own performance on a future test (i.e., predictions of performance; e.g., van Loon et al. 2014) or to predict their future recall or recognition of a recently studied item (i.e., judgments of learning [JOLs]; e.g., Kornell and Metcalfe 2006; Metcalfe and Finn 2008; Nelson and Leonesio 1988). Monitoring has also been examined retrospectively by asking students to rate their confidence of having learned successfully (i.e., retrospective confidence judgments [RCJs]; e.g., Dinsmore and Parkinson 2013; Hadwin and Webster 2013). Accuracy of these metacognitive judgments was estimated either in terms of the difference between judged and actual performance (i.e., *absolute accuracy* or *calibration*; see Alexander 2013) or in terms of the within-person correlation between judgments and performance, whereby the latter indicates how well students discriminate between better-learned and less-learned items (i.e., *relative accuracy* or *discrimination*; see Nelson 1996). In this study, we focused on absolute accuracy because it is crucial for the termination of learning when students are at liberty to choose the amount of learning material and the amount of study time invested (cf., Dunlosky and Metcalfe 2009).

Research has shown that metacognitive judgments are often inaccurate (e.g., Dunlosky et al. 2005; Dunlosky and Lipko 2007; Thiede et al. 2009). A possible explanation is that students do not base their judgments on predictive cues for their performance (see Koriat 1997). These cues may include, for example, domain knowledge (Griffin et al. 2009), cue familiarity (e.g., Metcalfe et al. 1993) or ease of processing (e.g., Begg et al. 1989). According to the domain familiarity hypothesis (Glenberg and Epstein 1987), students rely on prior knowledge-based cues rather than on cues derived from the actual learning situation (e.g., text-specific cues; Griffin et al. 2009). This is especially true for high prior knowledge students, who frequently show poorer relative monitoring accuracy compared to low prior knowledge students (see Glenberg and Epstein 1987). Griffin et al. (2009) found in a study measuring absolute as well as relative monitoring accuracy that high prior knowledge students had greater absolute accuracy than low prior knowledge students, whereas no effect of prior knowledge on relative accuracy was found. The authors explained this finding as a consequence of

high prior knowledge students' ability to use domain cues more effectively than low prior knowledge students, even if both groups used these cues to the same extent. Nietfeld and Schraw (2002) reported similar results in a study focusing on probability problem solving. In particular, students in the high-knowledge group showed higher absolute accuracy than students in the low-knowledge and mid-knowledge groups. The findings of these two studies indicate a positive relationship between prior knowledge and absolute monitoring accuracy. The open question is whether the abovementioned findings can be generalized when students are allowed to regulate their learning with problem-solving tasks.

Although there is a consistent finding that monitoring accuracy can be improved with specific instructional strategies (see Baars et al. 2014a, b, Experiment 2, for problem-based learning), the evidence for the relationship between monitoring accuracy and performance is far from conclusive (Dunlosky and Rawson 2012). Most studies that established a relationship between monitoring accuracy and performance focused on relative rather than absolute accuracy (e.g., Kornell and Metcalfe 2006; Thiede et al. 2003). Only a few studies have examined absolute accuracy (e.g., Baars et al. 2014a, b; Dunlosky and Rawson 2012; Pieger et al. 2016), but with mixed results. Some results suggest that increased absolute accuracy of judgments leads to better performance through effective allocation of study time (e.g., Dunlosky and Rawson 2012). Pieger et al. (2016), however, did not find a consistent relationship between judgments and study time allocation or text selection, and consequently no effects of better absolute accuracy on post-test performance were found. Studies using problem-solving tasks (Baars et al. 2014a, b) focused on the effect of absolute accuracy on control decisions (i.e., restudy choices), but did not consider the effects of these decisions on performance.

To investigate the effects of control decisions on performance, students need to be allowed to have control regarding what to (re)study and for how long. In the current study students were allowed to control the number and the order of tasks (i.e., study choices) as well as the invested time on solving tasks (i.e., study time), a choice that was not implemented in the abovementioned studies. More specifically, students were given either full or restricted control over learning based on the assumption that the degree of learner control may affect the relationship between monitoring accuracy, control processes, and performance. Kimball et al. (2012) tested this assumption in a series of experiments involving learning word pairs, by manipulating the degree of self-regulation. In the high self-regulation condition, students' restudy decisions were honored (i.e., students restudied only the items they had selected for restudying), in the low self-regulation condition, their decisions were dishonored (i.e., students had to restudy items they had not chosen or had to restudy items randomly selected by the computer), whereas in the no self-regulation condition, students were not allowed to restudy any items. The effects of self-regulation manipulation on recall performance were not more pronounced with higher than with lower relative monitoring accuracy. Put differently, relative accuracy of students' metacognitive judgments predicted recall performance but this prediction was not moderated by the manipulation of self-regulation. It is premature to draw any firm conclusions about the relationship between these variables, based on Kimball et al.'s (2012) findings, because in their study the focus was on relative rather than on absolute accuracy, and only one kind of metacognitive judgment (i.e., JOLs) was used. Furthermore, the authors used a restudy paradigm in which exactly the same learning material was used in the study and restudy phase. In educational settings including problem-solving tasks, the restudy phase consists of solving structurally similar, but not exactly the same tasks. Whether the findings of Kimball et al.'s (2012) study can be replicated under these conditions is an open question.

Monitoring and learner control

Effective self-regulated learning requires not only accurate monitoring of performance, but also responding adequately to the information provided by monitoring (i.e., effective control; Bjork et al. 2013). The two frequently investigated components of control are study choices and study time allocation (Metcalfe and Kornell 2005). Among the models of self-regulated learning dealing with the use of monitoring to regulate study choices and time allocation (for an overview, see Ariel et al. 2009), the discrepancy reduction model is of special importance for our study (Dunlosky and Hertzog 1998; Thiede and Dunlosky 1999). The model assumes that students choose to study and allocate more time to the most difficult items because they contain the largest discrepancies between actual and desired level of learning. Effective control of learning would therefore be indicated by a negative correlation between metacognitive judgments and study time, with items judged to be learned poorly receiving more study time than those judged as better learned (e.g., Nelson et al. 1994).

There is a good deal of evidence that when controlling their learning, students devote more time to the items judged to be more difficult, at least when there are no time constraints (see Son and Metcalfe 2000, for an overview). However, there is little evidence on whether students make an effective use of study time when allowed to control it. Koriat et al. (2006, Experiment 2) found that students in a self-paced condition allocated more study time to difficult items and less time to easy ones than students in a fixed-rate condition, which was yoked on the mean study time with the self-paced condition. However, this differential allocation of study time in the self-paced condition did not yield better recall performance compared to the fixed-rate condition. On the contrary, Tullis and Benjamin (2011) found that recognition performance was higher in a self-paced condition than in a fixed-rate condition even when total study time was equated. The advantage of self-pacing occurred, however, only for students who allocated more study time to objectively difficult items, which is consistent with the discrepancy reduction model. The findings of this study seem to support the assumption that students benefit from having control over time allocation but only when they are able accurately to monitor the item difficulty and allocate their study time appropriately (see also Son and Kornell 2008). Given the contradictory evidence it is difficult to conclude whether control of time allocation has potential benefits when learning word pairs, and whether this would also be the case when using more complex and ecologically valid requirements such as problem-solving tasks. Furthermore, an important factor that might affect the effectiveness of learner control (e.g., how long to study), but which has, to our knowledge, not yet been investigated when learning from problems, is the degree of control over learning.

Research on learner-controlled instruction has shown that high levels of control over learning (i.e., what tasks to be solved, in which order, and for how long; Corbalan et al. 2006) are detrimental to students with either low prior knowledge or poor metacognitive skills (Clark and Mayer 2003; Kicken et al. 2008). One explanation for this finding is that these students lack the understanding of what they know and how they learn (i.e., inability to monitor accurately their own performance), as well as the knowledge of performance standards (i.e., knowledge of what aspects of performance to evaluate, and what constitutes good performance on those aspects; van Gog et al. 2011). Inaccurate monitoring of their own performance makes students prone to choose inappropriate learning tasks or to terminate practicing too early because they erroneously believe they have reached the desired level of learning (see Kicken et al. 2008). Put differently, having good metacognitive skills seems to protect students from the negative effects of high levels of learner control (Clark and Mayer

2003). In order to help students with low metacognitive skills increase the accuracy of their monitoring and thus make effective task selections, the amount of learner control should be restricted (cf., Gerjets et al. 2009). The control over task selections should be limited in such a way to prevent students from working on more complex tasks before being able to master easier tasks (see van Merriënboer 1997), and to allow them to test whether they could perform tasks without any support (Kostons et al. 2012). This may contribute to a more accurate view of the performance level (e.g., not having reached the performance standards yet; Kicken et al. 2008), which may in turn help students to determine the appropriate level of difficulty and support of the next to-be-selected task(s).

Current study

The first aim of this study was to investigate whether and how the accuracy of metacognitive judgments moderates the effect of learner control on performance in the context of ecologically valid problem-solving tasks. In order to study this moderation, we manipulated the degree of learner control (full learner control vs. restricted learner control) and analyzed the effects of varying the degree of control on the relationship between accuracy of metacognitive judgments and performance in solving genetics tasks. The restricted control differed from the full control by asking students to solve at least one conventional problem (i.e., a traditional problem without any support) for each difficulty level before being allowed to proceed to another difficulty level in the learning environment. The rationale behind this restriction is based on assumptions of the four-component instructional design model (4C/ID model; van Merriënboer 1997) regarding the task selections. According to this model, learning tasks should be presented in a simple-to-complex sequence (i.e., from low to high difficulty) to reduce the cognitive load associated with the selection of future tasks. When the student's competence increases as a result of learning, his/her zone of proximal development (Vygotsky 1978) will be shifted accordingly towards higher difficulty levels, and the support given for a specific difficulty level would be gradually reduced and eventually faded out. That is, students should first receive high support (i.e., step-by-step demonstrations of how to solve a problem), then low support (i.e., a partial solution), and finally no support (i.e., conventional problems). Solving conventional problems can be seen as an independent assessment of students' individual performance indicating whether or not they are ready to perform more difficult tasks (see van Merriënboer and Sluijsmans 2009).

For students with low accuracy of metacognitive judgments, this restriction is expected to be beneficial as it ensures the selection of more appropriate tasks in terms of difficulty and support levels. Therefore, imposing restrictions on the self-regulated decisions of these students is expected to reduce the harmful effects of their inaccurate judgments on performance. However, for students with high accuracy of metacognitive judgments full control is assumed to be beneficial as these students are better able to select appropriate tasks based on their own judgments. This is why we hypothesized that restricted control would be more effective (i.e., higher post-test performance) than full control for students with low accuracy of their EOLs and RCJs, whereas the opposite would be true for students with high accuracy of their judgments (Hypothesis 1). Furthermore, we investigated whether students rely on prior knowledge-based cues when making metacognitive judgments in the context of problem solving. In line with the findings of Griffin et al. (2009), we predicted that students with higher prior knowledge would make more accurate use of knowledge-based cues than lower

prior knowledge students, which would result in a better absolute accuracy of their EOLs and RCJs (Hypothesis 2).

The second aim of this study was to examine the mechanisms underlying the effectiveness of learner control, that is, to investigate whether students base their control processes (e.g., time-on-task, invested mental effort) on metacognitive judgments (i.e., monitoring-based control hypothesis; Hypothesis 3). According to the discrepancy reduction model (e.g., Dunlosky and Hertzog 1998), students were expected to allocate more time to the tasks judged as being difficult (i.e., having low EOLs) as no time constraints were imposed. In addition, students were expected not only to spend more time but also to invest more mental effort for the tasks judged as being difficult (Hypothesis 3a). Moreover, students' RCJs were expected to predict the total number of tasks selected during training, as well as the total training time. In particular, it was assumed that students with higher RCJs would experience no need to study further, and thus would select a lower number of tasks during training and spend less total time in training (Hypothesis 3b). Finally, in line with the assumption that regulation affects test performance (Thiede et al. 2003), we hypothesized that control processes used during training would predict post-test performance. More specifically, students who solved more tasks and spent more time on training were expected to have higher post-test performance, regardless of the varying degrees of learner control (Hypothesis 3c).

Method

Participants and experimental design

There were ninety-one students of a small university in Germany who participated in this study. Five participants had to be excluded from the final analyses because they did not learn anything from the interaction with the learning environment, and performed worse than the chance level on the post-test. The remaining sample contained eighty-six participants, mainly females ($n = 62$, 72%) with a mean age of 23.10 years ($SD = 3.54$).

The topic to-be-learned by the participants was Mendel's laws of inheritance. All participants had at least some basic knowledge of Mendel's laws because this topic is a part of the biology curriculum for Gymnasium (i.e., secondary education level in Germany).

A pre-post-test experimental design with two conditions (i.e., varying degrees of learner control) was used, in which participants completed a prior knowledge test, and then solved the self-selected genetics tasks in the training phase, followed by a post-test. Participants were randomly assigned to a full control condition ($n = 41$), or a restricted control condition ($n = 45$). In the full control condition, students had complete responsibility for their learning process. More specifically, students received an overview of 45 genetics tasks (see Fig. 1) with an indication of their objective difficulty levels (from low to high) and support levels (high, low, and no support), and they could choose any task they wanted to solve, in which order, and for how long they wanted. The restricted control differed from the full control by asking students to solve at least one conventional problem (i.e., no support) for each difficulty level, correctly or incorrectly, before being allowed to proceed to another difficulty level. This restriction aimed to prevent students from selecting only genetics tasks with a high support level, which enabled them to test whether they are able to solve tasks without any support (for a similar procedure, see Kostons et al. 2010). Students were informed about this restriction at the beginning of training and were reminded of it each time they wanted to select a task from

Fig. 1 Selection table of all 45 genetics tasks

another difficulty level before solving a conventional problem in the current difficulty level. As compensation for their participation in the study, students received either 10€ or credit points towards their research experience requirement.

Materials

Electronic learning environment The computer-based learning environment developed for this study was based on Mihalca et al. (2011). All materials and instruments (e.g., prior knowledge test, post-test) were included in a database connected to the learning environment.

Genetics tasks The task database (i.e., the selection screen) consisted of 45 genetics tasks on the subject of heredity according to Mendel's laws (see Fig. 1), which were designed based on the 4C/ID model (van Merriënboer 1997). There were five levels of difficulty ordered in a simple-to-complex sequence, defined in cooperation with two domain experts based on several task characteristics: number of generations, number of possible correct solutions, and type of reasoning needed to find out the solutions (deductive and/or inductive). The tasks within one difficulty level had different surface features such as eye color or head shape of humans. Note that varying the surface features does not affect difficulty levels because these features are not relevant to how the task is solved (i.e., for the solution steps; Corbalan et al. 2006).

Within each difficulty level, support was gradually decreased (i.e., so-called completion strategy; van Merriënboer 1997) from (1) high support or incomplete worked examples which provided students almost all solution steps needed to solve the genetics tasks, except for the final solution itself, to (2) low support or completion problems which provided students only two out of five solution steps (see Fig. 2 for a screenshot of these two types of support), and to (3) no support or conventional problems in which students had to solve all the steps on their own. Note that all types of support were structured to emphasize the sub-goals by labeling the steps, as well as by visually isolating them (i.e., a sub-goal oriented type of support; see Mihalca et al. 2015). Moreover, students did not receive feedback concerning the correctness of their answers while solving the self-selected tasks.

Instruments

Performance measures The *prior knowledge test* and the *post-test* consisted of the same ten multiple-choice questions on the subject of heredity (i.e., Mendel's Laws), with four answer options of which one was correct (see Appendix). There were two questions for each difficulty level, that is, they required the same solution steps as the tasks in training but with different surface features and without any support (i.e., no solution steps were provided). The test has been successfully used in a prior study (Mihalca et al. 2015). The maximum score was 10 points, one point for each correct answer. The internal consistency measured with Cronbach's alpha was .44 for the prior knowledge test, and .75 for the post-test. As can be noticed, the reliability was good for the post-test, but low for the prior knowledge test. Nevertheless, we used the prior knowledge test scores for some of our analyses but readers have to keep in mind that non-significant results may be due to the low reliability of this test.
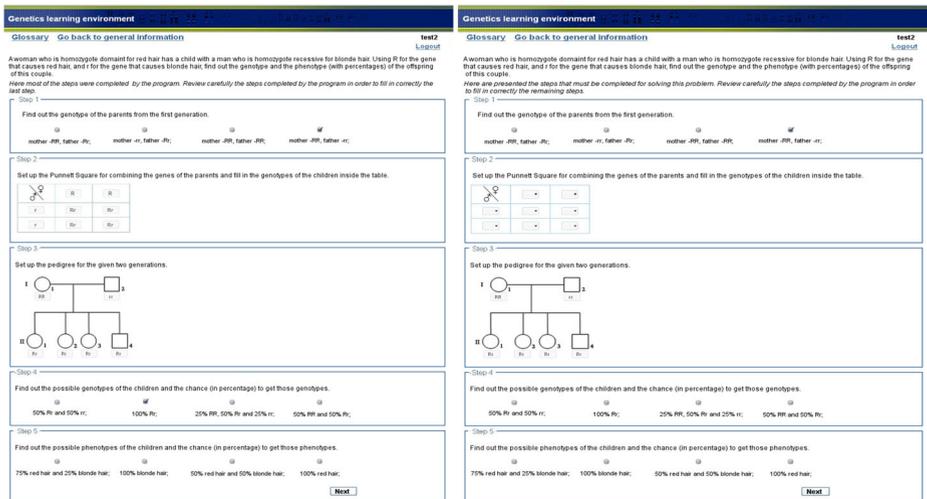

Fig. 2 Screenshots of an incomplete worked-out example (a), and a completion problem (b) from the electronic learning environment. In incomplete worked-out examples four steps out of five were completed by the program. In completion problems two steps out of five were completed by the program, and students had to fill in the other three steps

Perceived difficulty and mental effort rating scales The perceived task difficulty and invested mental effort were measured after each task during all the phases of the study (i.e., prior knowledge test, training, and post-test) on a 5-point rating scale with values ranging from 1 (very low) to 5 (very high). The item used to measure perceived task difficulty was "How difficult was it for you to solve the problem?" (cf., Ayres 2006), while the item for measuring invested mental effort was "How much mental effort did you invest in order to solve the problem?" (cf., Paas 1992; see also Mihalca et al. 2015).

Metacognitive judgments EOLs were obtained with an analogous rating scale from 0% confident to 100% confident using the following wording: "How confident are you that you will solve the upcoming task correctly?". For incomplete worked examples in which students had to complete only the final solution step (i.e., a multiple-choice question), the rating scale was from 25% to 100% because the probability of guessing correctly is 25% in the multiple-choice tasks with four answer options of which one is correct.

After solving each training task, participants were asked for RCJs using the following wording: "How confident are you that your answer is correct?" with an analogous rating scale from 0% confident to 100% confident (again for the incomplete worked examples the rating scale was from 25% to 100%).

Procedure

The experiment was conducted in a computer room in sessions that took between 45 and 100 min, with a maximum of 10 students per session. First, participants were given the prior knowledge test and then they read a basic introduction before the training phase started. The basic introduction included the main genetics concepts required for solving the tasks in training such as dominant and recessive genes, genotype, and phenotype, as well as an introductory text about Mendelian laws accompanied by examples of how to compute the probability of inheriting a specific trait using Punnett squares and genetic pedigrees. Participants were free to consult this basic introduction during the entire training session. During training, participants solved the self-selected genetics tasks from an overview of 45 tasks (see Fig. 1). The selection of tasks depended on the experimental conditions. After each selected task, but before the task was presented on the computer screen participants' EOLs were obtained. Furthermore, after solving each task in training, participants had to rate their confidence in the accuracy of the provided answers (i.e., RCJs), as well as their perceived task difficulty and invested mental effort before the program would let them proceed. At the end of the training phase, participants made a global JOL. However, as no learning took place after making this type of judgment, its accuracy is not important for performance, and hence data on global JOL are not reported. Participants could solve a maximum of 20 training tasks, and had to solve at least five tasks before they were allowed to terminate training. After training, participants performed the post-test, and the confidence in the accuracy of their answers (i.e., RCJs), as well as the perceived task difficulty and invested mental effort were rated after each post-test task.

Results

For all analyses the Type I error rate was set to .05. Cohen's (1992) taxonomy of effect sizes was used to classify effects as small, medium or large corresponding to values of .01, .06, .14

for (partial) eta-squared, respectively. To test interactions between conditions and continuous variables such as prior knowledge, we used moderated regression analyses following the procedures suggested by Cohen et al. (2003). Prior to conducting regression analyses, we checked whether the assumptions of regressions were met, that is, whether the distribution of the standardized residuals was normal by plotting histograms and q-q-plots and whether all tolerance values were above .10.

Preliminary analyses

To check whether there were significant differences between the two conditions before training, we conducted a MANOVA with performance, perceived task difficulty and mental effort invested during the prior knowledge test as dependent variables. Box's M test indicated no violation of the equality of covariance matrices, and Levene's test showed no significant inequality of variances between conditions for any of the dependent variables. The results of the MANOVA showed a significant effect of condition on the dependent variables using Pillai's trace, $V = .098$, $F(3, 82) = 2.98$, $p = .036$, $\eta_p^2 = .10$. However, univariate ANOVAs showed no significant effect on any of the dependent variables, except for a tendency of invested mental effort to be higher for the full control compared to the restricted control condition, $F(1, 84) = 3.89$, $p = .052$, $\eta^2 = .04$. The descriptive statistics are given in Table 1, and the correlations between the variables are presented in Table 2.

Accuracy of metacognitive judgments and performance in the post-test

Before testing our first hypothesis we checked whether students improved their performance from the prior knowledge test to the post-test, and whether these improvements were different between conditions. Therefore, a mixed ANOVA using condition as the between-subject

Table 1 Descriptive statistics for prior knowledge test, training, and post-test by condition

	Full learner control ($n = 41$)		Restricted learner control ($n = 45$)	
	M	SD	M	SD
Prior knowledge test				
Performance[a]	48.5	19.7	55.3	18.7
Perceived difficulty[b]	3.69	.66	3.60	.78
Mental effort[b]	3.53	.80	3.22	.63
Training				
Absolute bias of EOLs	39.7	19.5	35.7	20.1
Absolute bias of RCJs	30.7	25.0	22.7	21.3
Time on training[c]	1190	562	1330	605
Post-test				
Performance[a]	72.4	24.6	76.9	20.7
Perceived difficulty[b]	3.15	0.65	2.89	0.69
Mental effort[b]	3.35	0.81	3.02	0.89

EOLs Ease-of-Learning Judgments, *RCJs* Retrospective Confidence Judgments

[a] percentage of correct responses

[b] from 1 (very low) to 5 (very high)

[c] in seconds

Table 2 Pearson correlations between variables

	1	2	3	4	5	6
Prior knowledge test						
1. Performance	–					
2. Perceived difficulty	−.43**	–				
3. Mental effort	.07	.37**	–			
Training						
1. Absolute bias of EOLs	−.45**	.22*	−.24*	–		
2. Absolute bias of RCJs	−.58**	.31**	−.16	.62**	–	
3. Time on training	.28**	−.13	−.12	−.12	−.07	–
Post-test						
1. Performance	.52**	−.23*	.08	−.49**	−.59**	.11

EOLs Ease-of-Learning Judgments, *RCJs* Retrospective Confidence Judgments

*p < .05. ** p < .01

factor, and performance in the prior knowledge test and post-test as the within-subject factor was conducted (see Table 1 for the descriptive statistics). Results showed a significant effect for the increase in performance from the prior knowledge test to the post-test, $F(1, 84) = 102.40$, $p < .001$, $\eta_p^2 = .55$, but no significant effect of condition, $F(1, 84) = 2.05$, $p = .156$, $\eta_p^2 = .02$, or of the interaction between the two factors, $F(1, 84) = 0.27, p = .603, \eta_p^2 < .01$, was found.

For testing the hypothesis that the effect of varying degrees of learner control on performance is moderated by the accuracy of EOLs and RCJs in training (Hypothesis 1), we calculated the absolute bias for each of these two types of judgments. Absolute bias was calculated using the computation formula for absolute accuracy by Schraw (2009), and then taking the square root to compensate for the squaring in Schraw's formula. The resulting formula is:

$$\text{absolute bias} = \sqrt{\frac{1}{N} \sum_{i}^{N} (c_i - p_i)^2},$$

where c corresponds to the metacognitive judgments (i.e., EOLs and RCJs), p corresponds to the performance that has been judged, and N is the number of judged tasks. This formula results in a measure that is not sensitive to the direction of the bias (i.e., over-or underconfidence), indicating only the amount of deviations between judgments and performance scores. A value of zero indicates no absolute bias of the metacognitive judgments, whereas a value of 1.0 indicates the maximum possible absolute bias. The direction of bias is not relevant in the current study, because students who are overconfident may choose too difficult tasks or prematurely terminate their study (e.g., Dunlosky and Rawson 2012). By contrast, students who are underconfident may select too easy tasks or waste time studying tasks they already know and, as a consequence, do not improve their learning (Kostons et al. 2010). Therefore, despite their different mechanisms both over-and underconfidence may lead to suboptimal decisions about control of learning, and thus they are harmful to performance.

After computing the absolute bias of EOLs, we conducted a moderated regression analysis using the values of absolute bias, the condition (i.e., dummy-coded using full control as the reference category), and the interaction between condition and absolute bias of EOLs as predictors, and performance in the post-test as the dependent variable. The predictors

explained 29% of the total variance in the post-test performance, $F(3, 82) = 10.90$, $p < .001$. There was no significant main effect of condition on post-test performance, $b = -16.23$, $p = .080$, but a significant main effect of the absolute bias of EOLs, $b = -.81$, $p < .001$, and a significant interaction between condition and absolute bias of EOLs, $b = .49$, $p = .025$. These results are illustrated in Fig. 3. The simple slopes were $b = -.81$, $p < .001$, and $b = -.32$, $p = .029$ for the full control and the restricted control condition, respectively. To calculate the regions of significance for which there were differences in post-test performance between the two conditions, we used the Johnson-Neyman technique (see Hayes 2013, for more details). The Johnson-Neyman technique is analogous to a simple effects analysis using ANOVAs, but it combines a random effect with a fixed effect. If simple effects of the predictor on the dependent variable are indicated by simple slopes, the simple effects for the difference between two groups are computed using the Johnson-Neyman technique. More precisely, this technique is used to calculate the region of the predictor for which there are significant differences between the two groups. Results showed that the post-test performance was significantly higher in the full control compared to the restricted control condition within the region of]57.5%, ∞[. In other words, students who had an absolute bias score of EOLs higher than 57.5% (i.e., low accuracy) obtained a better post-test performance in the restricted control compared to the full control condition. An absolute bias of EOLs higher than 57.5% was found for 19.8% of the sample (i.e., 17 students).

In the next regression analysis, the absolute bias of RCJs was used as a predictor instead of the absolute bias of EOLs. In total, 39% of variance in the post-test performance was explained by the predictors, $F(3, 82) = 17.60$, $p < .001$. The main effect of condition was not significant, $b = -11.21$, $p = .063$, whereas the main effect of absolute bias, $b = -.76$, $p < .001$, and the interaction between condition and absolute bias of RCJs, $b = .42$, $p = .015$, were significant. To illustrate these results, we plotted the regression lines for both conditions in Fig. 4. The simple slope for both the full control, $b = -.76$, $p < .001$, and the restricted control condition, $b = -.33$, $p = .010$, was significantly different from zero, indicating that the more accurate the RCJs (i.e., the lower the absolute bias), the better the post-test performance. The effect of absolute bias on post-test performance was significantly stronger for the full control compared to the restricted control condition. The region of significance was]57.6%, ∞[, that is, for students who had an
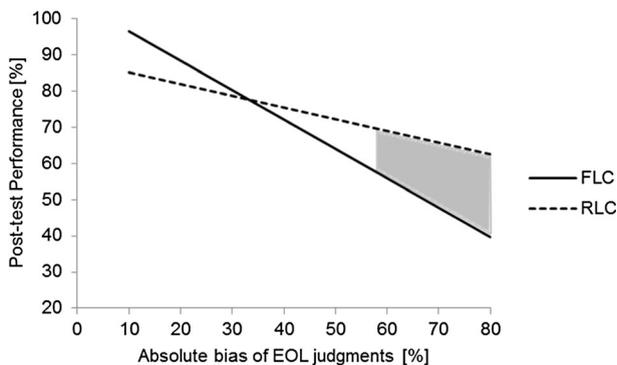

Fig. 3 Regression of post-test performance on absolute bias of EOLs in training. *Lower values* of absolute bias (i.e., approaching the value of zero) indicate higher accuracy, whereas *higher values* indicate over-or-underestimation of performance in training (i.e., deviation from accuracy). The *grey area* indicates the region of significance. EOL = Ease-of-Learning; FLC = Full Learner Control; RLC = Restricted Learner Control
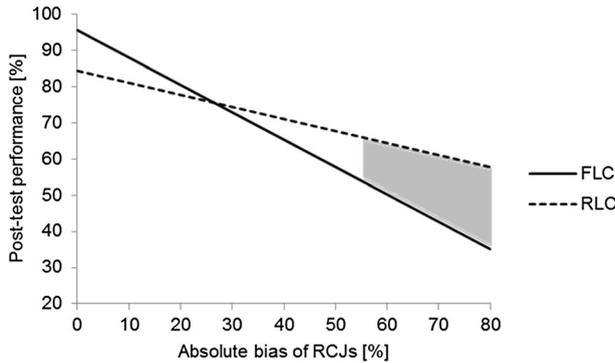
Fig. 4 Regression of post-test performance on absolute bias of RCJs in training. *Lower values* of absolute bias indicate higher accuracy, whereas *higher values* indicate over-or-underestimation of performance in training (i.e., deviation from accuracy). The *grey area* indicates the region of significance. RCJs = Retrospective Confidence Judgments; FLC = Full Learner Control; RLC = Restricted Learner Control

absolute bias score of RCJs higher than 57.6% (i.e., low accuracy), the post-test performance was better in the restricted control than in the full control condition. Students with an absolute bias score of RCJs higher than 57.6% accounted for 14.0% of the sample (i.e., 12 students).

Prior knowledge and accuracy of metacognitive judgments

According to the second hypothesis, higher prior knowledge students were expected to be more accurate in their metacognitive judgments made during training than lower prior knowledge students. To test this hypothesis, we conducted three regression analyses using prior knowledge, condition, and the interaction of prior knowledge and condition as the predictors, and the accuracy of EOLs or RCJs as the dependent variables (for descriptive statistics, see Table 1). As can be seen from Table 3, prior knowledge significantly predicted the absolute accuracy of both EOLs and RCJs. More precisely, the higher the prior knowledge the lower the absolute bias (i.e., higher absolute accuracy) of these two types of judgments. Thus, it can be concluded that higher prior knowledge was beneficial in making more accurate judgments. The effect of prior knowledge on the absolute accuracy of EOLs and RCJs was independent of the restrictions imposed in learner control as can be seen from the non-significant interaction effect between condition and prior knowledge.

Table 3 Regression of absolute accuracy of judgments in the training on prior knowledge and condition

	Absolute bias of EOLs		Absolute bias of RCJs	
	b	p	b	p
Constant	68.76	< .001	73.37	< .001
Restricted control	−15.75	.164	−23.31	.054
Prior knowledge	−0.60	< .001	−0.88	< .001
Restricted control x prior knowledge	0.29	.160	0.39	.077
R^2	.22	< .001	.37	< .001

EOLs Ease-of-Learning Judgments, *RCJs* Retrospective Confidence Judgments.

Metacognitive judgments and control processes

The results of the regression analyses indicated that absolute accuracy of both EOLs and RCJs was a predictor of post-test performance, and this accuracy was a stronger predictor for the full control than for the restricted control condition. As stated in our third hypothesis, these results may occur due to the use of metacognitive judgments as a basis for the control processes. Thus, it was expected that students' EOLs would correlate with the time spent on task and invested mental effort (Hypothesis 3a). To test this hypothesis, we first calculated gamma correlations between EOLs and time-on-task across all selected genetics tasks within each individual student. Next, the average across these gamma correlations between students was calculated. The results indicated that EOLs correlated significantly with the time-on-task, $\gamma = -.16$, $t(85) = -3.66$, $p < .001$. Specifically, the higher the EOLs, the less time students spent on tasks. The mean gamma correlations did not significantly differ between the two conditions, $t(84) = 0.83$, $p = .411$, $d = 0.18$, indicating that in both conditions students used their EOLs to allocate time on tasks.

In a similar way, we computed gamma correlations between EOLs and the mental effort invested during training. For the students who did not report different mental effort ratings across the training tasks, no gamma correlations were computed due to the lack of variance in these ratings. The results showed that the mean gamma correlation was significantly different from zero, $\gamma = -.16$, $t(67) = -2.32$, $p = .023$, indicating that the higher their EOLs, the less mental effort students reported to have invested into the completion of training tasks. No significant differences in mean gamma correlations between the two conditions were found, $t(66) = -0.76$, $p = .448$, $d = -0.19$.

Furthermore, we investigated whether students based the termination of study on their RCJs (Hypothesis 3b). For this purpose, we calculated a between-person correlation of the average of all RCJs across training tasks and the number of tasks selected during training. Because the number of tasks selected during training was not normally distributed, we computed a non-parametric gamma correlation. Results indicated that the average of RCJs across training tasks ($Md = 80.5$) did not predict the number of tasks selected in training ($Md = 7$), $\gamma = -.02$, $t(85) = -0.24$, $p = .808$. This result holds for both full control, $\gamma = -.07$, $t(40) = -0.52$, $p = .601$, and restricted control condition, $\gamma = -.07$, $t(44) = -0.67$, $p = .505$.

We also calculated the between-person correlation of the average of all RCJs across training tasks and the total training time (i.e., total time spent on solving all training tasks). Results showed that the average of RCJs across training tasks ($Md = 74.3$) did correlate significantly with the total time on training ($Md = 1094$ s) but only in the full control condition, $\gamma = .21$, $t(40) = 2.08$, $p = .038$, not in the restricted control condition ($Md = 1267$ s), $\gamma < .01$, $t(44) = 0.23$, $p = .982$.

Control processes and performance in the post-test

Based on Hypothesis 3c, we expected that the number of tasks selected in training and the total training time would predict the post-test performance. For testing this hypothesis, we conducted a regression analysis using the number of tasks selected during training, the condition, and the interaction between the number of training tasks and condition as the predictors, and post-test performance as the dependent variable. The predictors explained 7% of the variance in the post-test performance, but this variance was not significantly different from zero, $F(3, 82) = 2.08$, $p = .109$. Thus, the number of selected tasks did not predict the post-test performance in either full control or restricted control condition.

Finally, we conducted a regression analysis using time on training, condition, and the interaction of training time and condition as the predictors (for descriptive statistics, see Table 1). In total, 3% of the variance in the post-test performance was explained, but this variance was not significantly different from zero, $F(3, 82) = 0.70$, $p = .555$. Therefore, total training time did not predict post-test performance in either full control or restricted control condition.

Discussion

This study examined the relationship between accuracy of metacognitive judgments and performance in the context of learner-controlled problem solving, a topic that is still understudied in the literature. More specifically, we investigated whether the absolute accuracy (i.e., absolute bias) of EOLs and RCJs during training moderates the effect of varying degrees of learner control on performance in solving genetics tasks, and whether the magnitude of these judgments predicts control of learning, which in turn leads to better performance.

Accuracy of metacognitive judgments and post-test performance

The most important finding in this study is that the effectiveness of learner control interacted with the accuracy of students' EOLs and RCJs (Hypothesis 1). As predicted, students who had less accurate EOLs and RCJs during training benefitted more from the restricted control than from the full control. Another way to interpret this interaction is that absolute accuracy of EOLs and RCJs was a better predictor for the full control than for the restricted control condition. These results suggest that the implemented restriction helped students with less accurate judgments to avoid inappropriate task selections (see Kicken et al. 2008). In particular, this restriction might have ensured that students with low monitoring accuracy select tasks from a higher difficulty level only after having met the performance standards on a conventional problem at the current difficulty level (van Merriënboer and Sluijsmans 2009). By preventing students from proceeding too fast from one difficulty level to another one and from working only on tasks with high support, this restriction might have compensated for the bias of their judgments (e.g., by reducing the illusion of understanding caused by the tasks with high support or incomplete worked examples; Mihalca et al. 2015). Whether or not another type of restricted control would produce similar results should be addressed in future research.

Contrary to our expectations, both types of learner control were equally effective for students with more accurate metacognitive judgments. A possible explanation for this finding is that students who are more accurate in their judgments have a better understanding of how much they need to know to perform the tasks within a specific difficulty level, and when exactly they are ready to move to a higher difficulty level (cf., van Merriënboer and Sluijsmans 2009). As a result, they might be able to control their task selections more effectively and compensate for the lack of system control (Lee and Lee 1991).

Overall, the abovementioned findings provide further evidence that learner control is not beneficial for all students (e.g., Merrill 2002; Kopcha and Sullivan 2007). It seems to be beneficial only for students who are able to monitor their own performance with sufficient accuracy. Thus, the amount of learner control should be matched to the accuracy of students' metacognitive judgments (cf., Niemiec et al. 1996). In particular, full control should be provided only when students are sufficiently accurate in their metacognitive judgments. As suggested by our results, even minor restrictions in the degree of learner control (as for

example, not allowing students to choose another difficulty level before having solved a conventional problem at the current difficulty level) may be beneficial for students with inaccurate metacognitive judgments.

Prior knowledge and accuracy of metacognitive judgments

Our second hypothesis that students' prior knowledge predicts the absolute accuracy of their judgments was confirmed for both EOLs and RCJs. More specifically, higher prior knowledge students were less biased in their predictions (EOLs) and postdictions (RCJs) about their own performance. These findings are consistent with facilitative hypothesis (see Nietfeld and Schraw 2002) assuming that prior knowledge provides a conceptual basis for evaluating one's performance, which improves monitoring accuracy. Another possible explanation of the positive relationship between prior knowledge and monitoring accuracy would be that students with higher prior knowledge have, due to their more automated schemas, enough cognitive resources available for performing the learning tasks and accurately monitoring their own performance simultaneously (cf., van Gog et al. 2011).

A further explanation might be the hard-easy effect (e.g., Dunlosky and Metcalfe 2009; Lichtenstein and Fischhoff 1977) which shows that learners are underconfident with easy tasks and overconfident with difficult tasks. A learning task may appear to be easier than it actually is for higher prior knowledge students who are therefore underconfident, whereas the opposite applies to lower prior knowledge students who tend to be overconfident. Therefore, prior knowledge should negatively correlate with the bias of judgments. Students in our study showed only overconfidence (almost no underconfidence), and therefore the absolute bias and the bias of judgments were almost equal. Thus, the negative correlation between prior knowledge and the bias of judgments (i.e., the higher the prior knowledge the lower the overconfidence) may constitute an alternative explanation for our results.

In addition, our results indicated that prior knowledge indirectly influenced the impact of learner control on performance. In particular, higher prior knowledge students had more accurate EOLs and RCJs, and the accuracy of these judgments moderated the positive effects of learner control on performance. These results suggest that prior knowledge may effectively guide learner-controlled instruction because students with higher prior knowledge seem to be "better able to invoke schema-driven selections, wherein knowledge needs are accurately identified a priori and selections made accordingly" (Gall and Hannafin 1994, p. 222) than lower prior knowledge students. Further research is needed to disentangle both the direct interaction effects of prior knowledge and learner control on performance (i.e., an expertise reversal effect, Kalyuga et al. 2003; Schnotz 2010) and the indirect effects of prior knowledge on performance via the interaction of monitoring accuracy and learner control.

A limitation of the current study is that the relationship between prior knowledge and monitoring accuracy could be due to other hidden variables that differ between higher and lower prior knowledge students (see Rey and Buchwald 2011), which might explain the effects on monitoring accuracy. Therefore, future studies should manipulate stundents' prior knowledge experimentally.

Effects of monitoring on control

As stated in our third hypothesis, the impact of accuracy of metacognitive judgments on learner control effectiveness may be explained by the use of these judgments as a basis for

control processes. Indeed, the results indicated that students used their metacognitive judgments to make decisions about how much time and effort to invest in solving the tasks, but not for deciding how many tasks to solve during training. More specifically, the magnitude of EOLs predicted time-on-task and invested mental effort (Hypothesis 3a), that is, students allocated more time and invested more mental effort on the tasks that they perceived as being more difficult (i.e., low EOLs). These findings are consistent with the discrepancy reduction model (e.g., Thiede and Dunlosky 1999) indicating that perceived task difficulty is an important determinant of study time allocation and investment of mental effort in problem solving. No differences were found between the two types of learner control regarding the correlations between EOLs and time-on-task as well as invested mental effort. This means that regardless of whether students had high or restricted control over learning, their EOLs played a role in guiding their decisions about the time and the mental effort invested on tasks (cf., Son and Kornell 2008).

Furthermore, RCJs predicted the total time spent on training (Hypothesis 3b) but the correlation was in the opposite direction, that is, higher confidence judgments made after solving the tasks resulted in higher total training time. According to Koriat et al. (2014, p. 5) "when the regulation of effort is goal-driven, students invest more effort in studying the material until they reach a targeted degree of mastery." It is possible that students in our study invested more effort and hence more time solving the selected tasks, while reporting higher RCJs. However, this was the case only for the full control condition. In the restricted control condition, the total training time might have been mainly due to the conventional problems that had to be solved by the students within each difficulty level, rather than the effort invested in the self-selected tasks.

Moreover, students did not use their RCJs to decide about the number of tasks selected to-be-solved during training in either of the two learner control conditions. Task selection was operationalized in the current study as the total number of solved tasks, but these tasks were characterized by specific combinations of difficulty and support levels, as well as surface features. Therefore, students could follow multiple possible learning paths within one condition. A limitation of the present study is the inability to identify whether the learning paths followed by students consisted of appropriate tasks for them in terms of difficulty *and* support levels, and to what extent the selection of these individual learning paths was guided by metacognitive judgments.

It is also possible that RCJs had no predictive value for the number of tasks selected during training because there are other potential factors that might have influenced task selection, which are not related to metacognitive monitoring (see Son and Kornell 2008). In particular, task selection can be a function of students' learning goals (e.g., to obtain the minimum performance possible across all tasks, to solve a number of tasks above a specific threshold; cf., Son and Sethi 2006) or a function of their motivation and interest (cf., the findings by Son and Metcalfe 2000, showing that students spent more time on tasks perceived as pleasant rather than on those perceived as unpleasant). This latter explanation is in accordance with the findings of a study by Nugteren et al. (2015), in which participants reported that they selected the-to-be-solved genetics tasks based on their interest on the task surface features. Therefore, it is possible that in our study students used other agendas (e.g., interest on task surface features) than those based on perceived task difficulty when they selected tasks during training (cf., Ariel et al. 2009). When students select tasks based on surface features that are irrelevant for goal attainment, it is likely that the difficulty and support levels of those tasks are not adapted to their individual learning needs. This may also be the reason why the number of tasks

selected during training did not predict the post-test performance (Hypothesis 3c). Solving a higher number of tasks is not sufficient to improve learning performance. A higher number of tasks would increase performance only when task difficulty and support levels are aligned with students' individual needs in different moments of learning (cf., Mihalca et al. 2011).

In addition to the lack of predictive value of the number of training tasks for performance, the total time spent on training did not predict performance (Hypothesis 3c). This suggests that spending additional time in training is not sufficient to produce learning (e.g., Nelson and Leonesio 1988). According to Karpicke (2009), time itself does not cause learning, but the processes (e.g., control processes) that take place during this time. As we were not able to detach support levels of the tasks from their difficulty levels, it was not possible to get deeper insights into the control processes during training. Future research should examine whether the predictors of task selections are represented mainly by students' metacognitive judgments (i.e., monitoring-guided control) or also by the objective and/or perceived task difficulty and the type of support level (cf., control-guided monitoring, see Koriat et al. 2006).

Conclusion

Despite its limitations, this study contributes to the field in various ways. First, it extends the research on learner control in computer-assisted learning by showing that students who have a low accuracy of their metacognitive judgments benefitted more from a restricted control rather than from a full control over their learning. Accordingly, high level of learner control should only be provided for the students who are sufficiently accurate in their judgments, and this is more likely when students have higher prior knowledge (cf., Clark and Mayer 2003). Another option would be to adapt the degree of learner control dynamically to the accuracy of students' metacognitive judgments (cf., Mihalca et al. 2011).

Second, the current study extends research on monitoring accuracy by providing evidence that metacognitive judgments affect control processes and that absolute accuracy of these judgments predicts performance of self-regulated learning from problem-solving tasks. However, more research is required to investigate in-depth how metacognitive judgments influence self-regulated task selections during problem solving. Future studies need to investigate the underlying mechanisms of task selection using more simplified selections either in terms of task difficulty or in terms of support levels.

Third, the finding that the degree of learner control has different effects on the relationship between monitoring accuracy and performance is very important, mainly because the vast majority of studies did not experimentally manipulate the degree of control over learning. Furthermore, we allowed students to have control over multiple learning features such as which tasks to study, in which order, and for how long (cf., Corbalan et al. 2008). The degree of learner control over multiple features of learning has the advantage of being highly ecologically valid because in real classroom situations students are not constrained to select only specific features (e.g., the amount of restudy time). As this is the first study in the literature that, to the best of our knowledge, reports effects of manipulating the degree of learner control on the relationship between monitoring accuracy and performance in problem solving, it will be important to replicate these findings in future studies.

Educationally, our findings highlight the importance of monitoring accuracy when learning from problem-solving tasks, and the use of different types of judgments (i.e., EOLs and RCJs) in addition to prior knowledge as a basis for adapting the degree of learner control to the students' individual learning needs.

Appendix

Examples of Genetics Problems from Prior Knowledge Test and Post-test

1. The diastema (the space between the upper incisors) is a dominant trait (D), while the lack of it is a recessive trait (d). What phenotype (and what percentage) will the offspring of a heterozygous couple for the distema trait have?

   - 100% with diastema;
   - 50% with diastema and 50% without diastema;
   - 100% without diastema;
   - 25% without distema and 75% with diastema.

2. In humans, brown eyes (C) are dominant over blue eyes. What is the genotype of the father (who has brown eyes) of a blue-eyed child, if the child's mother has blue eyes?

   - CC;
   - cc;
   - Cc;
   - cC.

References

Alexander, P. A. (2013). Calibration: what is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction, 24*, 1–3.

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*(3), 432–447.

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction, 16*(5), 389–400.

Baars, M., Vink, S., van Gog, T., de Bruin, A., & Paas, F. (2014a). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. *Learning and Instruction, 33*, 92–107.

Baars, M., van Gog, T., de Bruin, A., & Paas, F. (2014b). Effects of problem solving after worked example study on primary school chidren's monitoring accuracy. *Applied Cognitive Psychology, 28*(3), 382–391.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28*(5), 610–632.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417–444.

de Bruin, A. B. H., & van Gog, T. (2012). Improving self-monitoring and self-regulation of learning: from cognitive psychology to the classroom. *Learning and Instruction, 22*, 245–252.

Clark, R. C., & Mayer, R. E. (2003). *E-learning and the science of instruction*. San Francisco: Jossey-Bass/Pfeiffer.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/ correlation analysis for the behavioral sciences* (3rd ed.). Mahwah: Lawrence Erlbaum.

Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2006). Towards a personalized task selection model with shared instructional control. *Instructional Science, 34*(5), 399–422.

Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2008). Selecting learning tasks: effects of adaptation and shared control on learning efficiency and task involvement. *Contemporary Educational Psychology, 33*(4), 733–756.

Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24*, 4–14.

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249–275). Mahwah: Erlbaum.

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*(4), 228–232.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks: Sage.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280.

Dunlosky, J., Hertzog, C., Kennedy, M. R. T., & Thiede, K. W. (2005). The self-monitoring approach for effective learning. *Cognitive Technology, 10*, 4–11.

Gall, J. E., & Hannafin, M. J. (1994). A framework for the study of hypertext. *Instructional Science, 22*(3), 207–232.

Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., & Eysink, T. H. S. (2009). Learning with hypermedia: the influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior, 25*(2), 360–370.

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15*(1), 84–93.

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*(4), 584–587.

Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Memory & Cognition, 37*(7), 1001–1013.

Hadwin, A. F., & Webster, E. A. (2013). Calibration in goal setting: Examining the nature of judgments of confidence. *Learning and Instruction, 24*, 37–47.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: The Guilford Press.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31.

Karich, A. C., Burns, M. K., & Maki, K. E. (2014). Updated meta-analysis of learner control within educational technology. *Review of Educational Research, 84*(3), 392–410.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*(4), 469–486.

Kicken, W., Brand-Gruwel, S., & van Merriënboer, J. J. (2008). Scaffolding advice on task selection: a safe path toward self-directed learning in on-demand education. *Journal of Vocational Education and Training, 60*(3), 223–239.

Kimball, D. R., Smith, T. A., & Muntean, W. J. (2012). Does delaying judgments of learning really improve the efficacy of study decisions? Not so much. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(4), 923–954.

Kopcha, T. J., & Sullivan, H. (2007). Learner preferences and prior knowledge in learner-controlled computer-based instruction. *Educational Technology, Research and Development, 56*(3), 265–286.

Koriat, A. (1997). Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370.

Koriat, A. (2012). The relationships between monitoring, regulation and performance. *Learning and Instruction, 22*(4), 296–298.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*(1), 36–69.

Koriat, A., Nussinson, R., & Ackerman, R. (2014). Judgments of learning depend on how learners interpret study effort. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(6), 1624–1637.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 609–622.

Kostons, D., van Gog, T., & Paas, F. (2010). Self-assessment and task selection in learner-controlled instruction: differences between effective and ineffective learners. *Computers & Education, 54*(4), 932–940.

Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction, 22*(2), 121–132.

Kraiger, K., & Jerden, E. (2007). A meta-analytic investigation of learner control: old findings and new directions. In S. M. Fiore & E. Salas (Eds.), *Toward a science of distributed learning* (pp. 65–90). Washington, DC: American Psychological Association.

Lawless, K. A., & Brown, S. W. (1997). Multimedia learning environments: Issues of learner control and navigation. *Instructional Science, 25*(2), 117–131.

Lee, S. S., & Lee, Y. H. (1991). Effects of learner-control versus program-control strategies on computer-aided learning of chemistry problems: for acquisition or review? *Journal of Educational Psychology, 83*(4), 491–498.

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 464–470.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know?. *Organizational Behavior and Human Performance, 20*(2), 159–183.

Loyens, S. M., Magda, J., & Rikers, R. M. (2008). Self-directed learning in problem-based learning and its relationships with self-regulated learning. *Educational Psychology Review, 20*(4), 411–427.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology, Research and. Development, 50*(3), 43–59.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15*(1), 174–179.

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52*(4), 463–477.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(4), 851–861.

Mihalca, L., Salden, R., Corbalan, G., Paas, F., & Miclea, M. (2011). Effectiveness of cognitive-load based adaptive instruction in genetics education. *Computers in Human Behavior, 27*(1), 82–88.

Mihalca, L., Mengelkamp, C., Schnotz, W., & Paas, F. (2015). Completion problems can reduce the illusions of understanding in a computer-based learning environment on genetics. *Contemporary Educational Psychology, 41*, 157–171.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51*(2), 102–116.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 676–686.

Nelson, T. O., & Narens, L. (1990). Metamemory: a theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–141). New York: Academic Press.

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*(4), 207–213.

Niemiec, R. P., Sikorski, C., & Walberg, H. J. (1996). Learner-control effects: a review of reviews and a meta-analysis. *Journal of Educational Computing Research, 15*(2), 157–174.

Nietfeld, J. L., & Schraw, G. (2002). The effect of knowledge and strategy training on monitoring accuracy. *Journal of Educational Research, 95*(3), 131–142.

Nugteren, M. L., Jarodzka, H., Kester, L., & van Merriënboer, J. J. G. (2015, August). How can tutoring support task selection? In S. Moser (chair), *Tutoring and E-Tutoring in Educational Settings.* Symposium conducted

at the 16th biennial conference of the European Association for Research on learning and instruction (EARLI), Limassol, Cyprus.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429–434.

Paas, F., van Merriënboer, J. J. G., & van Gog, T. (2011). Designing instruction for the contemporary learning landscape. In K. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: vol. 3. Application to learning and teaching*. Washington: American Psychological Association.

Pieger, E., Mengelkamp, C., & Bannert, M. (2016). Metacognitive judgments and disfluency- does disfluency lead to more accurate judgments, better control, and better performance? *Learning and Instruction, 44*, 31–40.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88*(1), 144–161.

Rey, G. D., & Buchwald, F. (2011). The expertise reversal effect: Cognitive load and motivational explanations. *Journal of Experimental Psychology: Applied, 17*(1), 33–48.

Schnotz, W. (2010). Reanalyzing the expertise reversal effect. *Instructional Science, 38*(3), 315–323.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4*(1), 33–45.

Son, L. K., & Kornell, N. (2008). Research on the allocation of study time: key studies from 1890 to the present (and beyond). In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 333–351). Hillsdale: Psychology Press.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 204–221.

Son, L. K., & Sethi, R. (2006). Metacognitive control and optimal learning. *Cognitive Science, 30*(4), 759–774.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: an analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 1024–1037.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73.

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 (Eds.), *Handbook of metacognition in education* (pp. 85–106). New York: Routledge.

Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language, 64*(2), 109–118.

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*(4), 584–587.

van Loon, M. H., de Bruin, A. B., van Gog, T., van Merriënboer, J. J., & Dunlosky, J. (2014). Can students evaluate their understanding of cause-and-effect relations? The effects of diagram completion on monitoring accuracy. *Acta Psychologica, 151*, 143–154.

van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: a four-component instructional design model*. Englewood Cliffs: Educational Technology Publications.

van Merriënboer, J. J. G., & Sluijsmans, D. M. A. (2009). Toward a synthesis of cognitive load theory, four-component instructional design, and self-directed learning. *Educational Psychology Review, 21*(1), 55–66.

Vygotsky, L. (1978). *Mind in society: the development of higher mental process*. Cambridge: Harvard University Press.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale: Erlbaum.