The benefit of reflection prompts encouraging learning with hints in programming education

Anonymous[a,*], Anonymous[a], Anonymous[a], Anonymous[a], Anonymous[a] and Anonymous[a]

[a]*University*

ARTICLE INFO

Keywords:
reflection prompt
hint
long-term performance

ABSTRACT

Giving learners hints is an often used scaffolding practice, but its impact may be limited if learners do not appropriately reflect. In this study, we examined if prompting learners to reflect along with receiving hints could improve learning outcomes, including short-term and long-term performance, perceived learning, and satisfaction. The study was conducted as a randomized field experiment in the context of learning programming in a four-week online course. We compared three conditions: prompting learners to reflect, giving them hints, and providing both reflection prompts and hints. The results showed no significant difference across conditions in terms of short-term performance and satisfaction. However, we did see differences in long-term performances of knowledge transfer and perceived learning. This work contributes to an evidence-based understanding of the importance of the reflection process in learning from hints and its benefit on long-term learning outcomes.

1. Introduction

1.1. Reflection prompts and hints

Providing reflection prompts to learners has been a widely adopted instructional practice aimed at improving learners' academic performance. Reflection is generally defined as a process of expanding and deepening one's understanding by critically analyzing what has been learned and how (Dewey, 1997; Moon, 2013; Boud et al., 1985; Lew and Schmidt, 2011). It scaffolds knowledge acquisition by facilitating the development of cognitive structures and making them available for problem-solving (Boekaerts, 1999; van den Boom et al., 2004). Reflection also leads learners to deliberately review their learning experiences and learn from them. Hence it tends to increase academic performance (Lew and Schmidt, 2011; O'Rourke, 1998; Dewey, 1997; Moon, 2013; Boud et al., 1985). Due to the clear benefits of reflection, researchers have been designing various kinds of interventions aimed at encouraging learners to engage in reflection during different phases of the learning process. An often-used intervention is metacognitive prompts. Over more than two decades, researchers have confirmed the effect of prompts on stimulating various metacognitive processes including reflection (Lin and Lehman, 1999; Chi et al., 1989; King, 1992; Lin, 2001).

Given the role of reflection in knowledge acquisition and academic performance, reflection prompts might be a useful intervention encouraging meaningful learning from hints and improving learning gain. Hints are commonly used to provide cognitive support explaining how to solve a given problem (Roll et al., 2011) or triggering a particular cognitive process (Nokes et al., 2011). Hints are often shown to be ineffective in increasing learning gain despite their design purpose to support learning. One reason for this is a lack of learners' mindful interactions with hints (Aleven et al., 2006; Roll et al., 2011). In general, learners do not deliberately activate metacognitive processes such as reflection when they learn (Lin and Lehman, 1999; Berardi-Coletta et al., 1995; Chi et al., 1989; Bransford et al., 1999). In addition, Aleven et al. (2006) showed that nearly half of undesirable interactions with hints are caused by learners aiming to find answers from hints without trying to thoroughly understand them (e.g., mindlessly clicking through hints). Follow-up work by Roll et al. (2011) found that guiding learners away from these behaviors did not make a significant difference in learning gains, suggesting that a solution might be instead to encourage mindful interactions with hints. Providing reflection prompts along with hints might be the missing piece.

The combined effect of hints and reflection prompts on task performance is still underexplored. Berthold et al. (2007) investigated a combination of cognitive and metacognitive prompts as means of promoting reflection and monitoring. Yet, their cognitive prompts were different from hints in that they were designed to activate organization

*Corresponding author
ORCID(s): 0000-0001-0000-0000 (Anonymous); 0000-0001-0000-0000 (Anonymous); 0000-0001-0000-0000 (Anonymous); 0000-0001-0000-0000 (Anonymous); 0000-0001-0000-0000 (Anonymous); 0000-0001-0000-0000 (Anonymous)

and elaboration strategies instead of directly explaining domain knowledge that learners might lack. Authors (date) showed that hints significantly increased immediate programming performance, but only when they were accompanied by self-explanation prompts. However, the self-explanation prompts did not seem to elicit reflection processes and instead were (1) designed primarily for general critical thinking rather than reflection and (2) given before letting learners apply hints to solve a problem. In fact, Authors (date) reported that during the post-study interview, one of the study participants suggested presenting self-explanation prompts after the task so that they could reflect upon the whole process including how they understood and used hints. It is clear that more investigations are necessary to comprehend the effect of reflection prompts on learners' interaction with hints and to examine the effect of such interactions on task performance.

This work addresses such a need and examines how hints and reflection prompts support programming learning, taking into consideration the duration of task complexity and prompt effects. To understand how providing learners with hints and reflection prompts relates to performance, we conducted an experiment combining three scaffolds: (1) hints alone, (2) reflection prompts alone, and (3) both hints and reflection prompts. We measured both short- and long-term effects on transfer task performance, perceived learning, and course satisfaction. Our findings show that combining reflection prompts and hints benefits long-term transfer task performance, but does not so for short-term performance. Our findings show that (1) reflection prompts alongside hints can support mindful interaction with hints, and (2) such mindful interaction with hints impacts long-term performance. These findings have design implications supporting the need to include reflection prompts with hints in the learning environments.

1.2. Previous studies on design factors

Previous studies suggest that it is important to consider factors that might affect the effect of reflection prompts on learning outcomes (Bransford et al., 1999; Barnett and Ceci, 2002; Day and Goldstone, 2012; Kim and Kendeou, 2021; Lin and Lehman, 1999; Lew and Schmidt, 2011; Schworm and Renkl, 2002; Bannert, 2006; Bannert and Reimann, 2012; Bannert et al., 2015; Krause and Stark, 2010). Previous research has shown that reflection prompts could increase performance most effectively when applied to knowledge transfer tasks, that is, tasks requiring learners to apply adopted knowledge to new contexts (Bransford et al., 1999; Barnett and Ceci, 2002; Day and Goldstone, 2012; Kim and Kendeou, 2021). Learners within the same intervention did not show significantly better performance on simpler tasks such as recall or knowledge comprehension tasks (Lin and Lehman, 1999; Lew and Schmidt, 2011; Schworm and Renkl, 2002; Bannert, 2006; Bannert and Reimann, 2012; Bannert et al., 2015). For instance, three studies by Bannert and colleagues presented significant increase in transfer task scores with moderate effect sizes (d=0.55, d=0.58, d=0.44) (Bannert, 2006; Bannert and Reimann, 2012; Bannert et al., 2015). Schworm and Renkl (2002) also used transfer tasks and found an increase in learners' task performance after providing self-explanation prompts which evoked reflections along with other metacognitive processes. Lew and Schmidt (2011) suggested that addressing reflection prompts assisted with the synthesis of new and prior knowledge, while reflection prompts alone did not significantly affect learners' performance on the knowledge acquisition tasks. While Krause and Stark (2010) did not find any impact of reflection prompts on task performance, they found that learners made significantly more progress on their tasks, with a large effect size (d=0.8), only when working on relatively complex problems. There was no significant difference in progress on simpler problems. Lin and Lehman (1999) observed meaningful increases in score only on more complex problems, which led learners to far transfer their conceptual understanding to dissimilar contexts. These findings clearly present a relationship between the complexity of tasks and the benefits of reflection prompts. Reflection prompts are highly likely to improve performance on transfer tasks or other tasks with higher complexity, whereas they do not affect performance on less complex tasks, such as recall or knowledge comprehension tasks.

Some of the previous work has also tried to reveal a duration of the effect of reflection prompts, with Bannert et al. (2015) being one of few studies that examined their long-term effect. They found that learners who received reflective prompts performed significantly better on their delayed transfer tasks compared to the control group. In contrast, Jeong et al. (2008), who studied reflection prompts in the context of concept map drawing tasks, found no effect of metacognitive prompts on their delayed task performance. Studies that reported short-term effects measured by immediate post-tests after learning sessions also showed conflicting results (Bannert, 2006; Bannert and Reimann, 2012; Krause and Stark, 2010; Lin and Lehman, 1999; van den Boom et al., 2004). For example, Krause and Stark (2010) presented learners with reflection prompts asking them to justify the choice of learning strategies, and found no significant effect on learners' performance. Similarly, van den Boom et al. (2004) showed no significant effect of reflection prompts on the immediate post-test score. Yet, Bannert (2006); Bannert and Reimann (2012) found significant positive effects of reflection prompts on immediate post-test scores. These mixed results show that the duration

of the effect might be an important factor to consider in reflection prompt study but it requires more investigation. The effect of reflection prompts might appear either immediately or after a certain time. To account for the potential effect of reflection prompts as well as to understand the duration of the impact from the prompts, it is necessary to measure both short-term and long-term effects while controlling for the task complexity.

1.3. Research questions

In this study, we aimed to understand the short-term and long-term effects of different combinations of hints and reflection prompts on task performance. The task complexity was controlled and thus only the transfer task was used. Furthermore, we designed a field study instead of a lab study, to collect the data in an authentic learning environment. We asked the following research questions:

RQ1 What is the short-term effect of the combined hint and reflective prompt intervention on transfer task performance?
RQ2 What is the long-term effect of the combined hint and reflective prompt intervention on transfer task performance?
RQ3 What is the short-term effect of the combined hint and reflective prompt intervention on meaningful interaction with the task?
RQ4 What is the long-term effect of the combined hint and reflective prompt intervention on meaningful interaction with the task?
RQ5 What is the effect of the combined hint and reflective prompt intervention on perceived learning?
RQ6 What is the effect of the combined hint and reflective prompt intervention on satisfaction on learning?

We posed RQ1 and RQ2 to investigate if there is any difference between the short-term (RQ1) and long-term effects (RQ2) of the interventions on achievement. RQ3 and RQ4 were added to understand learners' short-term and long-term behavioral changes in solving programming tasks measured through the number of task submissions. In this particular learning environment, learners were allowed to submit their tasks as many times as they wanted and could check their scores for each submission. If learners in a certain condition submitted tasks significantly more only to achieve the same or lower scores than other conditions, it would be the evidence of learners' less meaningful engagement with interventions and tasks and even the evidence of gaming the system by carelessly guessing answers and checking if their guesses were correct. RQ5 and RQ6 were posed to answer the effect of the interventions on perceived learning and satisfaction over short-term tasks.

This study addressed these questions through a three-condition experiment: (1) the *hint intervention* was provided while a learner was working on a task i.e., the hint condition, (2) a *reflection prompt intervention* would appear when a learner finished tasks i.e., the reflection condition, and (3) both *the hint and a reflection prompt intervention* present the programming task i.e., the hint-reflection condition. The hint condition was a control condition, while the hint-reflection condition was the main treatment condition. The reflection condition was additionally included to see if reflection prompts alone could impact the listed learning outcomes, when compared to the hint condition.

2. Pilot study

2.1. Study context

Prior to the main experiment, a pilot study was conducted to ensure the appropriate design of hints and reflection prompts. The primary design decisions to be made were (1) when to present reflection prompts - every time before or after the hint was presented or only after a learner completed the task, (2) which concepts or code elements to explain through hints, (3) which type of reflection prompts was effective in eliciting reflection, and (4) if learners could understand task questions, reflection prompt questions, and hints. For the pilot study, nineteen learners who had already taken this course were recruited.

2.2. Study procedure

A set of reflection prompts were designed based on previous literature (Ifenthaler, 2012a; Devolder et al., 2012; Davis, 2003; Zepeda et al., 2015). The authors and the course instructor reviewed and discussed the prompt designs to ensure that the prompts can activate reflection processes. During the pilot study, the first author explained the benefits of reflection prompts to the participants, to prevent potential annoyance and unwillingness from interacting with the prompts (Bannert and Reimann, 2012). Learners were then asked to think aloud while working on tasks with both

hint and reflection prompt interventions. Learners were provided with a reflection prompt before and after every hint request. Task work was followed by an interview with each learner.

2.3. Instruments

Follow-up interviews. The first author asked common questions and follow-up questions to clarify some of the statements made during each participant's think-aloud session. Common questions were as follows:

- Think about how you interacted with the prompts. Did prompts help you check your understanding with hints?
- Was it easy to understand what prompts asked you to think about?
- Think about how you interacted with hints while working on the Jupyter Notebook programming tasks. Did hints address what you wanted to know to solve your problem?
- Was it easy to understand the suggestions that hints gave?

Follow-up questions varied and were determined by some specific statements the participants said while self-reporting (think-aloud) on the task.

2.4. Pilot study results

Overall, pilot study participants did not comply with reflection prompt instructions when they saw prompts either before or after a hint. The reflection prompt given before each hint was designed to encourage them to reflect on what they already knew and what they were not sure of. However, 13 out of 19 participants ignored these reflection prompts and proceeded to see hints. During the follow-up interview, one participant (P2) said that they did not have enough motivation to engage with a reflection prompt when they were frustrated, but were eager to see a hint to resolve the frustration as soon as possible. Another participant (P12) added that using reflection prompts before seeing hints did not add anything to their understanding. One participant (P15) even described the prompt as 'an extra obstacle to the selection of hints.' Similarly, 9 out of 19 participants also considered reflection prompts given after every hint annoying and did not comply with prompts asking them to activate reflection.

The participants' annoyance due to the high frequency of prompting along with low compliance with prompts is not at all surprising considering findings from previous studies (Berthold et al., 2007; Bannert and Reimann, 2012). It has been frequently shown that learners often prefer to invest less effort in learning and wrongly believe that they choose an efficient way to increase learning gain (Bjork et al., 2013; Clark, 1982). Yet, it was concerning that most participants did not engage with reflection prompts even though the benefits of engaging with reflection prompts were explained to them, to motivate them for such engagements (Bannert and Reimann, 2012). To prevent learners' annoyance and to draw their attention to reflection prompts, for the main study we decreased the frequency of prompting to only once when learners completed their tasks. In the modified design, reflective prompts aimed to activate learners' reflection process over the entire tasks, including but not limited to their use of hints. Authors (date) suggested such a design to encourage reflections instead of eliciting them after each hint.

Based on learner feedback, we also included an additional hint to address a problematic area in the assignment. Otherwise, learners agreed that the assignment questions and reflection prompts were clear and understandable.

3. Method

3.1. Study context

A field study was conducted where learners received different combinations of hints and reflection prompts interventions while working on a programming transfer task. There were two tasks: (1) the immediate task measured the short-term effect of the interventions and (2) the delayed task measured the long-term effect of the interventions. Interventions (hints and reflection prompts) accompanied immediate tasks, but not the delayed tasks.

We used G*Power (Faul et al., 2009) to decide the estimated sample size, which showed that there should be more than 159 participants total i.e., at least 53 per condition for the medium effect size (Cohen's $f = 0.25$) with an alpha of .05 and power of .80 for one-way ANOVA tests. Since most previous studies reported their significant results with medium to large effect sizes, we determined to use the medium effect size based on the Cohen (2013).

Data were collected during two iterations of the same introductory data science course in the fully online Master's degree program of OMITTED FOR PEER REVIEW. The course materials included lecture videos, readings, and

programming assignments. Iteration 1 took place first which was followed by iteration 2 in the next semester. Each iteration was four weeks long and had a different number of enrolled learners. The instructor team was ethically obliged to provide a similar learning experience to all learners who enrolled in the course for credits in the same term. Hence, all learners enrolled in the first iteration were assigned in the reflection condition. All learners who enrolled in the second iteration of the course were split between the hint-reflection condition and the hint condition. This was based on discussions with the instructor who expected that giving only a portion of learners hints would discourage learners who did not get hints and even feel they were not fairly treated compared to other students in the same class.

The total number of learners who enrolled in the two iterations of the course was 432 before removing learners who did not complete even a single task. While each of the course materials was recommended in certain weeks, some learners worked ahead and finished materials earlier. That is, while they followed all the course materials in order as planned, they might have finished courses earlier than other students. IRB oversight was obtained through OMITTED FOR PEER REVIEW study ID OMITTED FOR PEER REVIEW.

3.2. Instruments

Questionnaire on metacognitive skills. To confirm that learners in different conditions did not significantly differ in metacognitive skills including reflection before engaging with the interventions, a questionnaire on metacognitive skills was conducted at the beginning of the course. Six question items were adopted from the Metacognitive Self-regulation section of the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich, 1991) and slightly reworded to follow the context of the course.

Tasks and assignments. To measure learners' performance on immediate and delayed knowledge transfer tasks, four weekly assignments were prepared in Jupyter Notebook, a web-based programming environment (Kluyver et al., 2016) used in programming courses. All assignments consisted of Python programming tasks that were automatically graded upon submission. All tasks were designed and reviewed by the course staff and the first author to confirm that they measured what learners were supposed to learn each week. Conceptually, each assignment was built upon the previous assignments. Therefore, it was expected that learners who did not perform well on earlier assignments would perform poorly on the subsequent assignments. Learners had to reach 80% of the full credit per assignment to pass the course; they were allowed to submit assignments as many times as they wanted prior to a weekly deadline.

The first two assignments, assignments 1 and 2, consisted of three separate tasks each and included an intervention. These tasks were used to measure the short-term effect of the intervention on learners' academic performance. Since each of these tasks was provided separately, learners who were provided with reflection prompts could activate their reflection per task. That is, these learners saw reflection prompts at least three times per assignment since they had to submit three different tasks for each of assignments 1 and 2. The last two assignments, assignments 3 and 4, included more than one task per assignment and did not offer any interventions to learners in any of the conditions. These assignments were used to measure the long-term effect of interventions received for assignments 1 and 2 on learners' academic performance.

Hints. Learners who were offered hints (i.e., the hint-reflection condition and the hint condition) could engage with the hint intervention by clicking the 'Show Hint' button while working on a task. When learners clicked the button, a pop-up with a list of summaries of available hints was displayed (Figure 1 (a), (b)).

When learners chose a hint and clicked the 'Next' button, they could see the full text of the chosen hint on the next pop-up. This full hint was inscribed below the associated task cell (Figure 1 (c)), so that learners could easily look up hints while working on a task even after closing the hint pop-up.

Reflection prompts. Learners assigned to the conditions with a reflection prompt intervention (i.e., the reflection condition and the hint-reflection condition) were given a reflection prompt when they clicked the submit button to submit their task and receive auto-graded credit (Figure 1 (d)).

All reflection prompts were designed as directed reflection prompts, i.e. prompts that offer specific instructions such as 'stop and think about what you misunderstood before seeing hints.' We opted for the directed prompts based on the pilot study participants' preference for them over the generic prompts that encourage learners' reflection by simply asking 'stop and think.' The distinction was proposed by Davis (2003) and no agreement currently exists on which of the two are more beneficial for the effective reflection process (Kramarski and Kohen, 2017; Ifenthaler, 2012b; Davis, 2003; Lee and Chen, 2009). Hence, upon task submissions, a learner would be provided with a reflection prompt statement randomly chosen from the following list:

- What steps did you take when solving the problem? Why? Provide a short justification for each step.

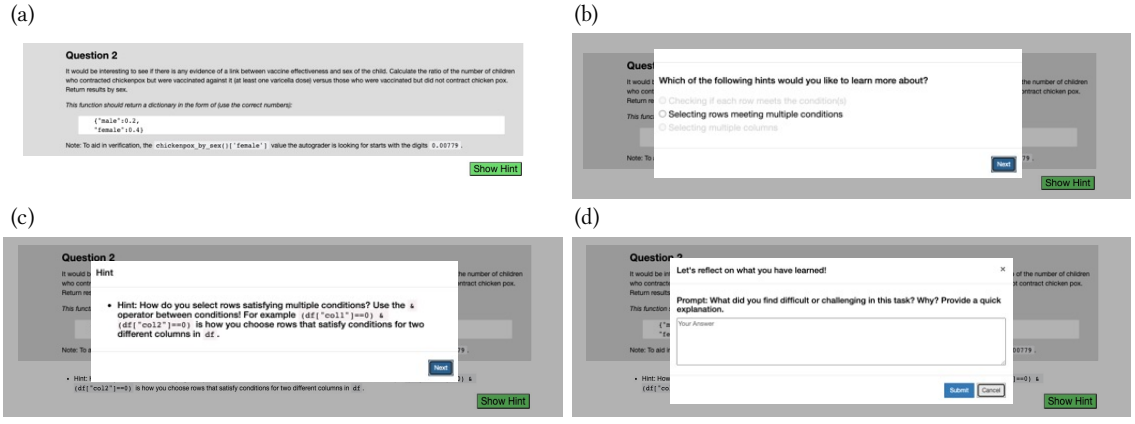

Figure 1: (a) Each task has a 'Show Hint' button below a question. (b) When a learner clicks the 'Show Hint' button, a pop-up appears and shows a list of available hints. Some hints are grayed out here since they have already been chosen by the learner in previous interactions. (c) The full text of the chosen hint is shown on a pop-up and also inscribed below the associated question so that a learner can see hints even after closing the pop-up. (d) When a learner clicks the submit button, a reflection pop-up appears.

Table 1
Question items for measuring perceived learning and satisfaction

Construct measured	Questionnaire items
Perceived learning	• The weekly assignments helped me learn more about the topic (e.g., regular expression) • I learned new things from the weekly assignments. • The weekly assignments helped me remember the things I learned. • The weekly assignments helped me apply the things I learned to other problems.
Satisfaction	• I enjoyed the weekly assignments. • I had fun working on weekly assignments. • Working on the weekly assignments was pleasant.

- What did you find difficult or challenging in this task? Why? Provide a quick explanation.
- What was the main thing you learned by completing this task?

Questionnaire on perceived learning and satisfaction. A questionnaire was also distributed to learners to measure perceived learning and satisfaction over the learning experience with the given intervention. The questionnaire had four 7-point Likert scale questions for perceived learning and three 7-point Likert scale questions for satisfaction, where both of the question sets asked how much a learner agreed with each statement. Statements were adopted from Barzilai and Blau (2014) and revised for this study as presented in Table1.

3.3. Study procedure

At the beginning of the first week, learners were asked to take the questionnaire on metacognitive skills. Learners assigned to the hint condition and the hint-reflection condition could use the hints while working on tasks included in assignments 1 and 2. Learners in the reflection condition or the hint-reflection condition were also asked to report their reflection process as a written response, upon submitting their tasks for assignments 1 and 2. Having completed week 1 and week 2 materials and prior to embarking on week 3, learners were required to take a questionnaire on perceived learning and satisfaction. The course was self-paced with fixed dates of assignment deadlines.

3.4. Data analyses

For data cleaning purposes, we removed incomplete submissions (e.g., questionnaires which were only partially filled out) while including learner data as long as they made at least one complete submission of a task or an assignment. For example, there were learners who received the grace period for a portion of assignments due to their personal circumstances and their tasks were manually graded with different criteria. In this case, we still retained their other submissions and questionnaire response data instead of dropping the entire submission data of the learner.

To address RQ1 and RQ2 related to short-term and long-term effects of interventions on the transfer tasks, we conducted a mixed-effect model analysis in R using the lme4 package (Bates et al., 2014). A mixed-effect model was adopted to account for scores of multiple assignment submissions per learner. Data of submitted assignment scores were split into two sets and respectively used to analyze the short-term effect (data of week 1 and week 2 assignments) and a long-term effect (data of week 3 and week 4 assignments). That is, there were two mixed effect models: one for the immediate task and the other for the delayed task. The experimental condition and assignment label for the week that the assignment was due (i.e., week1, week2, week3, and week4) were used as the fixed effects. For example, in the mixed-effect model for the immediate task, there were two assignment values, i.e. week1 or week2, included as a covariate. Learner IDs were used as random effects in the model. Regarding RQ3 and RQ4, a one-way Analysis Of Variance (ANOVA) was conducted over the number of submissions for measuring the short-term effect on performance (RQ3) and for measuring the long-term effect on performance (RQ4). For RQ5 and RQ6, we ran a one-way ANOVA over the questionnaire responses on perceived learning (RQ5) and satisfaction (RQ6).

The normality of the residuals of the assignment score data was visually inspected and confirmed by a quantile-quantile (QQ) plot. Considering that the submission count data were highly right-skewed, we applied reciprocal transformations to the immediate task submission count and logarithmic transformation to the delayed task submission count data. The data transformation reduced the skewness values respectively to 0.46 and -0.39 which are in the generally accepted range of normal distribution. Even though the perceived learning data and satisfaction data were also non-normally distributed and all showed high skewness values, statistical tests such as ANOVA and regression analysis are still statistically valid options for the non-normally distributed Likert-scale data (Norman, 2010; Harpe, 2015).

4. Results

4.1. Data overview

The total number of learners in the dataset used for analysis was 354 with the following breakdown across the conditions: 70 participants in the hint condition, 222 participants in the reflection condition, and 62 participants in the hint-reflection condition. Overall, the number of participants meets the target sampling size.

The overall mean of the submission score of immediate tasks for measuring the short-term effect on performance and delayed task for measuring the long-term effect on performance were 53.26 out of 100 ($SD = 49.89$) and 44.02 out of 100 ($SD = 34.71$). Means and standard deviations of the submission scores per condition are given in Table 2. It is important to note that each of assignments 1 and 2 was split into each task which makes them into six tasks (1-1, 1-2, 1-3, 2-1, 2-2, and 2-3) and therefore task score was either 0 or 100, which explains large standard deviations. On the other hand, assignments 3 and 4 were composed of multiple tasks. Average submission counts of immediate tasks and delayed tasks were respectively 2.08 ($SD = 2.30$) and 16.79 ($SD = 13.95$). Means and standard deviations of the number of submissions per condition are presented in Table 3. Learners' average satisfaction score was reported as 17.88 out of 21 ($SD = 3.26$) and the average score of perceived learning was 25.01 out of 28 ($SD = 2.87$). Means and standard deviations for the scores per condition are shown in Table 4.

An ANOVA was run on the questionnaire on metacognitive skills and it confirmed that the learners' prior metacognitive skills were not significantly different across conditions.

4.2. Study results

RQ1 and RQ2 asked whether hint and reflective prompt interventions had the short-term effect (RQ1) and the long-term effect (RQ2) on the performance of a transfer task. These results analyzed through mixed-effect modeling are reported in line with the guidelines suggested by Brown (2021).

Analysis to address RQ1 revealed that there was no significant difference across the conditions in the performance of the immediate transfer tasks that measured the short-term effect of the interventions (i.e., assignments 1 and 2) ($\chi^2 = 4.98$, $p = 0.08$). Yet, data analysis addressing RQ2 showed that there was a significant difference across the conditions in the long-term effect on the performance of the delayed transfer tasks (i.e., week 3 and week 4) ($\chi^2 = 7.84$,

$p = 0.01$). Task score of the hint-reflection group was on average an estimated 6.26 points higher than the hint group ($\hat{\beta} = 6.26$, $SE = 2.25$, $t = 2.77$). Tukey's post-hoc test confirmed the difference with a small effect size ($\hat{\beta} = -6.26$, $p = 0.01$, Cohen's $f = 0.19$). Furthermore, the task score of the reflection group was on average an estimated 2.34 points higher than the hint group ($\hat{\beta} = 2.34$, $SE = 1.75$, $t = 1.33$). Yet, the post-hoc test showed no significant difference between the reflection condition and the hint condition ($\hat{\beta} = -2.34$, $p = 0.37$). No significant difference was observed between reflection condition and the hint-reflection condition ($\hat{\beta} = -3.92$, $p = 0.08$).

Table 2
Means and standard deviations of assignment scores per condition

Assignment/task labels	Hint	Reflection	Hint-reflection
Task 1-1	67.48 (46.99)	73.83 (44.01)	74.02 (44.03)
Task 1-2	66.46 (47.36)	71.36 (45.26)	82.08 (38.54)
Task 1-3	34.60 (47.66)	36.82 (48.26)	40.38 (49.18)
Task 2-1	63.23 (48.38)	60.34 (48.97)	69.77 (46.11)
Task 2-2	53.30 (50.03)	58.85 (49.26)	68.91 (46.48)
Task 2-3	40.09 (49.11)	36.58 (48.20)	44.86 (49.87)
Assignment 3	40.32 (35.61)	47.43 (33.42)	48.18 (34.87)
Assignment 4	41.24 (35.38)	37.74 (35.34)	48.07 (34.43)

Note. Mean (standard deviation)

Table 3
Means and standard deviations of assignment submission counts per condition

Assignment/task labels	Hint	Reflection	Hint-reflection
Task 1-1	1.85 (1.30)	1.57 (1.01)	1.58 (1.22)
Task 1-2	1.85 (1.32)	1.54 (1.27)	1.32 (0.63)
Task 1-3	2.98 (2.68)	2.89 (3.00)	2.66 (2.17)
Task 2-1	1.80 (1.60)	1.85 (1.56)	1.61 (1.43)
Task 2-2	2.11 (1.65)	1.84 (1.99)	1.48 (0.99)
Task 2-3	2.72 (3.05)	3.00 (4.36)	2.31 (1.66)
Assignment 3	21.56 (17.38)	24.18 (16.44)	17.63 (12.20)
Assignment 4	8.69 (5.44)	12.22 (8.04)	7.00 (4.98)

Note. Mean (standard deviation)

RQ3 and RQ4 asked how each of the interventions affected the behavioral pattern of interacting with tasks measured by the number of immediate and delayed task submissions. An ANOVA showed that there was a significant difference across conditions in the number of immediate task submissions ($F(2, 2528) = 7.42$, $p = 0.0006$, Cohen's $f = 0.07$). While no significant difference was found in the follow-up Tukey post-hoc test, there was suggestive evidence showing that the hint-reflection group made fewer submissions than the hint group ($p = 0.06$, Cohen's $f = 0.11$). For effect on the number of delayed task submissions (RQ4), there was a significant difference across conditions in the number of delayed assignment submissions ($F(2, 805) = 14.99$, $p = 0.0000004$, Cohen's $f = 0.20$). The follow-up Tukey post-hoc test revealed that the reflection group submitted significantly more than both the hint-reflection group ($p < 0.0001$, Cohen's $f = 0.18$) and the hint group ($p < 0.006$, Cohen's $f = 0.11$). There was no difference in the submission count between the hint group and the hint-reflection group (Table 3).

RQ5 asked if the interventions affected perceived learning over assignments 1 and 2 by asking how much learners felt they learned. An ANOVA revealed a significant difference across conditions with small effect size ($F(2, 352) = 7.33$, $p = 0.0.0007$, Cohen's $f = 0.20$). Tukey's post-hoc test showed a significant difference between the hint-reflection condition and the hint condition with a small effect size ($p = 0.004$, Cohen's $f = 0.17$). There was also evidence for a significant difference between reflection and hint, ($p = 0.0001$, Cohen's $f = 0.19$). There was no significant difference between the hint-reflection condition and the reflection condition ($p = 0.91$) (Table 4).

Lastly, RQ6 asked how much learners enjoyed assignments 1 and 2. An ANOVA showed that there was no significant difference in satisfaction across conditions ($p = 0.09$) (Table 4).

Table 4
Means and standard deviations of perceived learning and satisfaction per condition

	Hint	Reflection	Hint-reflection
Perceived learning	23.87 (4.06)	25.27 (2.38)	25.43 (2.56)
Satisfaction	17.14 (3.86)	18.00 (3.06)	18.28 (3.13)

Note. Mean (standard deviation)

5. Discussion and future work

This field study deployed in an online programming course did not show substantial differences between hints, reflection, and hints-reflection conditions in terms of (1) short-term effect of interventions on the immediate task performance and (2) satisfaction. However, it provided evidence for the significant long-term effect on delayed transfer tasks performance when hints and reflection prompts are combined. This suggests that hints could have a long-term effect on achievement only when they are followed by reflection prompts.

Combined with the findings of the submission count, the effect of interventions on the long-term task performance becomes clearer. Compared to those in the hint-only condition, learners exposed to both hints and reflection prompts scored significantly higher, whereas maintaining a similar count of assignment submissions. These results can be interpreted in several ways. Considering that each task was conceptually built on the previous tasks, the hint-reflection intervention can be understood as beneficial in maintaining the knowledge obtained from the previous tasks activated through reflection processes. Maintaining knowledge could enable learners in the hint-reflection condition to perform better by making fewer incorrect submissions. The other interpretation, compatible with the former, is that the hint-reflection intervention led learners to interact with learning materials and tasks in a more meaningful manner. They might have built such interaction patterns, instead of undesirable interactions with hints (e.g., making 'wild' guesses) while working on the immediate tasks and might have maintained the patterns. It is also important to acknowledge that the reflection group made significantly more submissions than the hint-reflection group only to achieve a similar score on the long-term task. This might suggest that reflection prompts without hints or other cognitive support could not fill learners' knowledge gaps. That is, the combination of both reflection prompts, as metacognitive support, and hints, as cognitive support, could effectively improve learners' long-term performance.

The results have implications for how to encourage learners to learn from hints. The first implication is the importance of interventions eliciting reflection processes in learning with hints. Previous studies have been consistent in showing that hints are ineffective in increasing learner performance (Aleven et al., 2006; Roll et al., 2011). Their findings showed that more than half of learners adopt a mindless and passive approach to hints without directing themselves to learn from the hints (Aleven et al., 2006). The reflection prompts with hints might effectively address the issue by encouraging learners to learn from hints, by reviewing how they have applied hints while solving the given problems and what they learned from the hints. Furthermore, our results have shown that the perceived learning in both conditions with reflection prompts was significantly higher than in the hint condition, while there was no significant difference in satisfaction across conditions. This suggests that the current design of prompts with hints does not affect learners' annoyance with intervention design.

Another implication for design stemming from our results is that a reflection prompt does not need to follow every hint presented to the learner. Most previous studies, even without hints, designed a system which displayed reflection prompts more than once or even showed reflection prompts after every decision making (Krause and Stark, 2010; Bannert, 2006; Bannert and Reimann, 2012; Bannert et al., 2015; van den Boom et al., 2004). While this might have worked in a lab setting, this study showed (a) that in an authentic online classroom, learners are highly likely not to comply with reflection prompts due to annoyance, if prompts are shown before or after every hint and (b) that presenting a reflection prompt once at the end of each task could significantly increase task performance.

The importance of the reflection process on learning from hints as demonstrated in the current study suggests that future work can further examine the effects of other metacognitive processes on learners' use of hints. Considering the overall importance of metacognitive processes in learning (Lai, 2011; Wolters and Pintrich, 1998; Sternberg, 1998; Tobias and Everson, 2002), the reflection would not be the only metacognitive process encouraging mindful learning. It would be interesting to see if hints affect academic performance when they are accompanied by metacognitive prompts or other interventions that elicit different metacognitive processes. With such findings, more generalizable conclusions

on the importance of metacognitive processes could be drawn for the design of hint interventions.

In this study, we did not analyze the responses of individual learners which could have indicated the quality of reflection processes. For example, Engelmann et al. (2021) did not find a significant effect of self-created metacognitive prompts including prompts asking for reflection on performance, and their analysis showed that how learners utilized prompts changed the impact of the prompts. Accordingly, another natural extension of the current study would be to measure the quality of learners' use of prompts and examining its impact on the size or significance of the effects of reflection prompts provided with hints. Furthermore, while the study findings are aligned with (Bannert et al., 2015) who showed the benefit of reflection prompts on the long-term transfer task, we acknowledge that the current study findings have reduced power. While the study sample size was set for medium or large effects, the significant findings did not reach the threshold of medium effect size according to Cohen (2013).

6. Conclusion

This work investigated the short-term and long-term effects of reflection prompts and hints on the performance in transfer tasks, perceived learning, and satisfaction in the domain of programming education. We have demonstrated that (a) none of the combinations of metacognitive and cognitive support affected short-term task performance, and (b) the combination of reflection prompts and hints significantly affected long-term performance on a transfer task. This study poses critical design implications on how to design a learning environment that can lead to deeper learning from hints and increased task performance.

7. Declarations of interest

None.

References

Aleven, V., Mclaren, B., Roll, I., Koedinger, K., 2006. Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. International Journal of Artificial Intelligence in Education 16, 101–128.

Authors, date.

Bannert, M., 2006. Effects of reflection prompts when learning with hypermedia. Journal of Educational Computing Research 35, 359–375.

Bannert, M., Reimann, P., 2012. Supporting self-regulated hypermedia learning through prompts. Instructional Science 40, 193–211.

Bannert, M., Sonnenberg, C., Mengelkamp, C., Pieger, E., 2015. Short-and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. Comput. Human Behav. 52, 293–306.

Barnett, S.M., Ceci, S.J., 2002. When and where do we apply what we learn?: A taxonomy for far transfer. Psychological bulletin 128, 612.

Barzilai, S., Blau, I., 2014. Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences. Comput. Educ. 70, 65–79.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823 .

Berardi-Coletta, B., Buyer, L.S., Dominowski, R.L., Rellinger, E.R., 1995. Metacognition and problem solving: A process-oriented approach. J. Exp. Psychol. Learn. Mem. Cogn. 21, 205–223.

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

Bjork, R.A., Dunlosky, J., Kornell, N., 2013. Self-regulated learning: beliefs, techniques, and illusions. Annu. Rev. Psychol. 64, 417–444.

Boekaerts, M., 1999. Self-regulated learning: Where we are today. International journal of educational research 31, 445–457.

van den Boom, G., Paas, F., van Merriënboer, J.J.G., van Gog, T., 2004. Reflection prompts and tutor feedback in a web-based learning environment: effects on students' self-regulated learning competence. Comput. Human Behav. 20, 551–567.

Boud, D., Keogh, R., Walker, D., 1985. What is reflection in learning. Reflection: Turning experience into learning , 7–17.

Bransford, J.D., Brown, A.L., Cocking, R.R., 1999. How people learn: Brain, mind, experience, and school. National Academy Press.

Brown, V.A., 2021. An introduction to linear mixed-effects modeling in r. Advances in Methods and Practices in Psychological Science 4, 2515245920960351.

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., Glaser, R., 1989. Self-explanations: How students study and use examples in learning to solve problems. Cogn. Sci. 13, 145–182.

Clark, R.E., 1982. Antagonism between achievement and enjoyment in ATI studies. Educ. Psychol. 17, 92–101.

Cohen, J., 2013. Statistical power analysis for the behavioral sciences. Academic press.

Davis, E.A., 2003. Prompting middle school science students for productive reflection: Generic and directed prompts. Journal of the Learning Sciences 12, 91–142.

Day, S.B., Goldstone, R.L., 2012. The import of knowledge export: Connecting findings and theories of transfer of learning. Educational Psychologist 47, 153–176.

Devolder, A., van Braak, J., Tondeur, J., 2012. Supporting self-regulated learning in computer-based learning environments: systematic review of effects of scaffolding in the domain of science education. Journal of Computer Assisted Learning 28, 557–573.

Dewey, J., 1997. How We Think. Courier Corporation.

Engelmann, K., Bannert, M., Melzner, N., 2021. Do self-created metacognitive prompts promote short- and long-term effects in computer-based learning environments? Research and Practice in Technology Enhanced Learning 16, 1–21.

Faul, F., Erdfelder, E., Buchner, A., Lang, A.G., 2009. Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. Behavior research methods 41, 1149–1160.

Harpe, S.E., 2015. How to analyze likert and other rating scale data. Currents in Pharmacy Teaching and Learning 7, 836–850.

Ifenthaler, D., 2012a. Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. Journal of Educational Technology & Society 15, 38–52.

Ifenthaler, D., 2012b. Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. Journal of Educational Technology & Society 15, 38–52.

Jeong, H., Gupta, A., Roscoe, R., Wagster, J., Biswas, G., Schwartz, D., 2008. Using hidden markov models to characterize student behaviors in Learning-by-Teaching environments, in: Intelligent Tutoring Systems, Springer Berlin Heidelberg. pp. 614–625.

Kim, J., Kendeou, P., 2021. Knowledge transfer in the context of refutation texts. Contemporary Educational Psychology , 102002.

King, A., 1992. Facilitating elaborative learning through guided Student-Generated questioning. Educ. Psychol. 27, 111–126.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B.E., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J.B., Grout, J., Corlay, S., Others, 2016. Jupyter notebooks-a publishing format for reproducible computational workflows, in: ELPUB, books.google.com. pp. 87–90.

Kramarski, B., Kohen, Z., 2017. Promoting preservice teachers' dual self-regulation roles as learners and as teachers: Effects of generic vs. specific prompts. Metacognition and Learning 12, 157–191.

Krause, U.M., Stark, R., 2010. Reflection in example- and problem-based learning: effects of reflection prompts, feedback and cooperative learning. Educ. Res. Eval. 23, 255–272.

Lai, E.R., 2011. Metacognition: A literature review. Always learning: Pearson research report 24.

Lee, C.Y., Chen, M.P., 2009. A computer game as a context for non-routine mathematical problem solving: The effects of type of question prompt and level of prior knowledge. Comput. Educ. 52, 530–542.

Lew, M.D.N., Schmidt, H.G., 2011. Self-reflection and academic performance: is there a relationship? Adv. Health Sci. Educ. Theory Pract. 16, 529–545.

Lin, X., 2001. Designing metacognitive activities. Educ. Technol. Res. Dev. 49, 23–40.

Lin, X., Lehman, J.D., 1999. Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching 36, 837–858.

Moon, J.A., 2013. A handbook of reflective and experiential learning: Theory and practice. Routledge, London, England.

Nokes, T.J., Hausmann, R.G.M., VanLehn, K., Gershman, S., 2011. Testing the instructional fit hypothesis: the case of self-explanation prompts. Instr. Sci. 39, 645–666.

Norman, G., 2010. Likert scales, levels of measurement and the "laws" of statistics. Adv. Health Sci. Educ. Theory Pract. 15, 625–632.

O'Rourke, R., 1998. The learning journal: from chaos to coherence. Assessment & Evaluation in Higher Education 23, 403–413.

Pintrich, P.R., 1991. A manual for the use of the motivated strategies for learning questionnaire (MSLQ) .

Roll, I., Aleven, V., McLaren, B.M., Koedinger, K.R., 2011. Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. Learning and instruction 21, 267–280.

Schworm, S., Renkl, A., 2002. Learning by solved example problems: Instructional explanations reduce self-explanation activity, in: Proceedings of the Annual Meeting of the Cognitive Science Society, escholarship.org.

Sternberg, R.J., 1998. Metacognition, abilities, and developing expertise: What makes an expert student? 26, 127–140. URL: https://www.jstor.org/stable/23371269.

Tobias, S., Everson, H.T., 2002. Knowing what you know and what you don't: Further research on metacognitive knowledge monitoring. research report no. 2002-3. College Entrance Examination Board .

Wolters, C.A., Pintrich, P.R., 1998. Contextual differences in student motivation and self-regulated learning in mathematics, english, and social studies classrooms. Instructional science 26, 27–47.

Zepeda, C.D., Richey, J.E., Ronevich, P., Nokes-Malach, T.J., 2015. Direct instruction of metacognition benefits adolescent science learning, transfer, and motivation: An in vivo study. Journal of Educational Psychology 107, 954.