

Understanding and enhancing students' use of evaluation strategies during physics problem solving through reflective prompts

Elien Sijmkens^{1b}

*Faculty of Engineering Science, KU Leuven, Leuven, Belgium
and LESEC, KU Leuven, Leuven, Belgium*

Tinne De Laet^{1b}

*Faculty of Engineering Science, KU Leuven, Leuven, Belgium
and LESEC & LIVO, KU Leuven, Leuven, Belgium*

Mieke De Cock^{1b*}

*Department of Physics and Astronomy, KU Leuven, Leuven, Belgium
and LESEC & LIVO, KU Leuven, Leuven, Belgium*



(Received 24 May 2024; revised 17 April 2026; accepted 4 May 2026; published 9 July 2026)

The Disciplinary Learning Companion (DLC) is an online reflection tool designed to enhance students' engagement in metacognitive activities during physics problem solving. It achieves this through providing reflective discipline- and topic-specific prompts on the solving process. This study focuses on evaluation strategies and aims to understand whether, how, and when students use evaluation strategies during physics problem solving, whether the DLC's reflective prompts trigger the use of evaluation strategies and the extent to which this use is transferred to future problem solving. A total of twelve first-year bioscience engineering students each participated in four think-aloud interviews over a four weeks period. In these interviews, the students first solved a physics problem and then reflected on it supported by the DLC. This study focuses on students' use of evaluation activities as part of their metacognitive process. To assess students' current metacognitive levels, we examined the evaluation strategies they used spontaneously. In addition, we evaluated the impact of the DLC by studying the evaluation strategies students used after being prompted and those transferred to subsequent problem-solving attempts. The analysis revealed that students spontaneously applied evaluation strategies to numerical problems but rarely to symbolic problems. The DLC's reflective prompts successfully triggered the use of evaluation strategies, though infrequently. Similarly, there were few instances of transferring evaluation strategies from one think-aloud interview to another.

DOI: [10.1103/y6cq-bklb](https://doi.org/10.1103/y6cq-bklb)

I. INTRODUCTION

Since problem-solving performance is linked to students' metacognitive abilities [1] and their engagement in metacognitive activities [2], supporting metacognition is essential to help them to become proficient problem solvers. In fact, both monitoring progress [3,4] and assessing the potential of different approaches [5] are important to effectively address nonroutine problems. In addition, knowledge of various problem-solving strategies and knowing when and how to apply them is required to make

progress in solving a given problem [6]. Metacognition encompasses these aspects and in line with the framework of Flavell [7], it can be defined as “students' comprehension of their cognitive processes and their ability to regulate and monitor these processes based on the feedback received through learning outcomes” [8]. Both Schoenfeld [5] and Desoete *et al.* [9] found that engaging in metacognitive activities is characteristic for an expert-like approach to problem solving. Furthermore, it was shown to be beneficial for promoting transfer of newly learned strategies to future problem-solving endeavors [3,10,11]. In this study, we research the impact of a reflection tool, the Disciplinary Learning Companion, that aims at stimulating students' metacognitive abilities in the context of physics problem solving. We qualitatively study which metacognitive activities, in particular evaluation strategies, are elicited by this tool and to what extent they are transferred by students to new problems.

*Contact author: mieke.decock@kuleuven.be

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

A. Metacognition and reflection

A common approach to foster students' metacognitive abilities is making them reflect on their learning or problem-solving strategies. In fact, in their review on metacognitive instruction in science education Zohar and Barzilai [6] found that reflective writing, such as keeping reflective diaries or responding to reflective prompts, is the second most used approach for fostering metacognition after providing metacognitive prompts that remind students to employ particular metacognitive activities. This is not surprising, as metacognition is strongly linked to reflection. For example, Davis [12] describes metacognition as reflection on "goals or one's own thinking," while reflection on the content of a learning task itself refers to sense-making. Schraw [13] argues that reflection, together with extended practice, is essential to construct metacognitive knowledge and develop regulatory skills. Several studies that examined the effect of interventions using reflection to enhance students' metacognitive abilities and thereby their academic achievement confirmed that reflection can be an effective tool to do so. The research of Clark *et al.* [14] demonstrated, in the context of a fluid dynamics course, that engaging students in weekly reflection on their in-class problem-solving performance, combined with metacognitive instruction, enhanced students' academic achievement and self-reported use of regulatory skills. Similarly, Sabel *et al.* [15] found that students in an introductory biology course who engaged with answer keys supplemented with reflection questions had a better conceptual understanding of biological concepts than their peers who did not use these answer keys. Moreover, instructing students on how to use the answer keys and reflection questions resulted in additional learning gains compared to providing no instruction. This instruction entailed discussing the goal and various components of the answer keys and reflection questions and demonstrating how they might be used. Kim and Lim [16] found that supplementing an ill-structured problem-solving task with reflective scaffolds was more effective than providing supportive scaffolds for improving students' academic achievement and certain aspects of the problem-solving process, such as monitoring and evaluating. Here reflective scaffolds refer to scaffolds supporting underlying learning management processes by asking questions such as "What have you done?", while supportive scaffolds guide students in what domain-knowledge and ideas to consider and help students with creating associations between them [17]. Moreover, Kim and Lim [16] found that the effect of the reflective scaffolds on academic achievement was even higher for students who already had a high level of metacognition.

B. Problem solving and evaluation

Many generic and domain-specific problem-solving models have been presented in the literature. After analyzing a selection of generic and physics problem-solving

models, Gok [8] summarized them into three fundamental steps: (1) identifying the fundamental principles, which includes developing a qualitative and quantitative understanding of the problem based on one's own knowledge and experience; (2) solving, which basically entails using the qualitative understanding of the problem to set up a plan for a quantitative solution and carrying out this plan; and (3) checking (or evaluating), which is assessing whether the solution is correct, satisfactory, and does not contain any shortages. As an example, these three fundamental steps can be recognized in the logical problem-solving model of the University of Minnesota that consists of five steps: focusing the problem, describing the physics, planning the solution, executing the plan, and evaluating the answer [8]. 'Focusing the problem' and 'describing the physics', which includes developing a qualitative description of the problem and using this qualitative understanding to write down quantitative relationships, respectively, correspond to the first step in Gok's model. 'Planning the solution' and 'executing the plan' constitute Gok's 'solving' step. In these steps one uses the quantitative relationships to describe the given physics phenomena with a set of equations, determines whether this set of equations yields a solution and how this solution can be obtained, and carries out the calculations. Finally, the checking step aligns with Gok's 'evaluating the answer', which entails checking the reasonableness and appropriateness of the solution.

As checking or evaluating the solution is one of the three fundamental steps for problem solving as identified by Gok [8], it is clear that evaluation is an essential part of problem solving. Moreover, many researchers have identified evaluation, together with planning and monitoring, as one of the metacognitive skills needed in problem solving [8]. In the context of problem solving, evaluation can be divided into evaluation of the *product*, which includes the final answer and the (written) reasoning steps to obtain that answer, and evaluation of the *problem-solving processes*, encompassing the efficiency and validity of the strategies used [16]. Similarly, Meijer *et al.* [18] identified 'verifying' as a typical metacognitive problem-solving activity for the evaluation phase, where verifying refers to checking activities as well as outcomes. The 'checking' step in Gok's problem-solving model and 'evaluating the answer' in the University of Minnesota's model focus on evaluation of the product rather than assessing the problem-solving processes. One could argue, however, that evaluation of the product or solution obtained can help students to assess the efficiency or validity of their problem-solving process. Indeed, if students notice that their final answer is not correct, it can trigger them to evaluate the appropriateness of the strategies used. If students can correctly link the quality of the final answer obtained with the appropriateness of the strategies used, they demonstrate a good level of metacognitive abilities [19].

C. Evaluation strategies in Physics Education Research

Although evaluation is a central component of many problem-solving models, relatively few studies in Physics Education Research (PER) have explicitly examined whether and how students evaluate their solutions, even when they are prompted to do so, and the transfer of acquired evaluation strategies to future problem solving.

First, Wilcox and Pollock [20], Lenz *et al.* [21], Loverude [22] and Akinyemi *et al.* [23] all confirm that students rarely engage spontaneously in evaluation, even at the upper-division level. In a study of student difficulties with separation of variables, Wilcox and Pollock [20] found that without prompting students to do so, very few of them engaged in the Reflection step of the ACER framework, which includes the use of evaluation techniques (units and dimensional analysis, limiting or special cases, physical plausibility, and comparison with expectations or known results). When students did evaluate their results, they most commonly checked units. The authors also reported that students struggled to interpret limiting behavior, which may help explain why limiting-case analysis was rarely employed. Loverude [22] likewise found limited use of evaluation strategies: fewer than 5% of students employed limiting-case analysis when evaluating expressions in a mathematical methods in physics course. Even when prompted to perform unit analysis, many students experienced difficulties, particularly as expressions involved less familiar quantities, more complex notation, or mathematical operations such as differentiation.

Similarly, Lenz *et al.* [21] argued that students require explicit instruction in special-case analysis if they are expected to use it productively as an evaluation strategy. More recent work by Akinyemi *et al.* [23] examined student responses to three evaluation tasks across different physics contexts in an introductory calculus-based physics course, drawing on both written work and interviews. They reported three key findings: most students did not evaluate solutions using strategies that experts would recognize as evaluation; many students relied on primarily computational approaches; and many students employed noncanonical but potentially productive strategies beyond unit analysis, limiting or special cases, and reasonable-value checks (called plausibility by Wilcox and Pollock [20]). Finally, in a study closely related to those of Loverude [22] and Akinyemi *et al.* [23], Burkholder *et al.* [24] asked students to evaluate problem solutions using different question formats. Once again, unit checking emerged as the most prevalent strategy, while limiting-case analysis was used much less frequently. Notably, the strategies students employed depended strongly on how the evaluation task was framed.

Second, several studies have also investigated instructional interventions designed to support evaluation. Warren [25] studied the impact of explicit evaluation tasks (e.g., unit analysis and limiting or special cases) in an introductory physics course and found improved performance on

some common exam problems relative to a comparison group. However, the gains were largely limited to topics directly addressed in the evaluation tasks, suggesting limited transfer. Similarly, Sikorski *et al.* [26] explicitly introduced unit analysis, limiting cases, and reasonable-value checks in a junior-level electricity and magnetism course. Consistent with prior work, unit analysis was the most frequently used strategy. When evaluating limiting cases, students typically focused on the mathematical form of the limits without articulating corresponding physical expectations. Students also struggled to use reasonable values effectively, reflecting limited intuition for what values are reasonable in the context of electricity and magnetism. Burkholder *et al.* [24] again confirmed that careful scaffolding is provided to teach students more advanced evaluation strategies. The findings from existing work therefore demonstrate that engaging students in evaluation activities in order to improve their problem-solving performance is an interesting but nontrivial task.

II. BACKGROUND AND RESEARCH QUESTIONS

Our research focuses on improving students' metacognitive abilities in the context of physics problem solving. To this end, we developed an online reflection tool, the Disciplinary Learning Companion (DLC), that provides discipline-specific reflective prompts to students after solving a problem (see Sec. III). In our previous research [27,28], we found a positive relation between students' interaction with the DLC and their academic achievement, conceptual understanding, and strategic approach to problem solving. However, the relation with the latter, which was used as an indirect measure of students' metacognitive abilities, was no longer significant after controlling for prior academic achievement [29]. The sample size in this study might have been too small to detect an effect of the DLC on students' strategic approach on top of the effect of prior academic achievement. Hence, a potential impact of the DLC on students' metacognitive abilities could not be confirmed via this indirect measure. Additionally, we administered a self-reported questionnaire, the Motivated Strategies for Learning Questionnaire (MSLQ) [30,31], to measure students' metacognitive abilities. Although it is a commonly used approach, the literature suggests that such self-reported questionnaires might not accurately measure students' metacognitive abilities. They depend on students' perception of their learning strategies, while it seems that many students are not able to accurately judge their own metacognitive abilities [32]. This led us to seek for alternative ways to investigate how students' interaction with the DLC relates to their use of metacognitive activities during problem solving. Qualitative observations of which metacognitive activities are used by the students when solving problems and interacting with the DLC would provide new insights. Furthermore, the ecological setting of the previous studies made it hard to disentangle the effect of the DLC from the impact of other variables affecting

students' academic achievement and metacognitive abilities. Qualitative observations can improve our understanding of the interplay between students' interaction with the DLC and their use of metacognitive activities. Therefore, we carried out think-aloud interviews, an approach also used by, e.g., Veenman *et al.* [33,34] to observe students' metacognitive behavior during task performance.

In our think-aloud interviews students solved a problem and afterwards reflected on the solving process guided by the DLC. This allowed us to observe students' use of metacognitive activities during problem solving, how they interact with the DLC, and which activities it triggers. By inviting each of the students to participate in multiple think-aloud interviews, we could observe which activities triggered by the DLC are transferred to new problem-solving endeavors. Students took four interviews over a time span of more or less four weeks, making our study a longitudinal study. Since evaluation is an important aspect of problem solving, as outlined above, we focus in this study on this phase of the problem-solving process. In particular, we focus on strategies to evaluate the final answer, i.e., the product of the problem-solving process, as this is explicitly targeted by one of the DLC's reflective prompts (see Sec. III) and is more straightforward to observe and interpret than other metacognitive activities in the think-aloud interviews. The research questions of interest are:

RQ1. Which of the seven DLC evaluation strategies (Sec. III B) do students use spontaneously? (*spontaneously used evaluation strategies*)

RQ2. When a student didn't use an evaluation strategy spontaneously, did the DLC's evaluation prompts or feedback trigger its use? (*triggered evaluation strategies*)

RQ3. When the DLC triggered a student to apply an evaluation strategy in a think-aloud interview, did the student then use it spontaneously in later think-aloud interviews? (*transfer of evaluation strategies*)

III. DISCIPLINARY LEARNING COMPANION

The goal of the Disciplinary Learning Companion (DLC) is to foster students' metacognitive abilities in the context of problem solving. We try to achieve this by triggering reflection on the solving process through discipline-specific or even topic-specific reflective prompts and personalized feedback. The DLC comprises online reflection modules each delving into a specific standard physics textbook problem. In this study, we focus on problems on Newtonian mechanics. The idea of the DLC is that students first (try to) solve a problem, and afterwards can reflect on their solving process and solution by completing the corresponding reflection module addressing this particular problem.

A. Reflection prompts and feedback

To support the reading of the next sections, Fig. 1 provides an overview of the DLC's components and what prompts and feedback the DLC provides.

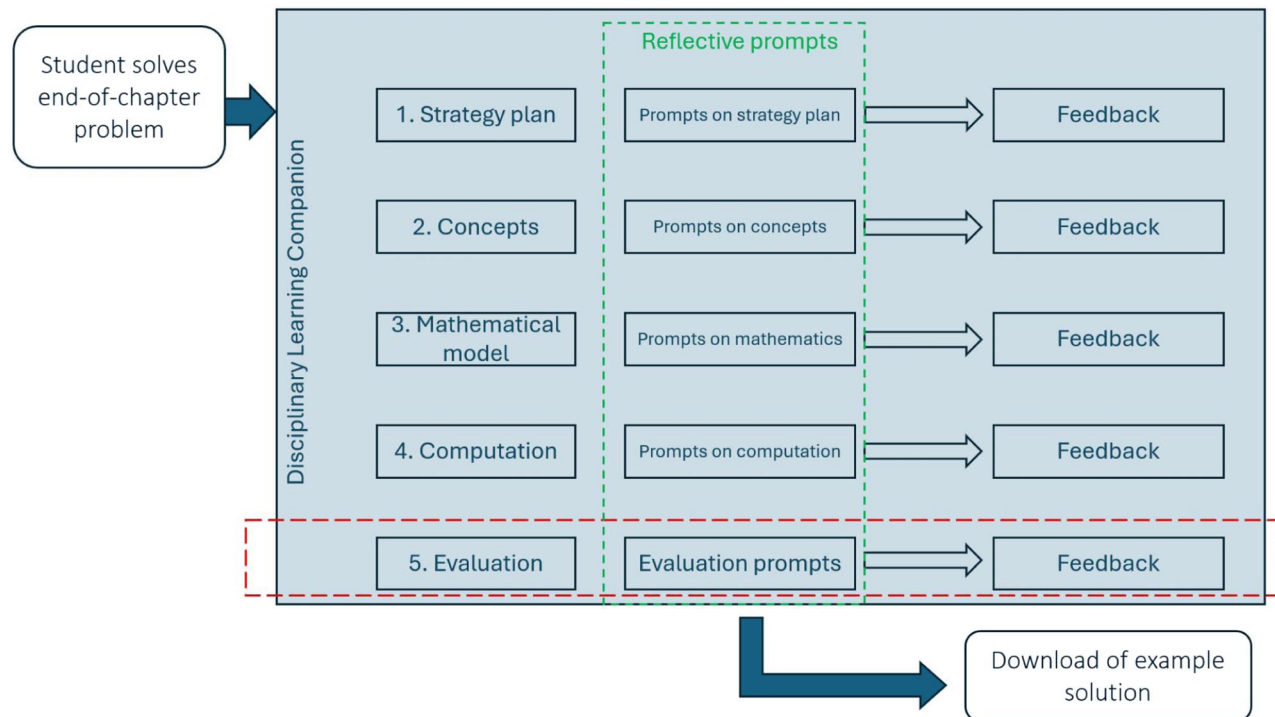


FIG. 1. Different components of the Disciplinary Learning Companion, and how prompts and feedback are used.

A reflection module of the DLC comprises reflective prompts that are organized according to five problem-solving dimensions that show similarities with the logical problem-solving model of the University of Minnesota [8]:

1. *Strategy plan*: setting up a well-considered strategy plan to address the problem at hand;
2. *Concepts*: identifying and comprehending the relevant discipline-specific concepts;
3. *Mathematical model*: translating the relevant discipline-specific ideas into a set of mathematical equations;
4. *Computations*: solving the obtained set of equations for the quantity asked for;
5. *Interpretation/evaluation*: interpreting and evaluating the obtained answer.

Each reflection module consists of 10–15 reflective prompts. We expect that students need 20–30 min to complete the reflection module after they have (tried to) solve the problem discussed in the reflection module. This was also confirmed by our observations in the think-aloud interviews.

The reflective prompts are presented in the form of multiple-choice questions, where the alternatives are based on common student difficulties and different possible strategies. Additionally, the options “something else” and “I do not know” are provided as alternatives in every question. In this way, students are not forced to choose an answer without being able to provide the reason for their choice. The reflective prompt about students’ interpretation of the final answer, referred to as the *evaluation prompt* in this study, asks about the strategies students used to check the correctness of their final answer. An example of this

evaluation prompt can be found in Fig. 2. Depending on the type of problem, different evaluation strategies are discussed in the reflection modules (see next Section).

Upon responding to a reflective prompt, the students receive feedback tailored to their response. The feedback addresses why the answer is (or is not) correct, or compares the advantages and disadvantages of different possible solving strategies. Moreover, the feedback includes actionable suggestions for students, e.g., prompting them to revise their solution approach based on the feedback provided. The *feedback to the evaluation prompt* discusses the advantages of checking the final answer and explains how different strategies can be applied in the problem at hand. Moreover, the feedback stimulates students to carry out the evaluation strategies they hadn’t tried before.

After finishing all topic-specific reflective prompts, students can download a model solution to the problem. Finally, students are asked to note something they have learned from solving the problem or completing the reflection module. This step is included to stimulate and facilitate the transfer of acquired concepts and strategies to future problem-solving endeavors.

B. DLC evaluation strategies

As we focus in this study on students’ use of evaluation strategies for assessing their final answer, we present in this section which evaluation strategies are covered by the DLC. In total, seven evaluation strategies (numbered E1–E7) occur in the DLC’s reflection modules. Not each of the seven evaluation strategies is discussed in each DLC

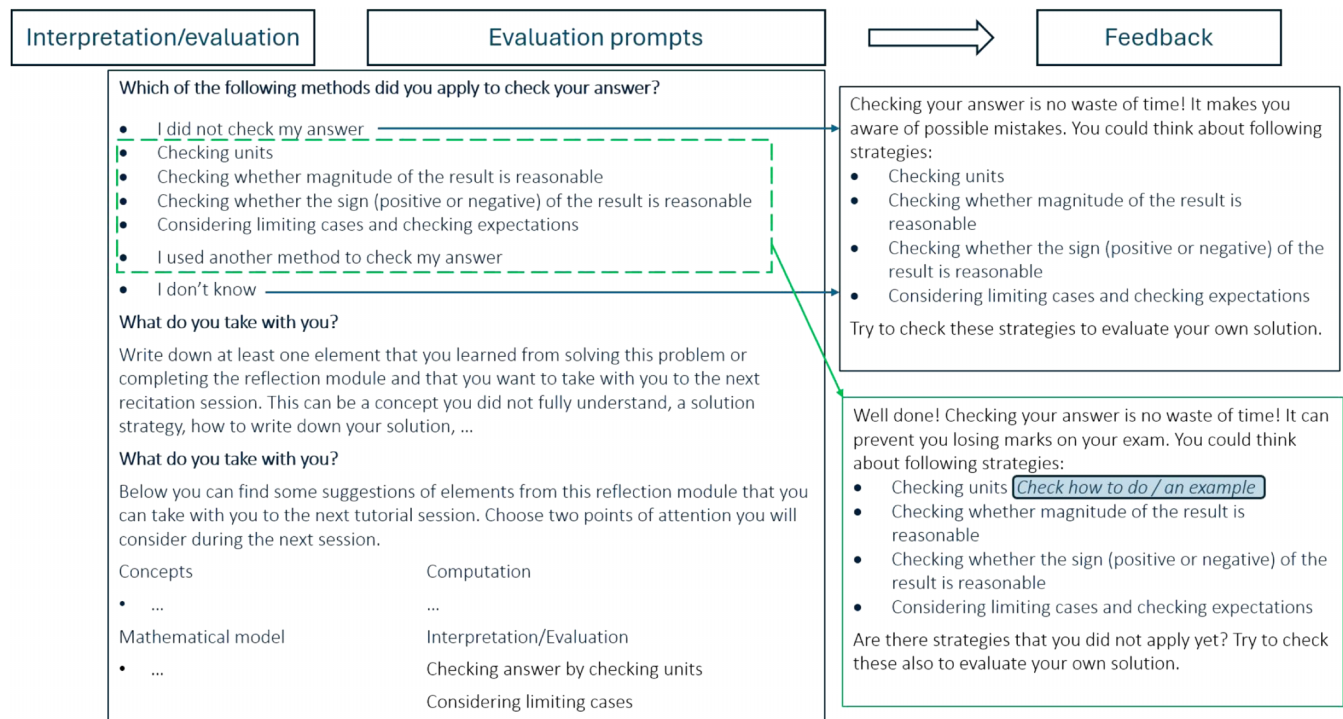


FIG. 2. Example of evaluation prompt and the feedback connected to the different answer options.

module as the applicability and usefulness of an evaluation strategy depends on the type of problem (numerical vs symbolic problem, one question vs multiple sub-questions). For example, some evaluation strategies might be more relevant for checking the solution to numerical problems, i.e., problems where numerical values are given for the known quantities and where a numerical solution is asked. Other strategies might be more relevant for symbolic problems, i.e., problems where no numerical values are given for the known quantities and one has to find a symbolic expression for the quantity asked for. Table I provides an overview of the evaluation strategies presented in the DLC's reflection modules students interact with in each think-aloud interview. For each evaluation strategy, we illustrate how it can be applied based on the problems solved in the think-aloud interviews (see Appendix A for the problem statements).

First, an evaluation strategy that occurred in every reflection module is checking the fundamental units, also referred to as unit analysis (E1) [25]. This strategy entails checking whether the equation or expression found is physically sensible by inspecting whether all its terms have the same units. For example, the problem for think-aloud interview 3 (Appendix A.3) asks to find an expression for a force F . The correct solution is $F = (m + M)g \tan \theta$, where m and M are masses of two objects involved, g is the gravitational acceleration and θ is an angle. Obviously, the unit of F is Newton (N). For the right-hand side, the unit of $m + M$ is kilogram (kg), the unit of g is meter per seconds squared (m/s^2), and $\tan \theta$ does not have a unit. Hence the unit of the term on the right is $\text{kg} \cdot \text{m/s}^2$, which is equal to N. We consider students to analyze units if they explicitly start from input units to check the output unit(s) algebraically. If a student's solution has inconsistent units, the student might have used incorrect equations, incorrectly recalled the units of one or more quantities involved, or made an error in solving the equations to the quantity asked for [25]. Note that the DLC does not discuss dimensional analysis [35], only unit analysis. This choice was made as the DLC targets an introductory physics courses, where students are typically not explicitly instructed about dimensions of quantities and equations [35].

TABLE I. Overview of the type of problem (numerical vs symbolic, one question vs multiple sub-questions) solved in each think-aloud interview (TAI) and the evaluation strategies discussed in the corresponding reflection module of the DLC.

TAI	Type of problem	Strategies
1	numerical, 1 question	E1, E2, E3, E4
2	numerical, 2 sub-questions	E1, E2, E3, E4, E5
3	symbolic, 1 question	E1, E2, E7
4	symbolic, 3 sub-questions	E1, E5, E6, E7

Second, there are some evaluation strategies that are mainly relevant for assessing the solution to a numerical problem. For these kind of problems, one can check whether the sign (E2) and the order of magnitude (E3) of the solution found is reasonable [36], i.e., whether it agrees with what one would expect based on experience and physics intuition. For example, the problem of think-aloud interview 1 (Appendix A.1) asks at which height above the ground two objects will collide with each other if one is dropped from a 50 m high rock and the other one is thrown up vertically with a certain initial speed. The problem asks to calculate a distance, which is always positive. As one of the objects is dropped from the 50 m high rock, the collision has to take place less than 50 m above the ground. This can be used to assess the reasonableness of the solution found. If either the sign or order of magnitude does not agree with expectations, one should be triggered to revise the equations used and/or the calculations done to solve the problem.

Third, one could insert the solution found into one of the equations (E4) that was set up to describe (part of) the problem and check whether the solution indeed satisfies this equation. This strategy is particularly interesting for detecting errors in the calculations done to solve the equations that were set up. For example, in the problem for think-aloud interview 2 (Appendix A.2), one has to find the minimum initial speed a car must have at the moment it leaves a ramp to clear a given point. To solve this problem, one has to set up an equation to describe the motion of the car in the x - and y -direction. After finding a value for the minimum speed, one can insert this value into the equation for the x - and/or y -direction to check whether the car indeed would reach the given point. In theory, this strategy could be applied to numerical as well as symbolic problems. However, in practice, applying this strategy for symbolic problems might become quite a hassle compared to applying it to numerical problems. Therefore, this strategy is included in the DLC only for numerical problems.

If a problem consists of multiple sub-questions that are closely related to each other, one can compare the solutions to the different sub-questions (E5). For example, we consider the problem for think-aloud interview 2 (Appendix A.2) again. The only difference between the first and second sub-question is the angle of inclination of the ramp. Based on experience and/or physics intuition, one can predict that the minimum speed should be larger when the car is leaving the ramp horizontally compared to when it is leaving the ramp with an upward vertical speed. Hence, the solution found to the first sub-question should be larger than the solution found to the second.

Finally, in the case of symbolic problems the DLC discusses two related evaluation strategies: considering functional dependencies (E6) and limiting cases (E7) (also referred to as special-case analysis) [36]. In fact, considering limiting cases is a special case of examining

functional dependencies. If one considers the functional dependencies in a symbolic expression for a quantity, one examines what happens with this quantity of interest if one of the other quantities in the expression changes in magnitude. Subsequently, one compares this with what would be expected based on experience or physical intuition. For example, the problem of think-aloud interview 4 (Appendix A.4) asks students to determine an expression for the angle of inclination θ that the banked turn of a frictionless bobsleigh track should have to make sure that a bobsleigh can take the turn with a given speed v_0 . The correct solution is $\theta = \frac{\arctan v_0^2}{gR}$ with R the radius of the turn and g the gravitational acceleration. If R becomes larger, i.e., the turn becomes less sharp, then θ has to be smaller, i.e., the banked turn has to be less steep. If v_0 becomes larger, i.e., the bobsleigh wants to take the turn at a higher speed, then θ has to become larger, i.e., the banked turn has to be steeper. Both observations agree with what can be expected based on experience. When examining limiting cases, the reasoning is similar, but with quantities approaching limiting values, such as 0 or infinity. Going back to the example of think-aloud interview 4, if R is approaching infinity, i.e., the bobsleigh track is approaching a straight track, then θ is approaching 0, i.e., approaching a track without inclination. Again, this agrees with expectations, as no inclination is needed to go straight on a frictionless track. If observations from the solution found do not agree with expectations, one should revise the solution regarding one's conceptual understanding of the problem, the equations used, and the calculations done.

IV. METHODS

A. Course context, participants, and study design

This study was carried out in the context of an introductory physics course, primarily focusing on Newtonian mechanics, for first-year bioscience engineering students ($N \sim 350$) at KU Leuven (Belgium). A standard textbook for introductory physics was used [37]. The course comprised two weekly 1, 5 h lectures, provided by a physics professor, and a 2 h problem-solving session per chapter, guided by teaching assistants. These problem-solving sessions take place after the lecture and consist of typical end-of-chapter problems in the textbook covering the topics handled in the previous lecture(s). During the first lecture, the first author explained the goal of the DLC and the aim of the research. Moreover, students were encouraged to volunteer to participate in the think-aloud interviews part of the research. It was motivated how they could benefit from the DLC as previous research indicated that interaction with the DLC is related to higher academic achievement [27]. Moreover, they were promised to receive additional feedback from the interviewer, i.e., the first author, on their problem-solving strategies after finishing all think-aloud interviews. Additionally, students who

completed all think-aloud interviews were rewarded with a gift card of 50 euros. Due to time limitations, twelve out of the about twenty volunteers were selected to participate in the research based on first come, first served base and agenda availability. In this study, these twelve students were given pseudonyms: Ana, Arin, Dana, Faith, Isabell, Jack, Kevin, Linn, Lola, Mary, Mia, and Mila. No further selection criteria were used although we have to remark that the students who volunteered can be assumed to be more motivated than average.

The think-aloud interviews allowed us to qualitatively observe which evaluation strategies students employ during problem solving, whether and how the DLC prompts and feedback trigger the use evaluation strategies. They consisted of two parts: *the problem-solving phase* where the student first solved a physics problem while thinking aloud, and *the reflection phase* where they reflected out loud on the solving process using the corresponding reflection module in the DLC. If a student was able to solve the problem, be it correct or incorrect, the student could indicate that they finished solving the problem and wanted to start the reflection module. If a student was not able to solve the problem, either the student could decide to quit solving the problem and continue with the reflection module, or after 25–30 min the interviewer would cue the student to start the reflection module. Students could decide at each moment during completion of the reflection module to return to solving the problem. During the interview, the interviewer only encouraged the student to think out loud but did not ask additional or follow-up questions. As we wanted to study how students interact with the tool, we aimed to mimic as much as possible the situation of a student reflecting individually, guided by the DLC. An interview (solving the problem followed by reflection guided by the DLC) typically lasted for approximately 1 h.

To study the evolution of students' use of metacognitive activities over a longer period of time in which they interact with the DLC, we invited the participants for four think-aloud interviews spanning about four weeks (see Fig. 3). First, students attended the first problem-solving session addressing the first chapter of the course. Between the first and second problem-solving session they took the first think-aloud interview, where they had to solve and reflect on a problem addressing the topic of the first problem-solving session. Similarly, they took the second and third think-aloud interview between the second and third problem-solving session and between the third and fourth problem-solving session, respectively. Finally, the students took the final think-aloud interview after attending problem-solving session 4 and 5, since these two sessions addressed closely related chapters. Only one of the twelve participants, Faith, quit the research after two think-aloud interviews.

To analyze the interviews, all think-aloud interviews were audio- and video recorded. Moreover, students wrote

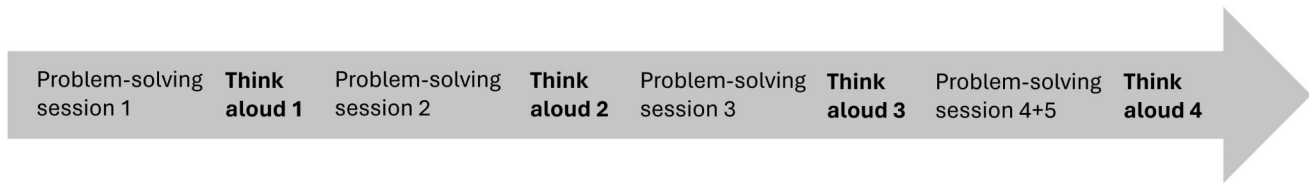


FIG. 3. Timeline of the think-aloud interviews.

with a smart pen while solving the problem and a screen recording was made while they were interacting with the online reflection module. Due to the COVID-19 pandemic, for two students, one of the interviews had to take place online. In this case, the video conference was recorded and students shared their screen while interacting with the reflection module. Moreover, after the interview, they emailed the interviewer a scan of their written solution to the problem.

B. Physics problems covered

The problems discussed in the DLC's reflection modules were standard physics textbook problems (Appendix A). The problem of the first reflection module (Appendix A.1) was related to the first chapter of the course on one-dimensional kinematics. The second reflection module (Appendix A.2) continued with two-dimensional kinematics and the problem that had to be solved consisted of two subquestions. The problems of the first two reflection modules were both numerical problems. The problem of the third reflection module (Appendix A.3) addressed the third chapter of the course, which was Newton's laws of motion. The last reflection module (Appendix A.4) dealt with friction and circular motion. This problem consisted of three subquestions, but only for the first and last subquestion students needed to set up and solve some equations to find the answer. The second subquestion only asked students to give an explanation of what would happen in the situation described with the aim of helping to solve the last subquestion. Table I presents an overview of the type of problems (numerical vs symbolic, one question vs multiple subquestions) solved in each think-aloud interview.

C. Data analysis

All think-aloud interviews were transcribed. Pseudonyms were used in the transcription and in this study (see Sec. IV A).

To answer our research questions, we analyzed whether students employed the DLC evaluation strategies spontaneously during the problem-solving phase (RQ1), whether the use of these strategies was triggered (RQ2) by the DLC's evaluation prompts or feedback during the reflection phase, or whether students did not use the evaluation strategies at all (RQ1).

While coding for spontaneous or triggered use, we observed that when students responded to the evaluation prompt, "Which of the following methods did you apply to check your answer?" (see Fig. 2), they often reported having used specific evaluation strategies without having verbalized them. This suggests that students might use these strategies implicitly, a limitation of the think-aloud protocol. To address this, we introduced the category "implicit spontaneous use." An example of implicit spontaneous use of checking the order of magnitude of the final results (E3) is illustrated in Ana's reaction:

[TAI 1, FS, E3, spontaneous—implicit] Ana: [checking for sense of] *reality. Yes, I did do that. I did it in my head, because I was like "28,7 meter is between 50 en 0, so it has to be between these two."*

Furthermore, we found that when triggered by a prompt, students sometimes discussed the potential of an evaluation strategy without actually implementing it. They would either elaborate on the strategy or briefly acknowledge it.

Based on the above observations, the following categorization was constructed:

1. **Spontaneous use:** The student (reported to) use the strategy without needing a trigger
 - (a) **Explicit spontaneous use:** the student used the strategy and carried out the strategy while thinking out loud during problem solving;
 - (b) **Implicit spontaneous use:** as a response to reflection prompts, the student reported to have used the strategy or having thought about it without mentioning it out loud during problem solving.
2. **Triggered use:** the student did not use the strategy spontaneously
 - (a) **Triggered actual use:** the student used the evaluation strategy after being triggered by the evaluation prompt and without needing any help of the feedback;
 - (b) **Elaborating on option/feedback:** the student did not actually use the strategy, but acknowledged it as possible strategy in the evaluation prompt and elaborated on it or on the feedback provided;
 - (c) **Reading option/feedback:** the student did not use the strategy, but acknowledged it as possible strategy in the evaluation prompt and/or reads feedback on this option without elaborating on it.

TABLE II. Number of think-aloud interviews in which a student used an evaluation strategy spontaneously (RQ1), summarized from the Table in Appendix B. E1–E7 are the evaluation strategies prompted by the DLC. En1–En4 indicate the other evaluation strategies used spontaneously by students. The table also includes the maximum possible number of occurrences [12 students $\times$ number of think-aloud interviews where strategy could be used (see Table I)].

Evaluation strategy	Description	Spontaneous use		Max. occ.
		Explicit	Implicit	
(E1) numeric	Checking units for numeric answer	3	12	24
(E1) symbolic	Checking units for symbolic answer	0	0	24
(E2)	Checking sign of final answer	3	12	36
(E3)	Checking magnitude of final answer	10	8	24
(E4)	Inserting solution in equation	6	0	24
(E5)	Compare solutions sub-questions	5	2	24
(E6)	Functional dependencies	0	0	12
(E7)	Limiting cases	1	0	24
(En1)	Checking solution answers question	2		48
(En2)	Checking symbolic answer contains known quantities	2		24
(En3)	Checking “reasonableness” without argumentation	5		48
(En4)	Revising (part of) the solution	3		48

3. **Not reading option/feedback:** the student did not use the strategy, did not mention it as possible strategy in the evaluation prompt, and did not read the feedback on it.

Each of the think-aloud interviews was coded using this categorization scheme. This was done iteratively. One of the researchers (first author) coded all student interviews. This coding was discussed during research meetings till consensus was reached.

For each think-aloud interview, we additionally recorded whether the student succeeded in solving the problem (FS: fully solved), succeeded in solving some subquestions of the problem (PS: partially solved), or did not solve the problem (NS: not solved). The solution did not need to be correct, but it had to address the problem statement.

By comparing these categorizations between all students and think-aloud interviews, and supplementing them with qualitative data from selected fragments, we formulate answers to our three research questions.

V. RESULTS

In this section, we discuss our observations on the use of different evaluation strategies by students. The categorization of students’ use of the different evaluation strategies over the four think-aloud interviews (TAIs) can be found in Appendix B.

As all interviews were conducted in Dutch, we translated the excerpts from the think-aloud interviews (TAIs) into English. To enhance readability, we adapted the students’ words, such as removing filler words like “um,” “wait,” “yeah,” and “uh.” We also included the students’ handwriting when it provided a better understanding of their quotes. To further enhance clarity, we rewrote the snippets of student responses to ensure that the handwriting itself is not a

hurdle in understanding the students’ written responses. In doing so, we stayed as close as possible to the original students’ answers, while translating Dutch words to English and removing parts that students had scratched out.

A. Spontaneous use of evaluation strategies (RQ1)

In this section, we discuss which of the seven DLC evaluation strategies (Sec. III B) were used spontaneously by students and how, which other evaluation strategies occurred, and which of the seven DLC evaluation strategies were never used spontaneously. Table II summarizes how often the different evaluations strategies were used spontaneously, both explicitly and implicitly.

1. Spontaneous use DLC evaluation strategies

The first strategy students used spontaneously is checking units (E1) for a numerical solution, although most students did so implicitly (Table II). One of the three students who checks the units of her solution spontaneously and explicitly is Mary.

[TAI 1, FS, E1, spontaneous—explicit] Mary: *I will quickly insert the units here [see Fig. 4]. This was in seconds, becomes seconds squared. Then you can cancel out this and then I indeed obtain meter. So that is the solution.*

After reading the evaluation prompt in the DLC, nine students over two TAI mentioned that they did check the

$$y_b = 50,0 \text{ m} - \frac{1}{2} \cdot 9,81 \frac{\text{m}}{\text{s}^2} \cdot (2,08 \text{ s})^2 = 23,78 \text{ m}$$

FIG. 4. Excerpt of Mary’s written solution to problem 1.

$$x = 50 - \frac{9.81}{2} \cdot 2.08^2 = 50 - 4.905 \cdot 4.34$$

$$= 50 - 21.289 = 28.7m$$

FIG. 5. Excerpt of Mila's written solution to problem 2.

units during the problem-solving phase without actually having verbalized this at that time, hereby representing 12 occurrences of implicit spontaneous use. From students' reaction to the evaluation prompt, it seems that not all of them understand what checking units entails and are not aware of this lack of comprehension themselves. For example, Mila indicates that she checked the units during the problem-solving phase when reading the evaluation prompt, while she actually only wrote down the expected unit in the final answer without checking for consistency throughout her calculations (see Fig. 5).

[TAI 2, FS, E1, spontaneous—implicit] Mila: [Reads evaluation prompt out loud] *Which of the following methods did you apply to check your answer? I did write down the units* [ticks option 'checking units' in response to evaluation prompt].

The second strategy that students used spontaneously (15 occurrences), albeit mostly implicitly (12 occurrences), was checking the sign of the final answer (E2). Students may have noticed during the problem-solving phase that the solution had a positive sign as expected, but did not mention it out loud. Similarly, students indicated that they checked the order of magnitude of the final result (E3) (18 occurrences) during the problem-solving phase, although this was only actually observed during the problem-solving phase in a subset of cases (10 occurrences). Again, students may have noticed that the obtained order of magnitude was reasonable without explicitly mentioning it. Students did use this strategy explicitly in cases where the order of magnitude of the solution found was not realistic, as, for example, in the first TAI of Faith.

[TAI 1, FS, E3, explicit—spontaneous] Faith: [Types into calculator] *And then 24 times 1.34, plus 4.905, times 1.34 squared. No, that is wrong anyway. Because I find that they collide on 40.97 meter. But the ball, well, the stone, that highest point that I calculated was only 23.39 meter. So one of the two is wrong, or both.*

Another strategy that students used spontaneously, but less frequently, was inserting the solution into an equation (E4) (6 occurrences). From students' reactions to the reflection question or feedback on this strategy, it seems that most students were familiar with this approach.

[TAI 2, NS, E4, triggered—elaborating] Faith: [Reads reflection question out loud] *Checking the answer by inserting the solution found in a given equation or any*

other equation found. Yes, probably I could insert that in $x(t)$ to see whether it became 22 meter.

Furthermore, some students compared the answers found to different subquestions (E5) (7 occurrences, 5 explicit and 2 implicit). Interestingly, this strategy was only spontaneously used in TAI 2 that dealt with a numerical problem, while it was never spontaneously applied in TAI 4 that dealt with a symbolic problem.

2. New evaluation strategies

Next to the strategies explicitly mentioned in the DLC, students used four other evaluation strategies, different from those covered by the DLC (notation En). Most of these strategies were rather superficial, meaning that they do not rely on physics knowledge or intuition and/or do not provide strong arguments to assess the solution's correctness. Nevertheless, they can still be valuable in helping students decide whether they are satisfied with their answer, whether the correct answer was found, or if more work is needed. Two students checked whether their solution answered the question in the problem statement (En1) (Table II); one of them was Arin.

[TAI 1, FS, En1, explicit—spontaneous] Arin: *This was also what was asked for: "how far." So this has to be* [the answer].

Checking whether their symbolic answer only contained known quantities (En2) was also done by two students (Table II).

Arin applies (En2) to decide whether he might have arrived at the correct solution of the problem of TAI 4.

[TAI 4, FS, En2, explicit—spontaneous] Arin: *The question was in function of R , θ , and μ_s . Yes, that g , that is just 9.81 meter per seconds squared. Yes, the same here. Other than that, I did express v in function of those other things* [see Fig. 6].

Although this strategy could be of value for textbook problems, its value for real physics problems is questionable.

A third evaluation strategy some students used is checking the "reasonableness" of their answer without providing any argumentation for it (En3, 5 occurrences).

$$\text{from } y: |F_n|(\cos\theta - \mu_s \sin\theta) = F_g$$

$$|F_n| = \frac{mg}{\cos\theta + \mu_s \sin\theta}$$

$$m x: \sqrt{\frac{1}{R} \left(\cos\theta + \mu_s \frac{mg}{\cos\theta - \mu_s \sin\theta} + \sin\theta \frac{mg}{\cos\theta - \mu_s \sin\theta} \right)} = v$$

||
v_min
v_max

FIG. 6. Excerpt of Arin's written solution to problem 4.

$$\frac{F_n \sin \theta}{m} = \frac{F}{m} \Leftrightarrow \boxed{F = \frac{m \cdot F_n \cdot \sin \theta}{m}}$$

FIG. 7. Excerpt of Isabell's written solution to problem 3.

This strategy was mainly applied to symbolic problems. Although we cannot know exactly what the students were thinking, it is possible that they assessed whether the form of their solution matched what they would expect from a standard physics textbook problem. They might have considered whether the solution looked too complicated or too simple or if it contained unexpected quantities. An example of this is Isabell's assessment of her solution to the problem in TAI 3.

[TAI 3, NS, En3, explicit—spontaneous] Isabell: *I really don't think that this is correct [see Fig. 7]. But right now I don't have ... Well, can I think a little more actually, well, till ...?*

Finally, there were some instances where students revised (parts of) their solution to check for mistakes (En4, 3 occurrences), such as Ana did for the problem of TAI 1. Although this strategy can help students detect small mistakes, such as calculation errors or the accidental use of wrong numerical values, it will not alert them to conceptual errors.

[TAI 1, FS, En4, explicit—spontaneous] Ana: *So I think, wait. Let's check for a moment. The velocity... [looks at solution], g squared, that is right. That is right, since g is negative. [Looks at solution] ...times t...divided by 25, by 24. That is 2.08, we have done 2.08. I will quickly ... [takes calculator], 2.08 times 2.08, divided by 2 times 9.8 times minus 1, plus 50 is ...Yes, I think that is correct.*

3. Rarely or never used evaluation strategies

Evaluation strategies that students did not or very rarely used spontaneously are checking units (E1) for a symbolic solution (0 occurrences), considering functional dependencies (E6) (0 occurrences), and considering limiting cases (E7) (1 occurrence) (Table II). Checking units for a symbolic solution appeared to be an unfamiliar strategy for students, while most seemed to know how to apply this strategy for numerical solutions. Some students, including Lola, even explicitly mentioned that they could not check the units in the case of a symbolic answer. If they do not read the feedback showing how to apply this strategy to the symbolic problem at hand, they unfortunately do not expand their repertoire for checking symbolic solutions.

[TAI 4, PS, E6, triggered—reading] Lola: *Checking the reality value or units or so is kind of hard, because it is all*

in symbols, so you can't really do that. It is not really poss...

Note that the strategies that were never or rarely used spontaneously by the students are mainly relevant for symbolic problems. In contrast, the strategies that were used spontaneously are primarily applicable to numerical solutions. Moreover, from the categorization (see tables in Appendix B), it can be observed that most students spontaneously checked their answers in some way for the problems in TAI 1 and TAI 2 (numerical problems), but seldom did so for the problems in TAI 3 and TAI 4 (symbolic problems). It should be noted that many students were unable to solve the problems in TAI 3 and TAI 4 and therefore could not demonstrate how they would evaluate their solutions to these symbolic problems.

B. Triggered evaluation strategies (RQ2)

In this section, we discuss which of the seven DLC evaluation strategies (Sec. III B) were used by students following a DLC evaluation prompt or feedback, and how they were applied. Table III summarizes the results.

One of the underlying goals of the evaluation prompt is to trigger students to evaluate their answer if they had not yet done so or to employ evaluation strategies they did not think about themselves. Unfortunately, this only occurred in 6 instances spread over 3 evaluation strategies. More often than actually using the strategy, students reacted to the evaluation prompt by elaborating on this strategy without actually using it (3 occurrences) or reading the feedback connected to this prompt but without elaborating on it (10 occurrences). Below, we elaborate on the instances where the DLC triggered the actual use of the evaluation strategy.

The actual use of the strategy of checking the sign (E2) was triggered by the evaluation prompt in two cases, one of which is illustrated below.

[TAI 1, FS, E2, triggered—actual] Jack: [Reads reflection question out loud] *Which of the following methods did you apply to check your answer? Checking sign of the result, positive or negative, for reality value. [Looks at solution] Sign of the result, positive or nega...Yes, that is right. It is a positive value, which it should be.*

As checking units (E1) for a symbolic solution is a strategy that was never used spontaneously, that students seemed to be unfamiliar with, and that some students even thought to be not applicable to symbolic solutions (see excerpt of Lola in Sec. VA 3), it is interesting to see if the DLC prompts or feedback would trigger the use of this evaluation strategy. Only one student, Kevin, used this strategy in TAI 4 after he was triggered by the reflection module to do so.

TABLE III. Number of think-aloud interviews in which a student used an evaluation strategy after being triggered by the DLC (RQ2), summarized from the table in Appendix B. The table also includes the maximum possible number of occurrences [12 students $\times$ number of think-aloud interviews where strategy could be used (see Table I)].

Evaluation strategy	Description	Triggered			Max. occ.
		Triggered actual use	Elaborating on feedback/option	Reading feedback/option without elaboration	
(E1) numeric	Checking units for numeric answer	0	5	6	24
(E1) symbolic	Checking units for symbolic answer	1	9	8	24
(E2)	Checking sign of final answer	2	3	10	36
(E3)	Checking magnitude of final answer	0	3	2	24
(E4)	Inserting solution in equation	0	13	3	24
(E5)	Compare solutions subquestions	0	4	7	24
(E6)	Functional dependencies	0	3	6	12
(E7)	Limiting cases	3	4	11	24

[TAI 4, PS, E1, triggered—actual] Kevin: *I did not check, yeah, my answer in this case, because it is all symbolic. I could not check the units. So I will see what we have here [see Fig. 8]. $v_{\max}$ is then the root of all symbols. θ has no units anyway, μ also has no units. g does have units, that is meter per seconds squared. And R is in meter. So we get meter squared per seconds squared as units. And the root of that is meter per seconds. So that is right [ticks option ‘checking units’ in reflection module]. So the units are okay.*

Another strategy that was only once applied spontaneously by one student in the problem of TAI 4 is considering limiting cases (E7). After being triggered by the feedback in the DLC’s reflection module, a few students knew how to reason about limiting cases and relate this to the solution found (3 cases of triggered actual use). The excerpt of TAI 3 of Jack gives a good example of this.

[TAI 3, FS, E7, triggered—actual] Jack: [Reads reflection question out loud] *Which of the following methods did you apply to check your answer? [...] I considered limiting cases and checked whether this agrees with my expectations. I would not directly know what I could expect as answer in this situation. [...] [Reads feedback out loud] Which force do you expect that is needed if θ is 0 degrees? [Student starts reasoning] So, if θ would be 0 degrees then there is no force needed. So if I fill in my formula [see*

$$\begin{aligned}
 & \left\{ \begin{array}{l} \cos\theta \cdot F_N - \sin\theta \cdot F_N \mu_s - mg = 0 \\ \cos\theta \cdot F_N \mu_s + \sin\theta \cdot F_N = \frac{mv^2}{r} \end{array} \right. \quad (F_N (\cos\theta - \sin\theta \mu_s)) \\
 & \quad \quad \quad = mg \\
 & \quad \quad \quad F_N = \frac{mg}{\cos\theta - \sin\theta \mu_s} \\
 & v_{\max} = \sqrt{\frac{r}{m} \cdot \frac{(\cos\theta \mu_s + \sin\theta) mg}{\cos\theta - \sin\theta \mu_s}}
 \end{aligned}$$

FIG. 8. Excerpt of Kevin’s written solution to problem 4.

Fig. 9], the tangent of 0 degrees is equal to 0. So then F is equal to 0. If that is 90 degrees, then the force [...] is infinite actually. Well, then it will never be able to carry it, it will always fall. And the tangent of 90 degrees is infinite, so then, you find infinite. So that is correct as well.

C. Transfer of evaluation strategies (RQ3)

A second underlying goal of the evaluation prompts and feedback is to enrich students’ repertoire of evaluation strategies. The feedback shows students how a particular evaluation strategy can be applied and how it helps assess the correctness of the solution. It should support students in transferring the evaluation strategy to future problem-solving endeavors. Unfortunately, we observed only four instances of such transfer, spread over three evaluation strategies in the TAIs. These instances occurred when students did not use a certain strategy spontaneously in an earlier TAI but used it spontaneously (explicit or implicit) in a later TAI.

A notable example of lack of transfer relates to checking units (E1) for a symbolic solution, one of the evaluation strategies that proved to be most difficult to students. As discussed in Sec. VA, this strategy was never used spontaneously. Students did not apply this evaluation strategy to the symbolic problem of TAI 4, even not after the DLC’s reflection module of TAI 3 introduced it explicitly.

We now focus on the cases where transfer was observed. Two students, Ana and Kevin, did not use the strategy inserting the solution in an equation (E4), a strategy that students seemed to be familiar with (see Sec. VA),

$$\frac{m_M + m_m}{F} = \frac{m_m}{m_m g \tan\theta} \Leftrightarrow F = (m_M + m_m) g \tan\theta$$

FIG. 9. Excerpt of Jack’s written solution to problem 3.

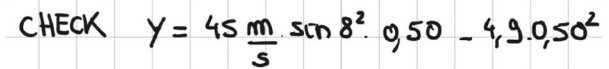
spontaneously in TAI 1 but interacted with the DLC's feedback on this strategy in TAI 1 and transferred it to TAI2.

[TAI 1, FS, E4, triggered—elaborating] Kevin: [Reads feedback out loud] *Checking answer by inserting the solution found in the given equation. [Looks at solution] I didn't do that. Ah, and now they ask to do it anyway. So I will quickly do so. [Reads feedback out loud] Checking whether the height of the ball and the stone are the same. [Student starts reasoning] So, I had the height, I had inserted the time, 2.08 seconds, in the height of the ball. So now I will also insert it in the height of the stone to check whether we find the same. So y is [types into calculator] 24 times the time, 2.083. Plus, the acceleration, minus 9.81, divided by 2, times the time, 2.083, squared. And then I become the same height, 28.7 meter.*

[TAI 2, FS, E4, explicit—spontaneous] Kevin: *So that was what we were looking for. So now I will quickly check whether that is correct. So I will see whether, if I eliminate the y-component of the velocity from this, that then indeed we ... With this equation, ... that then we have covered a distance of 1.5... No, I can not check it like that. How will I check it? [looks at solution]. I can eliminate the x-component of the velocity, I know x as well, so I can eliminate the t from it. And then I could optionally check the y. Ah, I will quickly check by the equation x is $v_{0,x}$ times t, I will quickly calculate the x, the $v_{x,0}$. I know x, then I eliminate the t from it and then I will quickly fill it in the y-equation to see whether I indeed find y is 0 after that many seconds. So $v_{0,x}$, check [writes]. [Starts doing calculations ...] And then we expect that we find 0 and that is also the case. So the check is also correct.*

[TAI 1, FS, E4, triggered—elaborating] Ana: [Reads feedback out loud] *Well done, checking your answer is definitely no waste of time. [...] You can always do the following things [...] sign of the result, checking answer by inserting the solution in the given equation or another equation found. More concretely, here it becomes whether $y_B(t)$ has the same height as $y_S(t)$ with t_1 the moment of the collision. Are there methods in this list that you did not apply yet? Try to apply them to assess whether your solution can be correct. [...] I did not really do that. I didn't feel like it, the y_B on moment t.*

[TAI 2, FS, E4, explicit—spontaneous] Ana: *Technically seen I can substitute that in my other, but than substi... Technically seen, I can do that. So your y ... So check here [writes, see Fig. 10]. So your y is than equal to v of 0, so that is that 45 meter per seconds, times that sine of 7 degrees, times 0.50. So that will be rounding off, so yes, to make it a bit easier with the calculator. Minus 4.9 times 0.50 squared. So y is equal to [types into calculator] 45 times the sinus of 7, times 0.5, minus 4.9 times 0.5 squared ... 1.517. That is close enough. That is strongly rounded off.*



$$\text{CHECK } y = 45 \frac{\text{m}}{\text{s}} \cdot \sin 8^\circ \cdot 0.50 - 4.9 \cdot 0.50^2$$

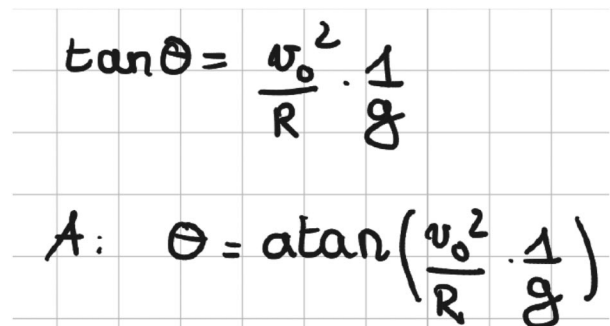
FIG. 10. Excerpt of Ana's written solution to problem 2.

One student, Arin, transferred limiting cases (E7), a strategy that was rarely used spontaneously by the students, to TAI 4 after interacting with the feedback about this strategy in the reflection module in TAI 3. In TAI 4, he used this strategy spontaneously to check whether his answer to the problem could be correct. He has more difficulties doing this for part c of the problem, but he tries to think about it (see Fig. 12).

[TAI 4a, FS, E7, explicit—spontaneous] Arin: [looks at problem statement] *That is question a. Then now question b. Ah no, quickly checking. My answer is just this [see Fig. 11]. Arctangent, so if θ is 0, then there is no force to the inside. So then the sled should just go straight. And if here θ is 0, then it means that tangent is 0. So then your speed should be 0 to be able to go in a circle. And if theta is 90, then your speed can normally be infinite. And yes, that is also correct.*

[TAI 4c, FS, E7, explicit—spontaneous] Arin: *Let's see whether that can be alright more or less. If θ is 0, then the velocity should be 0. θ 0, so then all this here is 0. Then this is 0, and then this is ... Ah no, then the velocity should not necessarily be 0, because there is still friction. The velocity should be very small, anyway. Then I find here that is something like square root μ_s divided by R. So that can be right. And if θ is 90, then your velocity can be infinite. So then this here is all 0. This here 90, so then this is 1. So then it says g divided by minus μ_s times sinus θ , minus μ_s . Then this becomes negative. [...] Other than that I would, I think, it should be alright. I'm just doubting now with these minus signs.*

One student, Mila, transferred checking units (E1) for a numeric solution, a strategy that some students used spontaneously (3 occurrences explicit vs 12 occurrences implicit), from one TAI to the next. After being triggered in



$$\tan \theta = \frac{v_0^2}{R} \cdot \frac{1}{g}$$

$$A: \theta = \text{atan}\left(\frac{v_0^2}{R} \cdot \frac{1}{g}\right)$$

FIG. 11. Excerpt of Arin's written solution to problem 4a.

$$\begin{aligned}
 \text{from } y: |F_n|(\cos\theta - \mu_s \sin\theta) &= F_z \\
 |F_n| &= \frac{mg}{\cos\theta + \mu_s \sin\theta} \\
 m a: \sqrt{\frac{1}{R} \left(\cos\theta - \mu_s \right) \left(\frac{mg}{\cos\theta - \mu_s \sin\theta} + \sin\theta \frac{mg}{\cos\theta - \mu_s \sin\theta} \right)} &= v \\
 &= v_{\max}
 \end{aligned}$$

FIG. 12. Excerpt of Arin's written solution to problem 4 c.

TAI 1, Mila adds the units to her final answers (which was without units until that point):

[TAI 1, FS, E1, triggered—elaborating] Mila: [reads the evaluation prompt out loud] *Try to apply these methods to check if your solutions are correct. Yes, the units. I will add them. Meter.*

Then, in TAI 2, again after being triggered, she states that she did use this strategy while solving the problem (implicit spontaneous use, student's solution shown in Fig. 5). However, note that Mila indicates that she did check the units as a response to the evaluation prompt, while she only wrote down the expected unit in the final answer without checking for consistency throughout her calculations.

VI. DISCUSSION

In this study, we discussed our observations from four consecutive think-aloud interviews with twelve students in an introductory physics course. In each interview, students solved a standard physics textbook problem and subsequently reflected on it by completing the corresponding reflection module in the DLC. We investigated which evaluation strategies were spontaneously used by the students during problem solving, which strategies were only used after being triggered by the evaluation prompt or feedback in the DLC, and which strategies were transferred to later think-aloud interviews.

All 12 students spontaneously applied evaluation strategies when solving physics textbook problems, hereby nuancing the earlier findings of Wilcox and Pollock [20], Lenz *et al.* [21], Loverude [22], and Akinyemi *et al.* [23], although our findings confirm that the repertoire applied was limited. The strategies students applied spontaneously were unit analysis for numeric solutions (E1), checking sign (E2) and order of magnitude (E3), inserting the solution into one of the equations (E4), comparing the solution to different subquestions (E5), and considering limiting cases (E7). This confirms the observations of Wilcox and Pollock [20], Sikorski *et al.* [26], and Burkholder *et al.* [24] who found that students most commonly used unit checking (E1) when solving physics problems. Similar to the conclusion of Akinyemi *et al.* [23] regarding reasonable-value checks, we find that for physics

problems in a mechanics context, checking sign (E2) and order of magnitude (E3) is also spontaneously used by students. We also confirm Wilcox and Pollock [20], Loverude [22], Lenz *et al.* [21], and Burkholder *et al.* [24] in their observations that spontaneous use of limiting cases (E7) is rare. Remark that only one of the seven targeted evaluation techniques was never used spontaneously: functional dependencies (E6). Additionally, we found that none of the students spontaneously applied unit analysis for symbolic solutions. This connects to our finding that the students spontaneously applied evaluation strategies predominantly when solving numerical problems. Only a few students evaluated their solutions to symbolic problems, and they used only a limited number of evaluation strategies to do so. This observation might suggest that students are familiar with strategies for checking numerical solutions but are still unfamiliar with strategies for checking symbolic solutions. This connects to the findings of Warren [25], who showed that students have a hard time applying evaluation strategies to topics not directly addressed in earlier evaluation tasks; of Loverude [22], who showed that, for instance, more complex notation affects students' evaluation strategy use; and of Burkholder *et al.* [24], who showed that how an evaluation task is framed strongly affects the evaluation strategies that students use. However, it is important to consider that the numerical problems for TAI 1 and TAI 2 were relatively easy compared to the other two symbolic problems for TAI 3 and TAI 4. From the categorization (see tables in Appendix B), it is indeed clear that more students were able to find a solution to the first two problems than to the last two problems. Students might be more familiar with the topic of the first two TAIs, which is kinematics, than with the topic of the last two TAIs, which is dynamics and circular motion. If students were unable to solve the problem, they had fewer opportunities to demonstrate the evaluation strategies they would use to assess the correctness of their final answer. This might explain why we observed more instances of evaluation for the numerical problems than for the symbolic problems. On the other hand, earlier research also indicated an impact of the representational format in which a problem is posed. Hung and Wu [38] not only report that students are more successful in solving numerical problems compared to symbolic ones, but they also found that most students perceived symbolic problems as more difficult than numeric ones. In their interviews, they found significant differences between both formats in how the interviewees performed when executing their plan and evaluating their solution. They related these results to greater cognitive load required to interpret the meaning of symbols and students' assumptions that a solution should be a number. That the use of symbols in problems tends to make them more difficult for students is also reported by Crisp and Grayson [39]. In other words, we should not conclude too easily that

students are more familiar with evaluation strategies for numerical problems than for symbolic problems. To make a fair comparison between students' familiarity with evaluation strategies for numerical and symbolic problems, future research should study which evaluation strategies students apply when solving numerical and symbolic problems on the same topic and of similar difficulty levels.

Remark that we introduced the category "implicit spontaneous use" to account for students who report, when being triggered by an evaluation prompt, to have used a particular evaluation strategy during the problem-solving phase although they did not verbalize this. The evaluation strategies that students reported using implicitly are indeed those that can be performed quickly in one's head: unit analysis for numeric solutions (E1), checking sign (E2) and order of magnitude (E3), and comparing the solution to different subquestions (E5). Students never claimed to have used evaluation strategies that are impossible to execute quickly in one's head, such as inserting the solution into one of the equations (E4).

Additionally, we observed that students were rarely triggered by the DLC's reflective prompts to immediately act and use the evaluation strategies suggested in the prompt. This was only observed in six instances [unit analysis for symbolic solutions (E1), checking sign (E2), and considering limiting cases (E7)], with five different students. This suggests that by the time students reach the evaluation prompt, they either do not know how to apply these strategies or do not feel the need to evaluate their answer further. Several explanations could account for this:

- Although it is possible that the teacher or TA mentions or models evaluation strategies during lectures or recitation sessions, this is not taught systematically. It is rather considered an implicit outcome and students are expected to develop these skills without explicit teaching. Not explicitly pointing toward particular evaluation strategies might imply that students do not learn how to apply them and do not realize their importance.
- Students might have already applied other evaluation strategies spontaneously, making them feel that using another one is unnecessary.
- The need to employ the evaluation strategies proposed in the DLC might depend on available resources, such as a model solution. The evaluation prompt appears at the end of the reflection module, when students have already responded to the prompts about the mathematical model and computations needed to solve the problem. The prompt about computations includes the correct final answer, helping students assess whether their calculations were correct. Thus, by the time they reach the evaluation prompt, students already know if their solution is correct or incorrect and may not feel the need to apply additional evaluation strategies.

- If students were unable to solve the problem themselves, applying these strategies to evaluate the answer they found with the help of the reflection module and the model solution might seem redundant. However, practicing these strategies could be beneficial for future problem solving.
- By the time they reach the evaluation prompt, students have already spent a considerable amount of time on the problem and reflection, which could have resulted in a low motivation level.

The limited effectiveness of the DLC's prompts confirms Loverude's [22] findings of remaining hurdles to use evaluation strategies after students were explicitly prompted to do so. Based on these possible explanations for the lack of triggered evaluation strategies, a potential improvement for the DLC is to include the evaluation prompt at the start of the reflection module rather than at the end. This way, students do not yet know whether their solution is correct, which could increase their motivation to evaluate their final answer. Additionally, if they realize through this evaluation prompt that their solution is incorrect, they might be encouraged to reflect more deeply on their solving process using the DLC's reflective prompts.

From the think-aloud interviews, it appeared that the transfer of evaluation strategies discussed in the DLC to new problems is limited. It was observed in only four students for unit analysis (E1), inserting the solution into an equation (E4), and considering limiting cases (E7). This aligns with the findings of Warren [25], who observed limited transfer of the benefits of providing evaluation tasks to teach students evaluation strategies to topics not covered in those tasks and with the call of Lenz *et al.* [21] for explicit instruction of evaluation techniques. Several reasons might explain the lack of transfer observed in the think-aloud interviews:

- The majority of the students were unable to solve the problems in TAI 3 and TAI 4. Consequently, they did not have the opportunity to demonstrate how they would apply newly acquired evaluation strategies to these problems, and thus, no transfer could be observed.
- We only tracked students during four think-aloud interviews. Some evaluation strategies were only relevant in two of the four interviews. More repetition might be needed to make these strategies part of students' problem-solving repertoire.
- Many students did not pay much attention to the feedback provided on the evaluation strategies. Literature shows that students perceive written feedback as more valuable when it is complemented with a feedback dialogue [40]. Additionally, although the feedback in the DLC is tailored to the problem at hand, it looks quite similar for every problem. Therefore, students might not recognize that it can teach them something new. More in-depth processing of the

feedback might be needed for students to internalize the evaluation strategies.

- They might not see the benefits of these strategies or might not feel the feedback is valuable for them. To stimulate students to engage with the feedback, they should have some motivation to do so. Including the evaluation prompt at the start of the reflection module could be an interesting idea, as students would not yet know whether their solution is correct and might feel more inclined to learn about strategies for assessing their final answer. Additionally, instructing students on how to use the reflection modules to improve their problem-solving performance might stimulate more in-depth processing of the feedback and enhance what students learn from it. Sabel *et al.* [15] found that providing instruction on the use of scaffolds resulted in additional learning gains beyond those achieved by just offering the scaffolds.

Our study highlights that the think-aloud protocol may not fully capture implicit cognitive processes. Students often reported using specific evaluation strategies in response to prompts without having verbalized them during the problem-solving phase. As they were not verbalized, they could not be captured by the think-aloud protocol. This suggests that some strategies are used implicitly. To address this, we introduced the category “implicit spontaneous use” to account for these unspoken strategies. We trusted that the students were accurately reporting their use of these strategies, but there is no way to ensure that this happened. This reliance on self-reporting introduces an element of uncertainty.

Additionally, the think-aloud protocol relies on students’ ability and willingness to articulate their thought processes. However, students sometimes discussed the potential of an evaluation strategy without actually implementing it, indicating a gap between recognition and application. This behavior suggests that while students may understand the value of certain strategies, they might not always integrate them into their problem-solving process due to various factors such as lack of familiarity, confidence, or perceived redundancy.

These limitations underscore the need for complementary methods to gain a more comprehensive understanding of students’ cognitive processes. Combining think-aloud protocols with other data collection techniques, such as retrospective interviews or eye-tracking, could provide deeper insights into how students approach problem solving and apply evaluation strategies.

Next to the above-mentioned limitations of the study related to the think-aloud interview method, we realize that all interviews were based on mechanics problems. This was not only a pragmatic choice. Mechanics typically is also the first introductory physics course at university where students are expected to develop problem-solving skills, and we think that supporting students to develop metacognitive skills makes sense early on. However, some of the evaluation strategies discussed in this study might be more suited for

mechanics problems than for problem in other domains like Electricity and Magnetism. As an example, whereas students might have some intuition on problems 1 and 3, this is not necessarily the case for problems related to electric or magnetic fields [26]. We do not claim that the evaluation strategies in the DLC are universal, but it was our aim to study to what extent the DLC supports students in implementing evaluation strategies by prompting them to reflect on it.

Finally, as mentioned in the Methods section, participation in the interviews was voluntarily, and we did not use any selection criterion. We simply invited all students in the course and selected participants on a first come, first served basis. We realize that this introduces a selection bias, and the findings probably reflect the behavior of highly motivated students that are willing to spend extra time on learning how to solve physics problems.

VII. CONCLUSIONS

In this study, we studied through qualitative observations the impact of the DLC on students’ use of metacognitive activities, in particular evaluation strategies, during problem solving. A total of twelve students participated in think-aloud interviews, where they first solved a standard textbook physics problem and afterward reflected on their solving process by responding to the reflective prompts in the DLC. Except for one student, each student participated in four think-aloud interviews, which allowed us to observe how students’ use of evaluation strategies evolved after (multiple) interaction(s) with the DLC. Analysis of the interviews revealed that evaluation strategies for assessing the solution to numerical problems were quite frequently and spontaneously used by the students, in contrast to strategies for evaluating symbolic solutions that students very rarely employed spontaneously. The reflective prompts in the DLC triggered students to use strategies they did not carry out spontaneously to assess the correctness of the solution found. However, this happened only in a few instances. Finally, a few students spontaneously used an evaluation strategy in a later think-aloud interview that they did not use spontaneously yet in an earlier think-aloud interview but was discussed in the DLC. This shows that the DLC has the capacity to stimulate transfer of evaluation strategies to future problem solving. Considering the rare instances of transfer or triggered execution of evaluation strategies, the goal of the DLC was not completely achieved. We provided possible explanations for this observation and what this implies for the design of the DLC. In summary, for several reasons students might not feel the urge to evaluate their final answer or to learn about strategies to do so. Starting rather than ending the reflection module with an evaluation prompt might increase students’ motivation to assess their own solution. In this way, at the point the evaluation strategies are discussed in the DLC, students do not know yet whether their solution is correct. This might make the usefulness of the strategies more clear to students.

ACKNOWLEDGMENTS

We would like to express our gratitude to the participants of this study, who dedicated their time to participate in the interviews and in this way showed interest in the research. Moreover, we would like to thank professor Hans Van Winckel and the teaching assistants for supporting our research in the context of their course and for motivating their students to participate in the research.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they contain sensitive personal information. The data are available from the authors upon reasonable request.

APPENDIX A: PROBLEMS SOLVED IN THE THINK-ALoud INTERVIEWS AND DISCUSSED IN THE DLC

The problems that were solved during the think-aloud interviews and discussed in the reflection modules of the DLC are presented below. These problems come from or are inspired by problems from Giancoli [37].

1. Think-aloud interview/reflection module 1

A ball is dropped from the top of a 50.0 m high cliff. At the same time, a carefully aimed stone is thrown straight up from the bottom of the cliff with a speed of 24.0 m/s. The stone and ball collide part way up. How far above the base of the cliff does this happen?

2. Think-aloud interview/reflection module 2

A stunt driver wants to make his car jump over 8 cars parked side by side below a horizontal ramp (see Fig. 13).

- With what minimum speed must he drive off the horizontal ramp? The vertical height of the ramp is 1.5 m above the cars and the horizontal distance he must clear is 22 m.
- If the ramp is now tilted upward, so that the “take-off angle” is 7.0° above the horizontal, what is the new minimum speed?

3. Think-aloud interview/reflection module 3

A small block of mass m rests on the sloping side of a triangular block of mass M which itself rests on a horizontal table as shown in Fig. 14. Assuming all surfaces

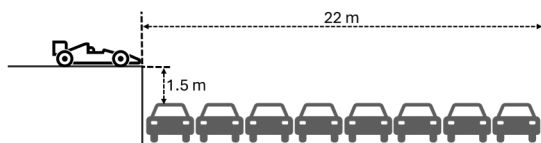


FIG. 13. Figure for problem think-aloud interview 2.

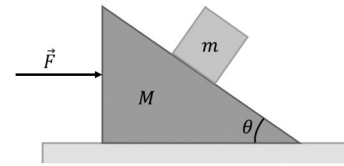


FIG. 14. Figure for problem think-aloud interview 3.

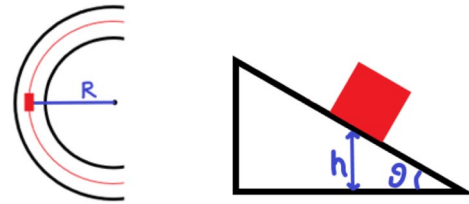


FIG. 15. Figure for problem think-aloud interview 4.

are frictionless, determine the magnitude of the force $\vec{F}$ that must be applied to M so that m remains in a fixed position relative to M (that is, m does not move on the incline). Hint: Take x and y axes horizontal and vertical.

4. Think-aloud interview/reflection module 4

The design of a new bobsleigh track contains a banked turn of radius R (see Fig. 15).

- One wants to design the turn such that a bobsleigh can take the turn with a given speed v_0 (without sliding across the track, i.e., the height h of the bobsleigh stays constant) and if there is no friction with the track. What should be the angle of inclination of the bobsleigh track?
- Suppose there is still no friction with the bobsleigh track. What happens if a bobsleigh wants to take the turn with a speed larger than v_0 ? What happens if a bobsleigh wants to take the turn with a speed smaller than v_0 ? Explain this. Hint: Notice that the normal force on the bobsleigh does not change.
- Suppose that now there is friction with the bobsleigh track and that the static friction coefficient is given by μ_s . Determine the minimal speed $v_{\min}$ and the maximal speed $v_{\max}$ at which a bobsleigh can take the turn (without sliding across the track) in function of R , θ and μ_s .

APPENDIX B: CATEGORIZATION OF EVALUATION ACTIVITIES

The categorization of students' use of the different evaluation strategies over the four think-aloud interviews can be found in the tables of Figs. 16. If the student only succeeded in finding a solution to one of the subquestions of the problem, it is underlined with a single line. If the student did not succeed at all in finding a solution, it is underlined with a double line.

	Explicit spontaneous use	Implicit spontaneous use	Triggered use	Elaborating on option/feedback	Reading option/feedback without elaboration	Not reading option/feedback
(E1) Units (numeric)	Isabell 1 Jack 1 Mary 1	Ana 1, 2 Arin 1, 2 Isabell 2 Jack 2 Kevin 1, 2 Linn 1 <u>Mary 2</u> Mia 1 Mila 2		<u>Faith 2</u> Lola 2 Mila 1	Dana 1 Faith 1 <u>Linn 2</u> <u>Mia 2</u>	<u>Dana 2</u> Lola 1
(E1) Units (symbolic)			<u>Kevin 4</u>	Ana 3 Arin 3 <u>Isabell 3</u> Jack 3, 4 <u>Kevin 3</u> Lola 3, 4 <u>Mila 3</u>	Ana 4 Arin 4 Dana 4 <u>Isabell 4</u> <u>Linn 3, 4</u> <u>Mia 3, 4</u>	<u>Dana 3</u> <u>Mary 3, 4</u> <u>Mila 4</u>
(E2) Sign	Ana 3 <u>Mary 2</u> Mia 1	Ana 1, 2 Arin 1, 2 Isabell 1 Kevin 1, 2 Linn 1 Lola 2 Mary 1 Mila 1, 2	Jack 1 <u>Mia 3</u>	<u>Faith 2</u> Jack 3 <u>Kevin 3</u>	Arin 3 Dana 1 Faith 1 Isabell 2, 3 Jack 2 <u>Linn 2, 3</u> <u>Lola 3</u> <u>Mila 3</u>	<u>Dana 2, 3</u> Lola 1 <u>Mary 3</u> <u>Mia 2</u>

FIG. 16. Categorization of students' use of the different evaluation strategies over the four think-aloud interviews. If the student only succeeded in finding a solution to one of the subquestions of the problem (PS: partially solved), it is underlined with a single line. If the student did not succeed at all in finding a solution (NS: not solved), it is underlined with a double line.

	Explicit spontaneous use	Implicit spontaneous use	Triggered use	Elaborating on feedback/option	Reading feedback/option without elaboration	Not reading feedback/option
(E3) Order of magnitude	Ana 2 Arin 1, 2 Dana 1 Faith 1 Lola 1 Mary 1, 2 Mia 1 Mila 2	Ana 1 Isabell 1 Jack 1 Kevin 1, 2 Linn 1 Lola 2 Mila 1		<u>Faith 2</u> Isabell 2 <u>Mia 2</u>	Jack 2 <u>Linn 2</u>	<u>Dana 2</u>
(E4) Inserting solution in equation	Ana 2 Arin 1, 2 Kevin 2 Mary 1 Mia 1			Ana 1 Faith 1, 2 Isabell 1, 2 Jack 1, 2 Kevin 1 Linn 1 Lola 1, 2 <u>Mary 2</u> Mila 1	Dana 1 <u>Mia 2</u> Mila 2	<u>Dana 2</u> <u>Linn 2</u>
(E5) Comparing subquestions	Ana 2 Arin 2 Isabell 2 Jack 2 Mila 2	Kevin 2 Lola 2		Arin 4 <u>Faith 2</u> <u>Mary 2</u> <u>Mia 2</u>	<u>Ana 4</u> Dana 4 <u>Isabell 4</u> <u>Jack 4</u> <u>Linn 2, 4</u> <u>Mia 4</u>	<u>Dana 2</u> <u>Kevin 4</u> <u>Lola 4</u> <u>Mary 4</u> <u>Mila 4</u>

FIG. 17. Categorization of evaluation strategies (continuation).

	Explicit spontaneous use	Implicit spontaneous use	Triggered use	Elaborating on feedback/option	Reading feedback/option without elaboration	Not reading feedback/option
(E6) Functional dependencies				Ana 4 Arin 4 <u>Kevin 4</u>	Dana 4 <u>Isabell 4</u> Jack 4 <u>Linn 4</u> Lola 4 <u>Mia 4</u>	<u>Mary 4</u> <u>Mila 4</u>
(E7) Limiting cases	Arin 4		Arin 3 Jack 3 <u>Mila 3</u>	Ana 3 <u>Kevin 3</u> <u>Lola 3</u> <u>Mia 3</u>	Ana 4 <u>Dana 3, 4</u> <u>Isabell 3, 4</u> Jack 4 Kevin 4 <u>Linn 3, 4</u> Mary 3 Mia 4	Lola 4 Mary 4 Mila 4
Checking solution answers question	Arin 1 Linn 1					
Checking symbolic answer only contains known quantities	Ana 4 Arin 4					
Checking "reasonableness" without argumentation	Ana 3, 4 Arin 3 <u>Isabell 3</u> Mila 2					
Revising (part of) solution	Ana 1, 2 <u>Isabell 3</u>					

FIG. 18. Categorization of evaluation strategies (continuation).

- [1] P. Rozencajg, Metacognitive factors in scientific problem-solving strategies, *Eur. J. Psychol. Educ.* **18**, 281 (2003).
- [2] P. Biryukov, Metacognitive aspects of solving combinatorics problems, *Int. J. Math. Teach. Learn.* (2004), <https://www.cimt.org.uk/journal/biryukov.pdf>.
- [3] B. Berardi-Coletta, L. Buyer, R. Dominowski, and E. Rellinger, Metacognition and problem solving: A process-oriented approach, *J. Exp. Psychol.* **21**, 205 (1995).
- [4] R. W. Hollingworth and C. McLoughlin, Developing science students' metacognitive problem solving skills online, *Australas. J. Educ. Technol.* **17**, 50 (2001).
- [5] A. Schoenfeld, Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics, in *Handbook for Research on Mathematics Teaching and Learning* (Macmillan Publishing Co., Inc., New York, NY, 1992), pp. 334–370.
- [6] A. Zohar and S. Barzilai, A review of research on metacognition in science education: Current and future directions, *Stud. Sci. Educ.* **49**, 121 (2013).
- [7] J. H. Flavell, Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry, *Am. Psychol.* **34**, 906 (1979).
- [8] T. Gok, The general assessment of problem solving processes in physics education, *Eur. J. Phys. Chem. Educ.* **2**, 110 (2010).
- [9] A. Desoete, H. Roeyers, and A. Buysse, Metacognition and mathematical problem solving in grade 3, *J. Learn. Disabil.* **34**, 435 (2001).
- [10] M. Bannert, C. Sonnenberg, C. Mengelkamp, and E. Pieger, Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance, *Comput. Hum. Behav.* **52**, 293 (2015).
- [11] E. Kapa, Transfer from structured to open-ended problem solving in a computerized metacognitive environment, *Learn. Instr.* **17**, 688 (2007).
- [12] E. A. Davis, Prompting middle school science students for productive reflection: Generic and directed prompts, *J. Learn. Sci.* **12**, 91 (2003).
- [13] G. Schraw, Promoting general metacognitive awareness, *Instr. Sci.* **26**, 113 (1998).
- [14] R. M. Clark, A. Kaw, and R. Guldiken, Metacognition instruction and repeated reflection in a fluid mechanics course: Reflective themes and student outcomes, *Int. J. Mech. Eng. Educ.* **51**, 243 (2023).
- [15] J. L. Sabel, J. T. Dauer, and C. T. Forbes, Introductory biology students' use of enhanced answer keys and reflection questions to engage in metacognition and enhance understanding, *CBE Life Sci. Educ.* **16**, ar40 (2017).
- [16] J. Y. Kim and K. Y. Lim, Promoting learning in online, ill-structured problem solving: The effects of scaffolding type and metacognition level, *Comput. Educ.* **138**, 116 (2019).
- [17] K. Cagiltay, Scaffolding strategies in electronic performance support systems: Types and challenges, *Innov. Educ. Teach. Int.* **43**, 93 (2006).
- [18] J. Meijer, M. V. J. Veenman, and B. H. A. M. van Hout-Wolters, Metacognitive activities in text-studying and problem-solving: Development of a taxonomy, *Educ. Res. Eval.* **12**, 209 (2006).
- [19] D. Lucangeli, P. E. Tressoldi, and M. Cendron, Cognitive and metacognitive abilities involved in the solution of mathematical word problems: Validation of a comprehensive model, *Contemp. Educ. Psychol.* **23**, 257 (1998).
- [20] B. R. Wilcox and S. J. Pollock, Upper-division student difficulties with separation of variables, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020131 (2015).
- [21] M. Lenz, P. Emigh, and E. Gire, Surprise! Students don't do special-case analysis when unaware of it, *presented at PER Conf. 2018, Washington, DC*, 10.1119/perc.2018.pr.Lenz.
- [22] M. Loverude, Quantitative reasoning skills in math methods, *presented at PER Conf. 2015, College Park, MD*, 10.1119/perc.2015.pr.046.
- [23] A. R. Akinyemi, M. E. Loverude, and J. R. Thompson, Solution evaluation strategies used by first-year physics students, *Phys. Rev. Phys. Educ. Res.* **21**, 020114 (2025).
- [24] E. Burkholder, L. Blackmon, and C. Wieman, Characterizing the mathematical problem-solving strategies of transitioning novice physics students, *Phys. Rev. Phys. Educ. Res.* **16**, 020134 (2020).
- [25] A. R. Warren, Impact of teaching students to use evaluation strategies, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020103 (2010).
- [26] T.-R. Sikorski, G. D. White, and J. Landay, Uptake of solution checks by undergraduate physics students, *presented at PER Conf. 2017, Cincinnati, OH*, 10.1119/perc.2017.pr.087.
- [27] E. Sijmken, M. De Cock, and T. De Laet, The disciplinary learning companion: The impact of disciplinary and topic-specific reflection on students' metacognitive abilities and academic achievement, in *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption* (Springer International Publishing, Cham, 2022), pp. 354–367.
- [28] E. Sijmken, M. De Cock, and T. De Laet, Scaffolding students' use of metacognitive activities using discipline- and topic-specific reflective prompts, *Metacogn. Learn.* **18**, 811 (2023).
- [29] E. Sijmken, M. De Cock, and T. De Laet, The impact of topic-specific vs generic self-reflection on students' metacognitive abilities, conceptual understanding, and problem-solving strategies, in *Technology Enhanced Learning for Inclusive and Equitable Quality Education. EC-TEL 2024*, edited by R. Ferreira Mello, N. Rummel, I. Jivet, G. Pishtari, and J. A. Ruipérez Valiente, Lecture Notes in Computer Science Vol. 15159 (Springer, Cham, 2024), 10.1007/978-3-031-72315-5_28.
- [30] P. R. Pintrich, and E. V. De Groot, Motivational and self-regulated learning components of classroom academic performance, *J. Educ. Psychol.* **82**, 33 (1990).
- [31] P. Pintrich, D. Smith, T. García, and W. McKeachie, A manual for the use of the motivated strategies questionnaire (MSLQ), Technical Report, University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning, Ann Arbor, MI, 1991, [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- [32] M. V. J. Veenman and D. van Cleef, Measuring metacognitive skills for mathematics: Students' self-reports versus on-line assessment methods, *ZDM Math. Educ.* **51**, 691 (2019).

- [33] M. V. J. Veenman, P. Wilhelm, and J. J. Beishuizen, The relation between intellectual and metacognitive skills from a developmental perspective, [Learn. Instr.](#) **14**, 89 (2004).
- [34] M. V. J. Veenman, R. Kok, and A. W. Blöte, The relation between intellectual and metacognitive skills in early adolescence, [Instr. Sci.](#) **33**, 193 (2005).
- [35] E. F. Redish, Using math in physics: 1. Dimensional analysis, [Phys. Teach.](#) **59**, 397 (2021).
- [36] E. Kuo, M. M. Hull, A. Elby, and A. Gupta, Assessing mathematical sensemaking in physics through calculation-concept crossover, [Phys. Rev. Phys. Educ. Res.](#) **16**, 020109 (2020).
- [37] D. C. Giancoli, *Physics for Scientists & Engineers with Modern Physics*, 4th ed. (Pearson Prentice Hall, Upper Saddle River, NJ, 2009).
- [38] C.-S. Hung and H.-K. Wu, Tenth graders' problem-solving performance, self-efficacy, and perceptions of physics problems with different representational formats, [Phys. Rev. Phys. Educ. Res.](#) **14**, 020114 (2018).
- [39] V. Crisp and R. Grayson, Modelling question difficulty in an A level physics examination, [Res. Pap. Educ.](#) **28**, 346 (2013).
- [40] M. Van der Schaaf, L. Baartman, F. Prins, A. Oosterbaan, and H. Schaap, Feedback dialogues that stimulate students' reflective thinking, [Scand. J. Educ. Res.](#) **57**, 227 (2013).