

Toward helping students develop error detection skills

Safana Ismael

*Department of Physics, North Dakota State University, 1340 Administration Ave, Fargo,
North Dakota, 58105*

Mila Kryjevskaja

*Department of Physics, North Dakota State University, 1340 Administration Ave,
Fargo, North Dakota, 58105*

Recent findings suggest that even those students who demonstrate relevant formal knowledge tend not to use it productively, especially on tasks that elicit intuitively appealing incorrect responses. Dual-Process Theories of Reasoning suggest that to catch a mistake, reasoners must engage in the process of error detection and override: recognize reasoning red flags, consider alternatives, and apply relevant knowledge to check their validity. It is, however, challenging for many novice physics learners to recognize what specific formal knowledge must be used as a criterion that needs to be satisfied for validating or rejecting a response. To help students develop skills necessary for error detection and override, we designed a sequence of systematic spaced practices in the context of Newton's 2nd law. We examined the effectiveness of this approach and identified specific factors that contribute to more productive engagement in error detection and override.

I. INTRODUCTION

Knowledge is central to productive reasoning but insufficient, especially in situations that present reasoning hazards [1–4]. Consider two students in an introductory mechanics course, Lisa and Danny (all names are pseudonyms). The students discuss forces acting on a box at rest on a horizontal surface while a constant horizontal 30N force is applied to the box, as shown in Fig. 1a [5,6]. Both students correctly argue that since the box is at rest, according to Newton's 2nd law, the force of friction must be 30N to the left. However, Lisa and Danny abandoned this line of reasoning on the follow-up question in Fig. 1b, where two identical boxes are now at rest on surfaces with different friction coefficients ($\mu_A < \mu_B$) while a horizontal 30 N force acts on each box. Lisa now argues that the force of friction on Box A is less than that on Box B because $\mu_A < \mu_B$. Danny agrees and supports this response with the expression for the maximum value of the static friction, $f_r = \mu N$. While this expression is not applicable in this case, it provides confirmation for Lisa's intuitively appealing but incorrect response. Neither student seems to question the validity of their answers by checking for consistency with Newton's 2nd law that they had just applied on a nearly identical problem presented without salient distractive features (i.e., different μ).

On question 1, Lisa and Danny demonstrated the knowledge and skills necessary to analyze forces acting on an object at rest. However, they did not transfer this knowledge to solve question 2 correctly. Inconsistent responses like Lisa's and Danny's often persist even after instruction [3,7–11]. To help students minimize reasoning inconsistencies, it is necessary to 1) understand the cognitive mechanisms responsible for productive and unproductive reasoning pathways and 2) pinpoint factors and instructional circumstances that help students enhance their reasoning skills necessary to validate or reject a response.

In this paper, we use Dual-Process Theories of Reasoning (DPToR) as a theoretical framework [1,2,12]. We describe a sequence of instructional interventions informed by DPToR and examine the roles of two factors that may impact student reasoning: the strength of relevant knowledge and the tendency toward cognitive reflection.

II. THEORETICAL FRAMEWORK

Research in cognitive psychology suggests that reasoning involves two processes: quick and subconscious process 1 and slow and deliberate process 2. Process 1 is often referred to as "gut feeling" or intuition. It immediately recognizes a given situation in a specific way based on prior

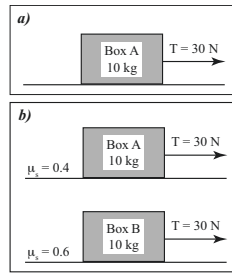


FIG. 1. Task 1

knowledge, experiences, and expectations. We agree with the parsimonious and pragmatic definition of intuition as “nothing more and nothing less than recognition. [2]” This recognition (often much more accurate for experts than novices) leads to the provisional

mental model, which becomes available for scrutiny by the slow, deliberate, and analytical process 2, as shown in Fig. 2. It could be argued that Lisa and Danny immediately and accurately recognized the task in Fig. 1a as “about balanced forces.” However, the salience of different μ in Fig. 1b may have overshadowed this approach and cued a provisional model involving static friction based on μ .

Process 1 cannot be turned off. As such, a provisional mental model is an entry point into any reasoning path, and process 2 is tasked with evaluating its validity. If a reasoner is confident in the provisional model, process 2 may be entirely circumvented such that a conclusion is reached via a path of cognitive frugality. Knowing when it is safe to jump to a conclusion via that path is linked to cognitive reflection skills, defined as a reasoner's tendency to mediate incorrect intuitive responses by engaging in process 2 analysis. The Cognitive Reflection Test (CRT) is often used to measure domain-general cognitive reflection skills [13–16].

If Process 2 intervenes, it still may not engage in productive error detection and override due to reasoning biases. Reasoners often look for evidence to justify the output of the intuitive process 1 if they already believe it is correct (i.e., confirmation bias) [17]. For example, Danny justified Lisa's comparison of the forces of friction by employing the mathematical expression for the maximum value of static friction that is not applicable in this case. If process 2 does scrutinize the provisional mental model and reasoning red flags are detected, then a new reasoning cycle begins by process 1 suggesting a new provisional mental model. The cycle repeats until a satisfactory answer is reached. In summary, to catch a mistake, a reasoner must engage in process 2, detect reasoning red flags, possess strong enough relevant knowledge to generate plausible alternatives, and assess their validity.

III. MOTIVATION AND STUDY DESIGN

The DPToR outlines cognitive mechanisms responsible for productive and unproductive reasoning pathways. In this study, we probed under what conditions students are more likely to reason productively and what factors may impact their reasoning approaches. We designed a longitudinal 8-week study in an introductory calculus-based mechanics

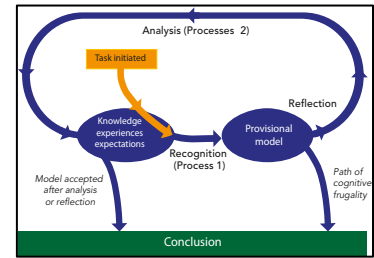


FIG. 2. Reasoning pathways according to dual-process theories

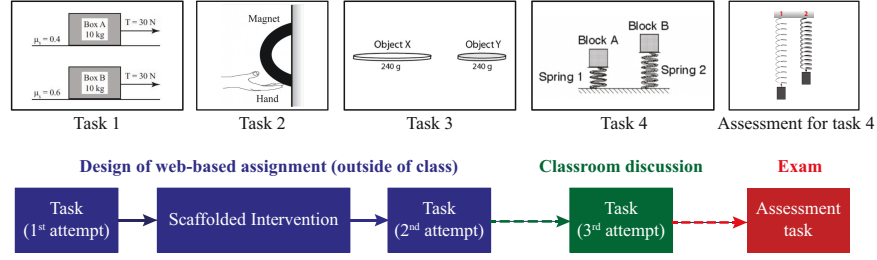


FIG. 3. Top row: tasks included in the systematic spaced practices; bottom row: instructional sequence for each spaced practice.

course. Students participated in 4 assignments spaced by ~ 10 -14 days and administered in a web-based format outside of class. Each assignment focused on a task similar to that included in the opening paragraph. Tasks for each assignment are shown in Fig. 3. Assignment 1 included task 1 and was administered after all relevant instruction on forces and Newton's laws. Assignments 2-4 included tasks 2-4, respectively, followed by scaffolded intervention.

The task from the opening paragraph is task 1. In task 2 a magnet weighing 10N is at rest on a refrigerator door while a hand supports the magnet from below with a 6N force; students determine the force of friction between the magnet and the door. In task 3, two pancake-like objects of different surface areas but the same mass fall to the ground after reaching terminal speeds; students compare the forces of air resistance on each object. In task 4, two identical blocks are at rest on different springs; students compare the forces on each block by a spring [18]. All tasks require the application of Newton's 2nd law to recognize that $F_{net}=0$ on each object, and therefore: in task 2, the force of friction between the magnet and the door points in the direction of the force by the hand; in task 3, the force of air resistance (F^{air}) is equal to the weight and therefore $F^{air}_1=F^{air}_2$; and in task 4, the force by the spring (F^{spr}) is equal to the weight so that $F^{spr}_1=F^{spr}_2$.

Many students gave incorrect answers, just like Lisa and Danny. On task 2, many reasoned that the force of friction "opposes the force by the hand" and must point downward. These students typically do not recognize the need to draw a free-body diagram and include the weight of the magnet or use Newton's 2nd law. When pointed to such an omission, many argue that they "forgot about the gravity." On task 3, many stated that the object with a larger surface area experiences a greater force of air resistance. Such responses often contain the mathematical expression $F^{air}=\rho ACv^2/2$ to justify the dependence of F^{air} on the surface area A . Students engaged in this line of reasoning neglect to recognize that this approach includes an inappropriate assumption that the objects move with the same terminal speed v . Similarly, many responses to task 4 are based on the salience of the different heights of the blocks on the springs. Students think that the different spring compressions, Δx , signify different F^{spr} and often justify this thinking with Hook's law while inappropriately assuming identical springs ($k_1=k_2$).

Through the lens of DPTOR, it could be argued that student responses of this nature stem from incorrect

provisional mental models cued by salient features of the tasks (e.g., different surfaces). Researchers argue that since the output of process 1 is subconscious and automatic, two approaches may be employed to improve performance. First, develop instruction focusing on a more accurate output of process 1 [19]. If the relevant knowledge is strengthened to the level of automaticity (as is often the case for physics instructors), then reasoners are more likely to immediately and subconsciously recognize its applicability correctly. Second, focus instruction on the more productive engagement of process 2 in error detection and override. Students should be able to recognize reasoning red flags and examine the validity of their provisional mental models by checking for consistency with more fundamental knowledge (e.g., Newton's laws). In our study, we create DPTOR-informed learning opportunities by 1) providing systematic spaced practices and 2) implementing scaffolded interventions for more productive process 2 engagement.

Scaffolded interventions were included in assignments 2-4 and followed each task as shown in Fig. 3. The interventions prompted the students to 1) consider alternative reasoning approaches, 2) apply relevant formal knowledge to help choose between alternative solutions, and 3) reconsider the initial response, if necessary. Below we focus on an intervention for task 4. The interventions for tasks 2 and 3 are based on similar principles.

First, students were asked to consider two expressions ($F^{spr}=k\Delta x$ and $F^{spr}=mg$) and determine which expression(s) must be used to compare the magnitudes of the forces on each block by the spring. This question was designed to nudge students to consider alternative reasoning approaches and, if appropriate, reject the reasoning based on the assumption that $k_1=k_2$. To further facilitate the error detection and override, students were asked to determine which one of the choices in Fig. 4 represents the correct free-body diagrams for the two blocks. This question was designed to make the information about the blocks being identical (i.e., equal W) more salient, thus prompting students to balance W and F^{spr} according to Newton's 2nd law and arrive at a correct answer. Finally, the students were asked whether they still agreed with their initial responses to task 4 and to elaborate on any changes in their reasoning.

A classroom discussion led by an instructor followed each web-based assignment, as shown in Fig. 3. Students considered a task from the assignment again (3rd attempt),

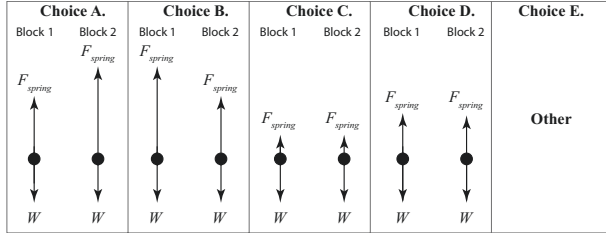


FIG. 4. Choices for task 4 intervention.

discussed their reasoning with peers, and submitted their individual answers via a classroom personal response system. The instructor then facilitated a discussion converging on a normative response. An assessment task shown in Fig. 3 was included on the course exam. Students considered two identical blocks hanging from springs and compared the forces by the springs on each block.

IV. PRELIMINARY RESULTS AND MOTIVATION FOR FURTHER INVESTIGATION

One common limitation of longitudinal studies is a reduced student response rate. In our case, 39 out of 60 students enrolled in the course completed all the assignments, reducing the sample size to $\sim 2/3$. Nevertheless, this potential selection bias provides an upper bound on our results since it could be argued that those students who completed all the assignments may be more motivated to receive a higher grade in the course.

As stated above, we designed this study to probe to what extent systematic spaced practices improve recognition of the applicability of Newton's 2nd law to novel situations and improve recognition of reasoning red flags that may lead to error detection and override. The expected desirable outcome was a higher success rate on each consecutive task.

The results in Table I show no clear improvement trajectory on the four tasks. The maximum success rate barely exceeds 50%. The scaffolded interventions do not appear to engage students in error detection and override successfully since only a few students improved on the 2nd attempt. The largest (but modest) improvement was observed between task 4 and the course exam assessment.

TABLE I. Results of the systematic spaced practices

Tasks	Correct (1 st attempt)	Correct (2 nd attempt)
Task 1	41%	-
Task 2	20%	33%
Task 3	54%	56%
Task 4	46%	49%
Assessment for task 4	67%	-

We examined performances on task 4 and the exam to gain further insights into student reasoning patterns. On tasks 1-3, $\sim 72\%$ of students answered at least one of the tasks correctly, demonstrating their abilities to recognize the applicability of Newton's 2nd law to a novel situation that presents reasoning challenges. This provides some evidence

that these students possess relevant knowledge and skills to solve task 4 correctly as well. Nevertheless, only $\sim 60\%$ of these students responded correctly to task 4 after the intervention. This leads to two hypotheses. First, students who possess relevant knowledge, but answer task 4 incorrectly, may have a higher tendency to jump to conclusions without engaging in process 2 (*i.e.*, have a lower tendency toward cognitive reflection). Second, the relevant knowledge is not simply present or absent. Instead, to reason productively, it must be instantiated to a greater depth, which may facilitate automatic recognition of its applicability (productive output of process 1) and/or increased confidence during error detection and override (productive engagement of process 2).

To test hypothesis 1, we used the cognitive reflection test developed and widely used in cognitive psychology and beyond [13–16,20]. The test consists of 3 items that cue intuitively appealing but incorrect responses that could be easily confirmed (or rejected) upon only brief reflection. For example, the first CRT item poses the question: “A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?” A solution based on basic arithmetic yields 5¢. Many, however, give a quick response of 10¢ without checking for its validity. A correct answer to each CRT item is assigned 1 point. Scores 2 or 3 indicate a stronger tendency to mediate intuition with analytical thinking.

To test hypothesis 2, we created a variable that indicates how many tasks 1-3 a student answered correctly after an intervention (upon 2nd attempt). In the following discussion, this variable, called *Strength of Knowledge*, provides a rough estimate of the level of knowledge instantiation. For example, a score of 1 indicates that a student not only possesses relevant knowledge but also was able to recognize its applicability to a situation eliciting intuitively appealing responses at least once. The higher the score, the deeper the knowledge is instantiated.

Since the most significant improvement in student performance occurred on the exam after an instructor-led classroom discussion, we explored how the shifts in performance between task 4 and the exam are linked to the *Strength of Knowledge* and a *CRT score*.

V. RESULTS AND DISCUSSION

The histogram in Fig 5a suggests no significant relationship between the performance on task 4 (labels *C* and *I* indicate correct and incorrect responses, respectively) and a CRT score. While more students with CRT=3 answered correctly and all students with CRT=0 answered incorrectly, the distributions of correct and incorrect responses for CRT=1 and CRT=2 are roughly the same. We used logistic regression analysis to formally verify this claim. Logistic regression is robust for a sample with more than 10 events (*i.e.*, number of correct responses) per predicting variable (*i.e.*, CRT score) [21]. The model for the probability of a

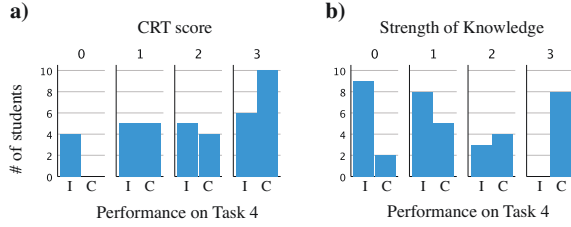


FIG. 5. Distribution of correct (indicated by C) and incorrect (indicated by I) responses to task 4 according to *CRT* score and *Strength of Knowledge* score.

correct response on task 4 as a function of a CRT score suggests that the CRT score is not a statistically significant predictor of success on task 4 ($\beta=0.6$, $p=0.07$) [15].

Even though our data do not support hypothesis 1, it does not mean cognitive reflection skills are irrelevant to productive reasoning. A replication study with a larger sample size is needed to verify the result. It is also possible that the strength of knowledge is a more powerful predictor in cases of systematic spaced practices. Indeed, a histogram in Fig. 5b does suggest a link between student performance and the strength of their knowledge. Students who consistently answered tasks 1-3 correctly had a 100% chance of correct response on task 4. Students with a knowledge score of 2 (or 1) were slightly more (or less) likely to answer task 4 correctly, and the students with a knowledge score of 0 were very unlikely to do so. The logistic regression model for the probability of answering task 4 correctly as a function of the *Strength of Knowledge* variable suggests a strong statistically significant relationship between the two variables ($\beta=1.4$, $p=0.002$), thus supporting hypothesis 2.

Finally, we examined the relationship between student performance on the exam, task 4, and a CRT score. The results suggest that nearly all students who answered task 4 correctly (bottom row in Fig. 6) also arrived at a correct answer on the exam. About half of the students who answered task 4 incorrectly (top row in Fig. 6) recovered on the exam and gave a correct response. There does not appear to be a dependence of the shifts in student performance between task 4 and the exam on the student CRT scores. As evident from the top row in Fig. 6, students who improved their reasoning on the exam were equally likely to do so regardless of their CRT score (except those with CRT=0 who consistently underperformed). The logistic regression analysis supports the conclusion that the performance on task 4 is a strong predictor of success on the exam ($\beta=2.3$, $p=0.007$) while a CRT score is not statistically significant.

We argue that the students who possess relevant knowledge do not always apply it successfully because intuitively appealing responses often overshadow its applicability. If students feel confident in their provisional mental models, they do not tend to recognize the need to apply formal knowledge to scrutinize intuition-based responses. This is consistent with prior findings that novices tend to compartmentalize their knowledge instead of reasoning from fundamental principles [22,23]. Many

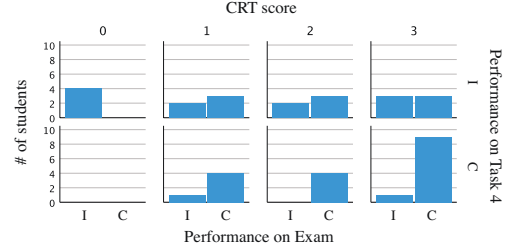


FIG. 6. Distribution of performance on exam according to a) CRT score and b) performance on task 4

students learn how to apply Newton's 2nd law to solve a variety of more computationally demanding problems but struggle to recognize how to use the same knowledge as a *criterion* that needs to be satisfied when validating or rejecting a response.

A classroom discussion incorporating peer-peer and instructor-student interactions helped some students to transfer correct reasoning to the situation presented on the exam. However, it is still an open question whether students will be more successful in applying this knowledge to similar tasks in different contexts (e.g., comparing buoyant forces on two identical blocks floating on surfaces of different liquids at different levels of depth).

VI. CONCLUSION

We employed DPTOR as a guide for developing instructional interventions to improve student reasoning on tasks that elicit intuitive incorrect responses. Analysis revealed that intuitive thinking has a strong hold on student reasoning even after systematic spaced practices. We examined factors that may impact student performance. In prior studies, cognitive reflection skills have been linked to productive reasoning on similar tasks. In this study involving systematic spaced practices, however, CRT scores do not appear to be a predictor of success. The strength of knowledge, measured by the success rate on similar tasks in different contexts, improves knowledge transfer to a different context. A classroom discussion also appears to facilitate a more productive application of relevant knowledge to a similar context. The improved performance may be attributed to two mechanisms: a strengthened recognition of the applicability of relevant knowledge and an improved recognition of reasoning red flags cued by a familiar context. However, more targeted instruction is needed to help students recognize how to apply fundamental knowledge as a criterion for validity checking. A replication study with a larger sample size is necessary to examine the validity of our findings and expand to different contexts.

ACKNOWLEDGMENTS

This material is based upon work supported by the national science foundation under the grants Nos. DUE-1821390, 1821123, 1821400, 1821511, 1821561.

References:

- [1] J. S. B. T. Evans and K. E. Stanovich, Dual-Process Theories of Higher Cognition: Advancing the Debate, *Perspect. Psychol. Sci.* **8**, 223 (2013).
- [2] D. Kahneman, *Thinking, Fast and Slow* (Farrar, Strauss, & Giroux, New York, 2011).
- [3] A. F. Heckler, The Ubiquitous Patterns of Incorrect Answers to Science Questions : The Role of Automatic, Bottom-up Processes, *Psychol. Learn. Motiv.* **55**, 227 (2011).
- [4] M. Kryjevskaja, P. R. L. Heron, and A. F. Heckler, Intuitive or rational? Students and experts need to be both, *Phys. Today* **74**, 28 (2021).
- [5] M. Kryjevskaja, M. R. Stetzer, and T. K. Le, Failure to Engage: Examining the Impact of Metacognitive Interventions on Persistent Intuitive Reasoning Approaches, 2014 *Phys. Educ. Res. Conf. Proc.* 143 (2015).
- [6] J. C. Speirs, R. Leuteritz, T. K. Lê, R. Deng, and S. W. Ell, Investigating the efficacy of attending to reflexive cognitive processes in the context of Newton's second law, *Phys. Rev. Phys. Educ. Res.* **19**, 10108 (2023).
- [7] A. F. Heckler and A. M. Bogdan, Reasoning with alternative explanations in physics: The cognitive accessibility rule, *Phys. Rev. Phys. Educ. Res.* **14**, 010120 (2018).
- [8] A. McInerney and M. Kryjevskaja, Investigating a collaborative group exam as an instructional tool to address student reasoning difficulties that remain even after instruction, *Phys. Educ. Res. Conf. Proc.* 327 (2020).
- [9] M. Kryjevskaja, M. R. Stetzer, and N. Grosz, Answer first: Applying the heuristic-analytic theory of reasoning to examine student intuitive thinking in the context of physics, *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **10**, 1 (2014).
- [10] W. N. Ferm Jr., J. C. Speirs, M. R. Stetzer, and B. A. Lindsey, Investigating student ability to follow and interact with reasoning chains, in *Proc. 2016 Phys. Educ. Res. Conf.*, edited by D. Jones, L. Ding, and A. Traxler (Sacramento, CA, 2016), pp. 120–123.
- [11] B. A. Lindsey, M. L. Nagel, and B. N. Savani, Leveraging understanding of energy from physics to overcome unproductive intuitions in chemistry, *Phys. Rev. Phys. Educ. Res.* **15**, 10120 (2019).
- [12] J. S. B. T. Evans, The heuristic-analytic theory of reasoning: extension and evaluation., *Psychon. Bull. Rev.* **13**, 378 (2006).
- [13] S. Frederick, Cognitive reflection and decision making, *J. Econ. Perspect.* **19**, 25 (2005).
- [14] M. E. Toplak, R. F. West, and K. E. Stanovich, Assessing miserly information processing: An expansion of the Cognitive Reflection Test, *Think. Reason.* **20**, 147 (2014).
- [15] C. R. Gette and M. Kryjevskaja, Establishing a relationship between student cognitive reflection skills and performance on physics questions that elicit strong intuitive responses, *Phys. Rev. Phys. Educ. Res.* **15**, 10118 (2019).
- [16] A. K. Wood, R. K. Galloway, and J. Hardy, Can dual processing theory explain physics students' performance on the Force Concept Inventory?, *Phys. Rev. Phys. Educ. Res.* **12**, 023101 (2016).
- [17] R. S. Nickerson, Confirmation bias: A ubiquitous phenomenon in many guises, *Rev. Gen. Psychol.* **2**, 175 (1998).
- [18] M. E. Loverude, C. H. Kautz, and P. R. L. Heron, Helping students develop an understanding of Archimedes' principle. I. Research on student understanding, *Am. J. Phys.* **71**, 1178 (2003).
- [19] B. D. Mikula and A. F. Heckler, Framework and implementation for improving physics essential skills via computer-based practice: Vector math, *Phys. Rev. Phys. Educ. Res.* **13**, 1 (2017).
- [20] M. Kryjevskaja, M. R. Stetzer, B. A. Lindsey, A. McInerney, P. R. L. Heron, and A. Boudreaux, Designing research-based instructional materials that leverage dual-process theories of reasoning: Insights from testing one specific, theory-driven intervention, *Phys. Rev. Phys. Educ. Res.* **16**, 20140 (2020).
- [21] E. Vittinghoff and C. E. McCulloch, Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression, *Am. J. Epidemiol.* **165**, 710 (2007).
- [22] E. Bagno, B.-S. Eylon, and U. Ganiel, From fragmented knowledge to a knowledge structure: Linking the domains of mechanics and electromagnetism, *Am. J. Phys.* **68**, S16 (2000).
- [23] M. Planinic, Assessment of difficulties of some conceptual areas from electricity and magnetism using the Conceptual Survey of Electricity and Magnetism, *Am. J. Phys.* **74**, 1143 (2006).