

Supporting sensemaking of physics equations: An intervention study in undergraduate coursesJulia Hofmann^{1,*}, Pascal Klein¹, Andreas Müller², and Josefine Neuhaus¹¹*Faculty of Physics, Physics Education Research, University of Göttingen, Friedrich-Hund-Platz 1, 37077, Göttingen, Germany*²*Faculty of Sciences, Department of Physics and Institute of Teacher Education, University of Geneva, Boulevard du Pont d'Arve 40, 1211, Genève, Switzerland*

(Received 19 September 2025; accepted 20 March 2026; published 22 May 2026)

A deep understanding of mathematical equations and their physical meaning is an important skill for physicists and therefore a central learning goal in university physics education. However, many students approach mathematical expressions in physics in a purely procedural manner, using them as calculation tools without critically evaluating their physical plausibility. Making sense of equations, that is, being able to interpret and evaluate equations within a physical context, is essential for the development of expertlike reasoning in physics. Frequently employed strategies that assist in the sensemaking process are dimensional, covariational, special, and limiting case analysis. Despite their importance, these strategies are rarely taught or practiced explicitly in university physics courses. The present intervention study systematically implemented sensemaking strategies within a university course for first-year physics students covering topics of electro- and magnetostatics as well as electrodynamics. Traditional exercises were enriched with tasks requiring students to critically evaluate the plausibility of self-derived equations. The tasks were integrated into recitation sessions and homework assignments. A total of $N = 67$ physics students participated in the study. A quasiexperimental pre-post design with two groups was employed to evaluate the effectiveness of the intervention approach. Results show that students highly valued all four strategies equally and reported an increased perceived competence in their application after the intervention. Significant improvements in performance were observed for dimensional ($d = 0.67$) as well as for special case analysis ($d = 0.50$). Students' performance in analyzing covariations only showed a tendency favoring the intervention group. No intervention-related impact on the performance of limiting case analysis was found, suggesting that this strategy may require more extensive practice. The findings underscore the importance and usefulness of explicitly teaching strategies to support students' sensemaking of physics equations.

DOI: [10.1103/mb7q-kf2s](https://doi.org/10.1103/mb7q-kf2s)**I. INTRODUCTION**

Understanding the meaning of a physical equation goes far beyond its computational use. For example, consider Coulomb's law, which describes the electrostatic force [1]:

$$\vec{F} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2} \hat{e}_r.$$

To an expert, this equation is not only a prescription for calculating the magnitude of the force acting along the straight line connecting two point charges q_1 and q_2 , separated by a distance r . Experts instead look beyond

the algebraic expression and build a meaningful connection to the underlying physical context. Central questions a physicist may ask in this context include: What happens, if the distance between the two charges becomes very large or approaches zero? Is the dimensional structure of the parameters involved physically plausible? How does the force depend on the distance between the two charges? Experts routinely engage in such considerations by applying several strategies to assess whether an equation is coherent with the physical context it describes, in particular in analyzing dimensions, covariations, special, or limiting cases [2]. These approaches are central for physicists to meaningfully interpret and make sense of physics equations [3].

In contrast, many undergraduate physics students frequently approach equations as mere computational tools. Students often engage in a “plug-and-chug” behavior by mechanically inserting values into equations without understanding the underlying physical principles or checking for plausibility [4–8]. However, mathematics plays a central role in physics that exceeds the pragmatic aspect of

*Contact author: julia.hofmann@uni-goettingen.de

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.

serving solely as a tool: It also has a communicative aspect as a “language” of physics and a structural function in terms of logical reasoning [6]. This highlights a central challenge in physics education: While mathematics is indispensable in physics, students frequently lack the skills and strategies needed to interpret mathematical expressions in a physical context. There is a broad consensus in the literature that it is important to support students in developing reasoning skills in order to make sense of mathematical expressions [4,9]. However, these skills are rarely taught explicitly in university courses and in many cases remain implicit [10,11].

Therefore, the present intervention study aims to explicitly integrate physics-specific strategies into undergraduate physics courses in order to foster students’ ability to evaluate the plausibility of equations and bridge the gap between mathematical expressions and their physical context. Exercises were developed to help students engage in the evaluation process for equations and were implemented within a course for first-year physics students. We systematically investigated and evaluated the effectiveness of this intervention approach.

II. BACKGROUND

A. Sensemaking

Sensemaking in the context of science refers to the active process of interpreting, understanding, and integrating scientific information in order to construct meaningful explanations of phenomena. Odden and Russ define sensemaking as a “dynamic process of building or revising an explanation in order to “figure something out”—to ascertain the mechanism underlying a phenomenon in order to resolve a gap or inconsistency in one’s understanding” [12] (pp. 5–6). From a cognitive science perspective, this involves forming connections between existing mental models and knowledge or integrating new knowledge into established cognitive frameworks. (This definition reflects just one of three strands of sensemaking identified in the literature by Odden and Russ [12]: stance toward science learning, cognitive processes, and discourse practices).

In physicists’ daily work, sensemaking is omnipresent. For example, in interpreting experimental data [13] or solving (theoretical) problems in general [3]. In problem-solving contexts, experts routinely ask themselves if the chosen approach is appropriate, if the results fit into a given context, and if the equations used and derived are consistent with their understanding of the physical situation. Equations play a pivotal role in problem solving, as they serve as a bridge between the mathematical formalism and the physical world. Unlike mathematicians, physicists “load physical meaning onto symbols” [4] (p. 3) and approach equations from a perspective that encompasses both the mathematical and the physical aspects of the systems they describe. Researchers argue that mastering mathematical tools in isolation, such as in traditional

mathematical methods courses, is not sufficient for physics learning [4,14]. It is essential to develop the ability to connect mathematical expressions with the underlying physical concepts and to make sense of how the variables and their relationships reflect physical behavior. This process of connecting equations to physical meaning heavily relies on prior knowledge, including formal understanding, everyday knowledge, and intuition [12].

B. Physics-specific strategies

Various physics-specific strategies can be employed to support the sensemaking process of physical equations. In the literature, such strategies are often referred to as sensemaking strategies [2,3], evaluation strategies [15,16], or quantitative reasoning skills [17]. Akinyemi emphasizes the importance of such strategies, stating that the “position of evaluation in models of problem solving in physics, and mathematical reasoning make evaluation strategies a great avenue for integrating physics and mathematics in ways that make sense physically” [15] (p. 21). A broad range of strategies has been identified in the literature, including graphical analysis [10], use of analogies [2,18], order-of-magnitude reasoning [19], and consideration of conservation laws [20]. In the present study, an intervention was developed and its effectiveness was studied on four strategies that stand out as expertlike approaches in research and the literature for assessing the physical plausibility of an equation: dimensional, covariational as well as special, and limiting case analysis. Here, not all of the previously mentioned strategies could be included due to practical constraints (in particular, limited teaching and learning time). Our selection is motivated by focusing on strategies that are explicitly and directly tied to the formal structure of equations, which facilitates their application across domains. In contrast, e.g., order-of-magnitude reasoning presupposes a wide range of highly context-specific knowledge about typical values and physical scales, which makes it more demanding to implement and harder to generalize across different areas.

The dimensional (or unit) analysis (DiA) is an approach that verifies the consistency and accuracy of an equation by examining the dimensions (or units) involved [15,21]. It is based on the principle that the dimensions on both sides of a correct physical equation are identical.

The relationship and variation of two or more quantities in an equation can be examined by means of covariational analysis (CoA). In particular, it comprises, how an alteration in one parameter influences another in terms of direction and rate [22,23]. Covariational reasoning is a well-established concept in the field of mathematics education research [22], however, its application in physics education research (PER) remains comparatively limited. Concepts that are often referred to in Physics Education Research as functional dependencies or relationships [2,8,24] and proportional reasoning [25] can be subsumed

under this framework. In this research, we refer to the analysis of covariations, as it is the commonly used term in mathematics education research [22].

Special case analysis (SCA) involves assigning specific values to a parameter in order to reduce or simplify a physical system to a well-known reference situation [2,3]. The aim is not to explore mathematical extreme scenarios, but to create a reduction to a well-understood and simpler special case against which the plausibility can be checked. Unlike arbitrary numerical values, special cases are deliberately selected to create a reference situation against which correctness and consistency can be evaluated. Consider, for example, an object on an inclined plane with inclination angle α . The object experiences an acceleration that can be described by $a = g(\sin \alpha - \mu \cos \alpha)$, where g denotes the gravitational acceleration and μ the coefficient of friction. In the special case of an inclination angle of $\alpha = 90^\circ$, the inclined plane becomes vertical, the friction term vanishes, and the situation reduces to the elementary case of a free fall with acceleration g . This reduction connects the formal expression to prior physical knowledge and enables a check of physical plausibility.

Limiting case analysis (LCA), also referred to as extreme case reasoning [18,26], examines the behavior of physical systems in the mathematical limits of parameters [9,27]. Nersessian characterizes limiting case analysis as thought experiments that idealize a physical situation: “Isolating the physical situation in thought allows us to manipulate variables beyond what is physically possible, and this creates data we did not possess before” [27] (p. 298). In contrast to special cases, limiting cases explicitly explore the behavior of a system as parameters approach extreme values in order to reveal structural properties of the underlying model. Consider, for example, a street light of mass m suspended by two identical cables of length l between two walls separated by a distance d . The force on one cable is given by $|\vec{F}| = \frac{mgl}{2\sqrt{l^2 - (d/2)^2}}$, with angle α (see Fig. 7 in the Appendix). In the limiting case where the distance approaches twice the cable length ($d \rightarrow 2l$), the cables become horizontal. In this limit, the force diverges ($|\vec{F}| \rightarrow \infty$), since an increasingly small vertical component of the force must balance the weight of the lamp. This limiting case reveals characteristic features of the situation and provides a powerful check of physical plausibility.

We acknowledge that the distinction between special and limiting cases is discussed inconsistently in the literature. Some authors use limiting case analysis as an umbrella term [9,28], while others treat limiting cases as a subset of special cases [16]. Following other authors [2,20], we consider special and limiting cases as two distinct sensemaking strategies in order to further investigate them systematically.

While checking the plausibility of an equation and making sense of it represent a central application of the

aforementioned strategies, their usefulness extends beyond this purpose. These strategies serve as tools for checking the consistency of an equation after it is derived (*a posteriori*) and for predicting the dependencies of different parameters (*a priori*) [29]. Thus, they are promising approaches for unknown or partially forgotten equations [29,30]. The *Buckingham–Pi theorem* formally establishes this predictive use of the analysis of dimensions [31]. In the following, however, the strategies are specifically employed as tools for sensemaking and for establishing coherence between the mathematical expression and the physical situation.

C. State of research

Students often learn mathematical tools as isolated procedures without developing a deep understanding of their connection to underlying physical concepts. However, the ability to interpret these tools within the physical context they describe is essential for meaningful learning in physics [4]. Research shows that students rarely engage in spontaneous reflection on answers. For example, Wilcox *et al.* [5] investigated students’ problem-solving process in the context of the application of Coulomb’s law. Strikingly, the authors found that only 8% of the students made any attempt to reflect upon their own result unprompted or to ensure it was physically plausible. Similarly, in an interview study by Lenz *et al.* [28], none of the participants spontaneously employed SCA when asked to reflect on equations, regardless of whether they had taken part in a course with integrated instruction on sensemaking strategies or a “traditional course.” Most of the students drew free-body diagrams, reflected upon the involved physical concepts, and some considered the dimensions. Moreover, White *et al.* [9] emphasize that students are more likely to engage in reflective practices when such steps are visibly valued and rewarded in university courses, for example, through structured prompts in recitation sessions or homework exercises.

Despite its importance, sensemaking is not systematically emphasized in university physics courses, and strategies supporting sensemaking are rarely taught explicitly and often remain implicit in traditional university courses [10]. Researchers emphasize the importance of explicitly incorporating reflection, evaluation, and sensemaking processes into university physics instruction [3,5,11,15,17]. Some research has been conducted to investigate the effective implementation of sensemaking strategies. The following paragraphs provide an overview of instructional approaches that have been implemented to incorporate sensemaking strategies within the context of university physics courses. The results give first insights into the potential for explicitly integrating such strategies into physics education.

As part of examining sensemaking strategies in physics instruction, White *et al.* investigated how students engage

in limiting case and dimensional analysis, using a design-based research approach [9]. The authors used LCA as a concept that includes the consideration of special cases as well as functional dependencies. In an electricity and magnetism course, exercises were implemented that prompted students to evaluate the plausibility of equations by analyzing limiting cases as well as dimensions. The study provides qualitative insights into how students handle these tasks and indicates that students can learn to apply such strategies productively. The implementation suggests that LCA can foster physical understanding and promote flexible, adaptive thinking. However, some students still approached the tasks merely algorithmically, without fully interpreting the equations in terms of the underlying physical situation. Another approach to integrating evaluation strategies in physics courses was conducted by Warren [16]. Evaluation strategies, such as DiA, SCA, and LCA, were embedded in homework and recitation tasks, and feedback was provided using the “scientific abilities” developed by Etkina *et al.* [32]. According to the rubrics, students’ abilities, e.g., regarding “evaluation,” were rated on a scale ranging from 0 to 3. The tasks included (correct and incorrect) equations that students evaluated using specific strategies, followed by a justification of the plausibility of the results. A control group received similar tasks but without the evaluation component. The findings indicate that the use and application of evaluation strategies can be effectively fostered through such instructional approaches. One main finding of the study is that students who received the intervention demonstrated significantly improved problem-solving performance on multiple-choice exam questions. They also employed evaluation strategies throughout the course more frequently than the control group. Warren concludes that strategies such as SCA, LCA, and DiA can be promoted through this approach and even help students benefit on exams.

A similar instructional approach was pursued by Hahn *et al.* [2], who integrated sensemaking strategies into a mechanics course. Students were introduced to various strategies, including DIA, SCA, and LCA, as well as reasoning about functional dependencies (CoA). Those strategies were embedded in traditional problem-solving tasks within weekly homework assignments, which outlined the specific strategy to be employed. After a few weeks, the explicit prompts were gradually faded out to encourage the independent use and selection of beneficial sensemaking strategies. To assess students’ spontaneous engagement with the strategies, the researchers analyzed the unprompted use of sensemaking strategies in homework assignments using an emergent coding scheme. The results showed that, among others, DiA was the most frequently used strategy (25%), followed by SCA (22%), CoA (14%), and LCA (4%). Hahn *et al.* conclude that explicitly embedding these strategies in standard courses

can effectively support students in developing a habit of evaluating and interpreting equations, thus enhancing their sensemaking skills in physics.

In summary, there is strong evidence that students require explicit support to develop reasoning skills that enable them to gain a deeper understanding of physical equations and to meaningfully connect mathematical expressions with physical contexts [4,10,14]. While some studies have demonstrated promising instructional approaches to foster such skills, these efforts remain limited in scope and have not yet been systematically implemented or rigorously evaluated. Some of the existing research has been primarily exploratory and diagnostic in nature, often lacking systematic design criteria for intervention tasks, established instruments, or manipulation checks. Moreover, little attention has been paid to students’ perspectives on the value of such strategies and to their perceived competence in applying them.

D. Research questions and hypotheses

Despite the importance of sensemaking strategies, the state of research reveals that the integration of these strategies has not been systematically investigated so far. While a few studies indicate that the application of sensemaking and evaluation strategies can have positive effects on students’ understanding and use of mathematical expressions in physics [3,9,16,33], empirical evidence remains scarce. The present study addresses this research gap by systematically implementing and empirically evaluating an intervention in the first year of an undergraduate physics degree program. The aim is to facilitate the development of a critical stance among students toward mathematical expressions with regard to the physical plausibility, thereby transcending the perception of equations as mere mathematical tools.

The objective of the study is to examine the feasibility of integrating four fundamental strategies: the analysis of dimensions, covariations, as well as special, and limiting cases. Given that students’ perception of the usefulness of these strategies has not been systematically investigated in previous studies, we address the following research question:

RQ1: How do first-year physics students perceive the usefulness of physics-specific strategies?

To assess the effectiveness of the systematic intervention approach, we examine students’ perceived competence as well as their performance in applying the four strategies. In light of findings reported in previous studies and the theoretical considerations outlined above, we propose that the implementation of explicit exercises concerning the application of sensemaking strategies ...

H1: ...will result in increased perceived competence among first-year physics students.

H2: ...will lead to improved performance measured by a performance test among first-year physics students.

TABLE I. Students' demographic characteristics for the sample at T0, the core sample in total, and for both intervention groups (mean and standard deviation).

	Sample T0	Core sample	DiACoA	SCALCA
Sample size	116	67	25	42
Female	26	12	7	5
Age (in years)	20.4 ± 2.5	19.8 ± 1.3	20.0 ± 1.4	19.6 ± 1.2
Grade of high school diploma ^a	1.7 ± 0.6	1.5 ± 0.5	1.5 ± 0.5	1.5 ± 0.4
Number of semester studied	2.5 ± 1.3	2.1 ± 0.5	2.2 ± 0.8	2.0 ± 0.0
Physics major	70%	79%	64%	88%
Physics teacher training	17%	12%	32%	0%

^aRanging from 1.0 (best score) to 4.0, indicated by students. Average grade of high school diploma in 2024 in Germany: 2.32 [35].

III. METHODS

A. Setting, sample, and data collection

During the summer semester of 2024, the study was conducted at the Faculty of Physics of the University of Göttingen as part of a standard experimental physics lecture for second semester students. The course primarily addressed the topics of electrostatics, magnetostatics, and electrodynamics, following the textbook by Demtröder [34]. In addition, the subjects of hydrostatics and fluid dynamics were covered during the initial two weeks of the lecture.

Data were collected at two occasions: at the beginning of the semester (T0, $N_0 = 116$) and midsemester (T1, $N_1 = 72$), with an interval of seven weeks between measurements. Demographic characteristics of the sample are summarized in Table I. The majority of students pursued a major in physics, while some participated in the teacher education program for physics teacher training. Teacher education students are required to enroll in the same introductory physics courses, but the emphasis merely indicates their intended study track. The analysis was performed with a core sample of $N = 67$ matched datasets from T0 and T1.

At the beginning of the course, the students were randomly assigned to one of two intervention groups: group DiACoA with $N = 25$ students focusing on the analysis of dimensions and covariations and group SCALCA with $N = 42$ students focusing on special and limiting case analysis. Due to conceptual similarities

between SCA and LCA, the two strategies were combined into one intervention group.

B. Study design and procedure

At the beginning of the semester (T0) and after a seven-week intervention period (T1), students completed a pre- and post-test (approx. duration 35 min) designed to assess the ability to apply the four strategies. Furthermore, students completed a questionnaire to assess their perceived competence and usefulness of each strategy. An overview of the study's design is presented in Fig. 1.

Throughout the semester, students participated in weekly recitation sessions, which included working on weekly assignments, a requirement for gaining admission to the exam.¹ The students of a particular recitation session all belonged to the same intervention group. As part of the intervention, the weekly exercises were enriched with tasks requiring students to evaluate the physical plausibility of equations using the strategies corresponding to the intervention group. Student solutions were corrected by tutors and discussed during recitation sessions. In designing the intervention, special attention was given to the integration of the four strategies with equivalent frequency throughout the seven weeks, and each strategy was applied to different equations. The topics of all exercise sheets as well as the corresponding intervention tasks are presented in Table II.

At the beginning of the course, all four strategies were introduced and discussed using worked examples, in accordance with the principles of example-based learning [36,37]. All students had access to a strategy guide, which served as a support throughout the semester. The guide included detailed, step-by-step instructions and worked examples for each of the four strategies.

Participation in the study was voluntary, and students generated personal pseudonymous codes so that their data could be matched from different time points without revealing their identities.

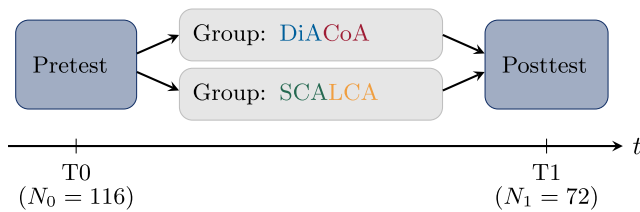


FIG. 1. Design of the study with two intervention groups DiACoA and SCALCA with an accompanying pre- (T0) and post-test (T1) within seven weeks.

¹To be admitted to the exam, students had to achieve at least 50% of the total points on all weekly assignments.

TABLE II. Topics addressed on weekly exercise sheets and intervention tasks during the study.

Sheet no.	Weekly exercise sheet topic - Topics of corresponding intervention task
1	Hydrostatics - Pressure of bike tire on an airplane
2	Fluid dynamics - Volume flow rate of a water jet - Magnus force acting on a rotating cylinder
3	Coulomb's law - Electric field of an infinite line of charge - Electric field of a uniformly charged ring
4	Gauss's law - Electric field of an infinitely long rod - Work done by moving a charge near a charged plate
5	Electrostatic potential and capacitor - Force between parallel plates of a capacitor - Capacitance of a coaxial cable
6	Dielectric and resistance - Voltage of a capacitor with dielectrics - Temperature-dependent resistance

C. Materials and measures

The translated version of all study material (originally German) is included in Supplemental Material [38]. It consists of the pre- and post-test, intervention tasks, and a questionnaire.

1. Intervention tasks

As a fundamental element of the course, weekly recitation sessions assist students in deepening their

comprehension of the lecture contents. Given that recitation sessions represent a promising location for the implementation of innovations [39], the intervention was implemented in this setting. Traditional exercises were enriched with intervention tasks that were inspired by preliminary work [2,9,16]. The students were assigned to check the physical plausibility of the equations they had derived. The assignments were structured and worded in the same way for each strategy to ensure comparability: “Check the equation derived in the previous part of the task for the corresponding physical quantity by analyzing dimensions/dependencies/special or limiting cases. Based on your analysis, is the equation physically plausible?” Figures 2 and 3 show examples of analogous tasks for both intervention groups regarding the movement of a point charge near a charged plate and the electric field of a rod.

Previous studies have shown that students particularly struggle to select a relevant parameter for a given situation and equation to investigate dependencies, special, or limiting cases [9]. In order to ease the selection process of relevant parameters for students, a relevant parameter was specified for the application of CoA, SCA, and LCA. At the end of each tasks, a reference to the strategy guide was made.

2. Performance tests

Students' ability to apply the four strategies was assessed using paper-based performance tests with open response as well as true-false tasks before (T0) and after (T1) the intervention. The tests covered different topics (pre: focus on mechanics; post: electrostatics). The wording of the test items was inspired by previous research [2,9,16,33]. In both performance tests, students were presented with equations and asked to evaluate their physical plausibility using the four strategies. To ensure consistency and

Task Homogeneously charged plate and point charge

Consider an infinitely extended plate with homogeneous surface charge $\sigma > 0$ oriented in the x - y plane.

- a) Determine the electric field $\vec{E}(\vec{r})$ of the plate using Gauss's law.

Now also consider a point charge $q > 0$, which is located in front of the plate.

- b) Calculate the work required to move the point charge q from point $\vec{x}_0$ to $\vec{x}_1$ in front of the plate as shown in the sketch. Let the distance be $|\vec{x}_1 - \vec{x}_0| = D$.
- c) Check the equation derived in b) by carrying out a dimensional or unit analysis. Based on your analysis, is the equation physically plausible?
- c) Check the equation derived in b) by considering the special case of a charge moving parallel to the plate. Based on your analysis, is the equation physically plausible?

(Note: You can read how to proceed when considering dimensions, special and limiting cases and dependencies at any time in the “Strategy Guide”.)

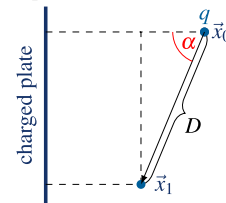


FIG. 2. Example task regarding the motion of a point charge near a charged plate with analogous tasks of the two groups (subtask c)). DiACoA considered dimensions (left) and SCALCA special cases (right).

Task Gauss's law (1)

- a) Given an infinitely long, non-conducting rod with radius R and an inhomogeneous charge carrier distribution

$$\rho(\vec{r}) = \begin{cases} \alpha r, & 0 \leq r \leq R \\ 0, & \text{otherwise,} \end{cases}$$

where r is the distance to the center of rotation of the rod and α is an arbitrary constant.

- | | |
|--|---|
| i) Use Gauss's law to calculate the electric field both inside and outside the rod. | |
| ii) Check the equation derived in i) by considering the electric field as a function of the distance r . Based on your analysis, is the equation physically plausible? | ii) Check the equation derived in i) by considering the limiting case of a distant observation point. Based on your analysis, is the equation physically plausible? |
| iii) How does the electric field change inside and outside the rod if it is an ideal conductor? | |

(Note: You can read how to proceed when considering dimensions, special and limiting cases and dependencies at any time in the "Strategy Guide".)

FIG. 3. Example task regarding the electric field of a rod with analogous tasks of the two groups (subtask ii). DiACoA considered covariations (left) and SCALCA limiting cases (right).

comparability, all items of the tests and the intervention exercises employed similar wording: "Check the dimensions/a relevant special or limiting case/a dependency between two parameters of the equation."

The pretest served as a "baseline measure" of students' initial competencies as well as a predictor. To ensure familiarity and to avoid confounding by unknown content, it consisted of eight tasks embedded in contexts from mechanics and some high-school electrostatics and magnetostatics topics (e.g., see Fig. 7 in the Appendix). In total, each strategy was applied twice.

Additionally, for DiA and CoA, students worked as a second task on true-false statements in order to accurately assess their abilities. For the dimensional analysis, the students had to identify whether a given expression had the dimensions of force or energy. One of the 14 true-false items was excluded from further analyses due to a high item difficulty ($P = 0.12$), indicating it was too challenging for most students. The true-false statements on CoA were inspired by the *Physics Inventory of Quantitative Literacy* by Brahmia *et al.* [25], which included, among others, tasks on covariational reasoning.

The post-test consisted of four tasks and reflected the topics addressed during the lecture (see Fig. 8 in the Appendix). Each strategy was covered with one task.

3. Perception of strategies and intervention

The pre- and post-tests were accompanied by a questionnaire comprising 11 items for each of the 4 strategies, resulting in a total of 44 items. Each strategy was introduced with a brief example to ensure clarity. The translated version of the items can be found in Table V in

the Appendix. Every item should be answered on a Likert-type rating scale ranging from 1 ("strongly disagree") to 4 ("strongly agree"). For each strategy, four items assessed students' perceived usefulness (e.g., "I believe that mastering the strategy is important"), while another four items measured their perceived competence in applying the respective strategy (e.g., "I am good at applying the strategy"). The questionnaire was developed by adapting items from the Intrinsic Motivation Inventory [40]. The items were modified by the authors to align with the specific aims of the study. To verify the implementation of the intervention, three additional items were included serving as a manipulation check and addressing the degree to which each strategy had been integrated in previous courses (e.g., "The strategy was explicitly subject of previous courses"; at T1 with the addition "including this course").

All three scales demonstrated satisfactory psychometric properties. Exploratory factor analysis, item-total correlations, and internal consistencies supported the reliability of the questionnaire. For each strategy and scale, the internal consistency (Cronbach's alpha) was higher than $\alpha > 0.72$, which is considered acceptable [41]. A translated version of the questionnaire, along with internal consistency coefficients, is provided in Table V in the Appendix.

D. Analysis methods and data processing

Standard quantitative methods were used to conduct the statistical analyses. All statistical analyses were performed using IBM SPSS Statistics software (version 29.0) and RStudio (version 2024.09.1 + 394).

Cohen's d was calculated to determine the effect size of comparisons of means (e.g., t test), providing a standardized

estimate of the magnitude of observed effects and enabling comparisons with previous research [42]. Effect sizes for paired values were calculated with consideration of the standard deviation of the predata (T0) [43].

To examine the effects of the intervention, analyses of variance and covariance were conducted. To determine group differences for the four strategies at T0, 2×4 analyses of variances were carried out. If pre- and post-time points were taken into account, the $2 \times 4 \times 2$ analysis of variance considered the factors group, strategy, and time point. The impact of the intervention on the performance was investigated with analyses of covariance controlling for the pretest score due to different pre- and post-tests used. Effect sizes for analyses of (co-)variance were reported using total η^2 [42]. Where applicable, *post hoc* tests were employed to identify statistically significant differences between groups, strategies, or time points. The corresponding requirements for the designated tests were verified accordingly. If the requirement of sphericity (Mauchly test) was not met, a correction was implemented, employing either the Greenhouse-Geisser or Huyn-Field correction corresponding to the ϵ value [44].

The qualitative data obtained from the open response tasks of the performance tests were analyzed using deductive category assignment in accordance with the approach described by Mayring and Fenzl [45]. The coding framework was based on the *Scientific Abilities* of Etkina *et al.* [32]. Specifically the rubric “evaluation” served as an orientation for rating students’ application of all four strategies in the pre- and post-test. The rubrics have been employed in similar contexts, for example in Warren’s study [16], particularly to score the application of DiA and SCA. While rubrics existed for DiA, SCA, and LCA, a new rubric for CoA was developed following the same structure and principles.

Each strategy was assessed using a rating scale ranging from 0 to 3 (0: no analysis, 1: analysis attempted, 2: flawed analysis, 3: flawless analysis). The adapted rubrics are documented in Table VI in the Appendix. To ensure the reliability of the coding process, two independent raters evaluated a subset of responses (approximately 20% of all pretest data). The interrater reliability was assessed using Cohen’s weighted kappa [46]. For all four strategies, Cohen’s weighted kappa was very high after an initial rating ($\kappa_w \geq 0.68$) and an intermediate discussion ($\kappa_w \geq 0.84$). Detailed reliability statistics are reported in Table VII in the Appendix. For further quantitative analyses, scores for each strategy were calculated with respect to all relevant test items and linearly transformed to a scale ranging from 0 (low level of performance) to 1 (high level of performance). To maintain coherence, the scales for the perceived competence, usefulness, and the manipulation check were linearly transformed to [0, 1] as well.

IV. RESULTS

In the following, the results are addressed according to the research questions and hypotheses. Descriptive values for all scales can be found in Table VIII in the Appendix. All test statistics of the conducted analyses of variances with the factors group, strategy, and time point are presented in Table III.

The core sample is representative in terms of demographic variables, but it shows a significant difference in the overall performance compared to the initial sample at T0 [$N_0 = 116$; $t(114) = -3.10$, $p = 0.002$, $d = 0.59$]. This indicates a systematic study drop out of low-performing students from T0 to T1.

The two intervention groups did not differ from each other with regard to demographic variables as well as the overall pretest performance [$t(65) = 1.08$, $p = 0.29$].

TABLE III Results of analyses of variances with factors group, strategy, and time point for several target variables as the dependent variable (DV), including test statistics and effect sizes (degrees of freedom, d.o.f.). Bold values indicate statistical significance.

DV	Interaction/factors	<i>F</i>	d.o.f.	<i>p</i>	η^2
Manipulation check	Group $\times$ strategy $\times$ time	14.88	186	<0.001	0.11
	Group	...	...	...	...
	Strategy	...	...	...	...
	Time	...	...	...	...
Usefulness	Group $\times$ strategy $\times$ time	0.72	186	0.54	0.01
	Group	2.58	62	0.11	0.04
	Strategy	0.54	186	0.65	0.01
	Time	2.18	62	0.15	0.03
Perceived competence	Group $\times$ strategy $\times$ time	4.08	162.56	0.01	0.03
	Group	...	...	...	...
	Strategy	...	...	...	...
	Time	...	...	...	...

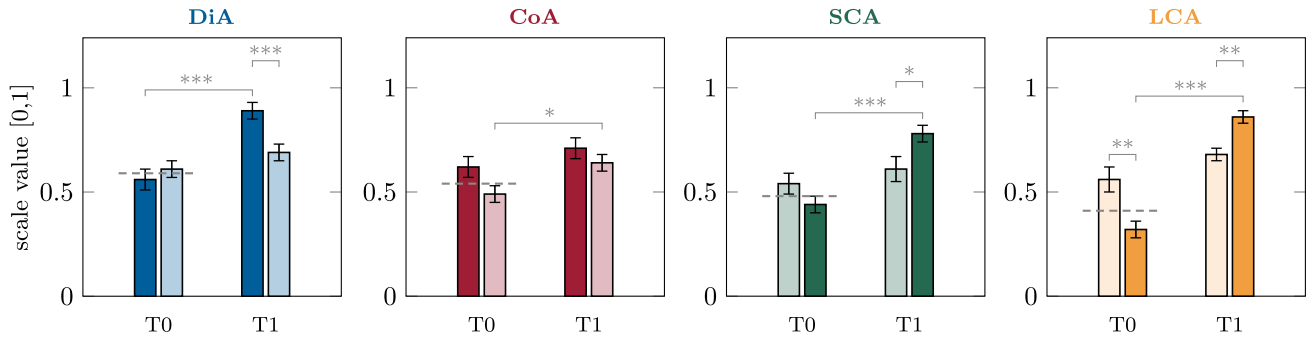


FIG. 4. Students' indicated perceived integration of the four strategies (mean and standard error, mean for T0 in gray) in previous (and for T1 also current) courses for the two intervention groups for T0 and T1. Dark color: group that received the strategy-related intervention, light color: nonreceiving group.

A. Perceived integration in previous and current courses (manipulation check)

At the beginning of the semester (T0), students stated that the overall integration of all strategies in previous courses was moderate (0.50 ± 0.19). An individual examination of each strategy reveals that LCA was the least prevalent (see Fig. 4). An analysis of variance with the factors group and strategy showed a small-sized interaction effect [$F(3, 189) = 3.79, p = 0.01, \eta^2 = 0.05$]. Students in the SCALCA group had the impression that the analysis of limiting cases received less attention in previous courses in comparison to the DiACoA group ($p = 0.002, d = 0.80$).

After the intervention (T1), the perception of the overall integration of all strategies in previous and current courses increased with a large-sized effect [$t(63) = 8.51, p < 0.001, d = 1.03$]. A medium-sized interaction effect was observed between the factors group, strategy, and time point. *Post hoc* analyses revealed a significant temporal difference for the DiACoA group for dimensional analysis ($p < 0.001, d = 1.54$) and for the SCALCA group for special and limiting case, as well as for covariational analysis (SCA: $p < 0.001, d = 1.26$; LCA: $p < 0.001, d = 2.23$; CoA: $p = 0.01, d = 0.56$). For time point T1, the *post*

hoc tests showed significant differences between the two groups for all strategies except covariational analysis (DiA: $p < 0.001, d = 0.94$; SCA: $p = 0.02, d = 0.61$; LCA: $p = 0.01, d = 0.68$).

B. Usefulness (RQ1)

All students perceived all four strategies across both time points as highly useful (0.82 ± 0.12 , i.e., close to the maximum value of 1). The results of the core sample and the four strategies for T0 and T1 are shown in Fig. 9 in the Appendix.

An analysis of variance before the intervention (T0) with the factors group and strategy showed no significant interaction, as well as no main group related between or main strategy related within difference [interaction: $F(2.64, 167.32) = 1.07, p = 0.36, \eta^2 = 0.02$; between: $F(1, 63) = 2.33, p = 0.13, \eta^2 = 0.03$; within: $F(2.64, 167.32) = 1.98, p = 0.13, \eta^2 = 0.03$].

The postsurvey (T1) revealed no significant change in the usefulness scores. There was neither a significant interaction effect between the factors time point, strategy, and group nor main effects regarding temporal within, strategy related within, or group related between differences.

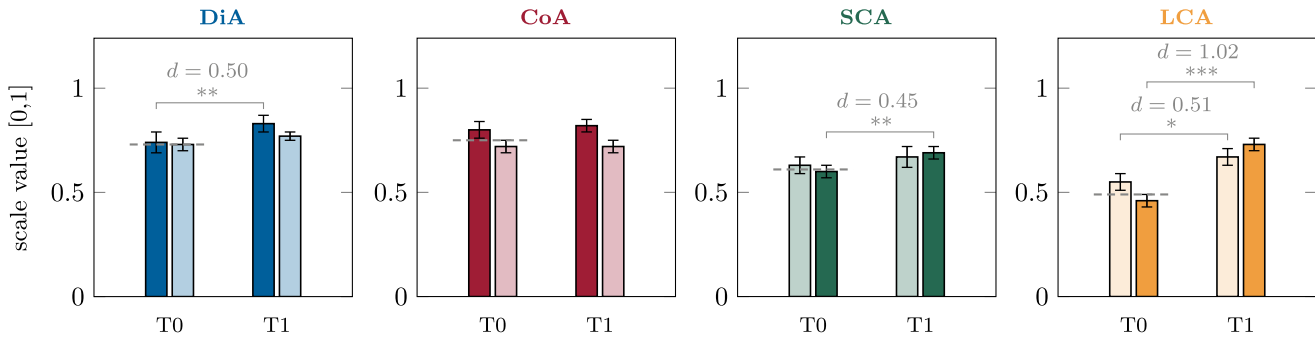


FIG. 5. Students' indicated perceived competence regarding the four strategies (mean and standard error, mean for T0 in gray) for the two intervention groups for T0 and T1 (effect size: Cohens' d). Dark color: group that received the strategy-related intervention, light color: nonreceiving group.

TABLE IV. Results of analyses of covariance with post-test performance scores as the dependent variable (DV) and prescores as a covariate, including test statistics and effect sizes (degrees of freedom d.o.f.=64). Bold values indicate statistical significance.

DV	Factor/covariate	F	p	η^2	d
Overall	Group	0.11	0.75	0.001	...
	Pre	9.58	0.003	0.13	...
DiA	Group	7.13	0.01	0.10	0.67
	Pre	3.57	0.06	0.05	...
CoA	Group	2.16	0.15	0.03	...
	Pre	4.49 ^a	0.04	0.06	...
SCA	Group	4.49	0.04	0.06	0.50
	Pre	4.49	0.04	0.06	...
LCA	Group	0.08	0.78	0.001	...
	Pre	0.34	0.56	0.01	...

^aDue to rounding, results for CoA and SCA appear similar.

C. Perceived competence (H1)

At the beginning of the intervention (T0), students indicated a perceived deficiency in their competence to analyze limiting cases in comparison to all other strategies (see Fig. 5, and for a detailed impression of T0, see Fig. 10 in the Appendix). When comparing the two intervention groups at T0, no interaction effect was found between the factors group and strategy as well as no main effect regarding the factor group [interaction: $F(3, 189) = 0.91, p = 0.44, \eta^2 = 0.01$; between: $F(1, 63) = 1.49, p = 0.23, \eta^2 = 0.02$]. Only a large-sized main effect regarding the factor strategy was found [$F(3, 189) = 35.41, p < 0.001, \eta^2 = 0.35$]. *Post hoc* tests show significant differences with medium- and large-sized effects between DiA and SCA ($p < 0.001, d = 0.64$), DiA and LCA ($p < 0.001, d = 1.04$), SCA and LCA ($p < 0.001, d = 0.53$), SCA and CoA ($p < 0.001, d = 0.67$), and LCA and CoA ($p < 0.001, d = 1.06$).

The overall perceived competence at the beginning of the semester (T0) regarding all four strategies was moderate and increased with a medium-sized effect to a significantly higher level after the intervention [T0: 0.64 ± 0.16 ; T1: 0.73 ± 0.14 ; $t(63) = -4.60; p < 0.001, d = 0.55$]. When looking at each strategy individually, a small-sized interaction effect was observed for the factors group, strategy, and time point. *Post hoc* tests revealed that the perceived competence in applying CoA after the intervention (T1) differed between the two groups ($p = 0.03, d = 0.59$). In addition, a temporal difference was observed for dimensional and limiting case analysis for DiACoA, each with a medium-sized effect (DiA: $p = 0.01, d = 0.50$; LCA: $p = 0.03, d = 0.51$). A difference between both time points was determined for SCALCA regarding special and limiting case analysis with a small and large effect size (SCA: $p = 0.01, d = 0.45$; LCA: $p < 0.001, d = 1.02$).

D. Performance (H2)

The performance analyses yielded results analogous to those obtained in the perceived competence analyses (see Fig. 6). The analysis of limiting cases was the most difficult in comparison to the other strategies. This finding is supported by a large-sized main effect for the factor strategy [$F(2.67, 173.55) = 35.12, p < 0.001, \eta^2 = 0.34$]. Significant differences were apparent between DiA and SCA ($p = 0.002, d = 0.46$), DiA and LCA ($p < 0.001, d = 1.14$), LCA and CoA ($p < 0.001, d = 0.95$), and LCA and SCA ($p < 0.001, d = 0.71$). Neither an interaction between the factors group and strategy nor a group effect was observed [interaction: $F(2.67, 173.55) = 1.10, p = 0.35, \eta^2 = 0.01$; between: $F(1, 65) = 1.16, p = 0.29, \eta^2 = 0.01$].

To determine the impact of the intervention, the post-test performance scores were adjusted by controlling for the prescore as a covariate. The results of the analyses of covariance for each strategy are presented in Table IV and Fig. 6, accompanied by the indication of adjusted means.

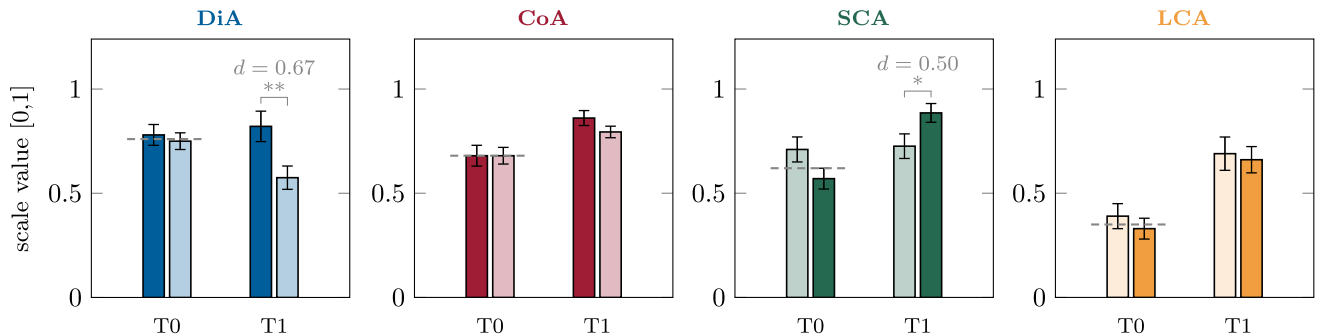


FIG. 6. Students' performance regarding the four strategies in comparison of the two intervention groups with adjusted means for T1 (effect size: Cohen's d). Note that the pretest (T0) covered mostly mechanics and the post-test (T1) advanced electrostatics topics. Dark color: group that received the strategy-related intervention, light color: nonreceiving group.

A medium-sized group effect has been observed for dimensional and special case analysis [DiA: $F(1, 64) = 7.13$, $p = 0.01$, $d = 0.67$; SCA: $F(1, 64) = 4.49$, $p = 0.04$, $d = 0.50$]. For covariational and special case analysis, the prescore has been demonstrated to contribute to a significant reduction in error variance (for both medium-sized effects: $\eta^2 = 0.06$).

V. DISCUSSION

The implementation of the intervention can be considered successful. The manipulation check, which assessed the perceived integration of the strategies into past and current courses, thereby validated the intervention for three strategies. It confirmed for DiA, SCA, and LCA that the implemented strategies were indeed perceived as being specifically integrated into the course. This suggests that the intervention effectively emphasized these strategies within the exercises and recitation sessions. Notably, CoA did not show a comparable pattern in the manipulation check. In contrast, students from the SCALCA group who did not perform CoA perceived the integration into the course as significantly higher after the intervention. The following sections discuss possible explanations in accordance with the research questions and hypotheses, taking target variables into account.

A. Usefulness (RQ1)

Students perceived all four strategies as highly relevant at the beginning of the course and after the intervention. No particular strategy was identified as being more beneficial than another. This is consistent with common science literature and experts who attribute high relevance to all of these strategies [1,47].

The consistently high ratings across the intervention (T0 and T1) suggest that students are already aware of the usefulness of the four strategies at an early stage of their studies. This indicates a general appreciation for sense-making approaches, and students recognize their value for understanding equations beyond thinking of them as mere tools. Previous studies report an overall positive feedback from students on the use of such strategies in instruction, and some students seem to gain further insights into physics concepts [9,10,16].

B. Perceived competence (H1)

An initial indication of the intervention's effectiveness is reflected in an increase in students' perceived competence over time. Specifically, students reported greater confidence in applying DiA, SCA, and LCA to assess the plausibility of equations. In contrast, no significant change in perceived competence was observed for CoA. These findings are discussed in the following paragraphs.

Subsequent to the intervention (T1), students in the respective group reported an increased perception of

competence over time in DiA, SCA, and LCA, supporting the instructional approach. Interestingly, students in the DiACoA group, who had not explicitly practiced LCA, also reported an increase in their perceived competence regarding LCA. This unexpected enhancement in both groups may be caused by potential intergroup contamination due to common challenges in field studies embedded in university courses. Despite efforts to maintain distinct intervention conditions, exchanges among students or overlapping emphases by tutors may have blurred the boundaries between the groups. Tutors may have unintentionally drawn attention to limiting behaviors during recitation sessions regardless of the respective intervention group. Additionally, limiting cases of electrical and magnetic phenomena were frequently and naturally embedded in the lecture material, which may have reinforced familiarity with LCA independent of the intervention group.

In contrast, the perceived competence regarding CoA did not show a significant increase over time, except for a minor (and nonsignificant) increase in the DiACoA group, where CoA was practiced. Several factors may explain this outcome. First, CoA was initially rated highest in perceived competence among all four strategies, which may have limited the potential for further improvement. Second, inconsistencies in terminology across materials may have introduced confusion. While the strategies DiA, SCA, and LCA were explicitly named in questionnaires as well as intervention and test tasks, intervention tasks on CoA referred to "functional dependencies" (see Figs. 2, 7, and 8). Covariational reasoning was not explicitly labeled within the intervention tasks as "the analysis of covariance." However, the questionnaire used this specific term, which students may not have associated with the intervention tasks, potentially leading to inaccurate self-assessments, even though a comprehensive example and explanation was always given in advance. This explanation is supported by manipulation check results, indicating that the DiACoA group did not perceive a significant increase in the integration of CoA within the course.

At the beginning of the semester (T0), differences across the strategies became apparent when examining students' perceived competence. Perceived confidence was lowest for LCA and highest for DiA. These findings align with prior research suggesting that DiA is more applicable and intuitive for students and therefore is mastered and acquired more rapidly in comparison to SCA and LCA. [16,33]. This is further supported by Hahn *et al.* [2], who found that students exhibit a tendency to gravitate toward DiA when no specific strategy is required, indicating its perceived ease of use.

C. Performance (H2)

The analyses of the performance data revealed strategy-specific effects of the intervention. Students in the respective intervention groups outperformed their peers in applying

DiA and SCA, indicating a positive effect of the intervention approach. For CoA, a nonsignificant trend favoring the DiACoA group was observed, while no group-related difference in performance was apparent for LCA. It is important to note that, due to different pre- and post-tests, a time-based comparison of performance results is not reasonable. The results for DiA and SCA are consistent with students' self-reported perceptions, as reflected in the manipulation check, where an increased integration of these strategies into the course was noted. These findings also align with prior work by Warren [16], which suggests that explicit instruction on DiA and SCA seems to improve students' use of those strategies and significantly enhance multiple-choice exam performance. Although previous studies have suggested that LCA can also benefit from explicit instruction [9,16], no significant performance difference was found for LCA in the present study. This lack of effect is surprising, especially given that students reported greater familiarity with LCA in the manipulation check. Although DiACoA showed a slight tendency toward an enhanced performance of CoA, the SCALCA group unexpectedly reported an increased integration of this strategy into the course, despite not having practiced this strategy explicitly. The following paragraphs will discuss possible explanations for the aforementioned findings.

An analysis of students' performance at T0 provides important insights into their prior proficiency before the intervention began. At the beginning of their second semester, students demonstrated the highest performance in applying DiA and the lowest in applying LCA. These results mirror the perceived competence findings and indicate that LCA posed the greatest challenge among the four strategies even before the intervention. Prior research supports this observation, identifying DiA as the most applicable strategy [16,33]. One reason for the high performance of DiA may be the relatively low cognitive demands associated with the strategy. The analysis of dimensions requires an algorithmic approach of substituting dimensions and performing arithmetic operations without requiring a deeper conceptual understanding. In contrast, the lower performance for CoA, SCA, and especially LCA is likely to reflect the complexity of these approaches. In particular, the construction of mental models, or thought experiments, in which variables must be systematically manipulated [9,27]. The outcomes then must be compared to physical intuition or prior knowledge. Burkholder *et al.* [8] observed as well that novice students engage in more algorithmic procedures (e.g., rederiving solutions, "plug-and-chug," dimensional analysis) while advanced students tend to apply more sophisticated strategies and build connections to their conceptual knowledge (e.g., SCA and LCA).

The persistent low performance, in combination with the low perceived competence regarding LCA (even

postintervention), may be attributed to limited prior exposure. Compared to the other strategies, LCA appears to be underrepresented in university physics courses, which could explain why students struggle with its application. Warren also observed that, in contrast to DiA, students require additional practice in applying LCA [16].

An intervention effect with regard to the performance of LCA could not be identified. Both groups showed a comparable level of performance at T1. This finding is unexpected in light of prior research, which indicates that the ability to perform LCA can be enhanced through recitation sessions and practice in homework assignments [9,16,33]. Given that LCA-related intervention and test tasks were formulated clearly and in the same manner, especially in relation to SCA, it is unlikely that the design of the materials contributed to the absence of a measurable intervention effect. A more plausible explanation may lie in the practical limitations of the design of the field study. As previously discussed, a complete separation of students from different intervention groups could not be ensured. Discussions in collaborative work on homework assignments across intervention groups or unintentional emphasis of limiting behaviors in recitation sessions could have led to "informal" learning. Such intergroup communication may have led to unintended exposure to LCA, thus contributing to a similar performance level after the intervention. This interpretation is further supported by the increase in perceived competence in LCA observed across both intervention groups (for DiACoA with a medium effect size and for SCALCA with a large effect size).

Group-specific performance differences after the intervention (T1) could not be found for CoA, with only a slight tendency favoring the respective intervention group. One possible explanation lies in the conceptual overlap of CoA and LCA. This became evident during the analysis of the pre- and post-test data. Students occasionally engaged in LCA, while CoA was required. Although we treated these as two distinct strategies, both involve reasoning about the relationship between two variables when one of them is changing. CoA shares some key features with the dynamic aspect of the concept of limits as discussed in particular in mathematics education research [48]. The dynamic aspect of limits refers to the infinite mental process of gradually approaching a limit. This is similar to the reasoning involved in CoA, where two variables have to be coordinated while one is changing [49]. Moreover, prior research by Zietsman and Clement [26] suggests that reasoning in limiting cases can strengthen students' understanding of the ordinal relationships between two variables. This conceptual overlap may have enabled students in both intervention groups to benefit from the implementation of each strategy. As a result, both groups may have improved their performance in CoA by the end of the intervention (T1) leading to no group-specific difference. This is further supported by

the finding that students in the SCALCA group reported a stronger perceived integration of CoA in the course, despite not explicitly performing it. A reciprocal influence, where CoA training in the DiACoA group enhanced performance and perceived competence of LCA, is likewise conceivable.

The SCALCA group showed a notably low performance in DiA postintervention, not only compared to the DiACoA group but also relative to its own prescore at T0. Even though different pre- and post-tests were used, the low performance level stands out. This decline appears counter-intuitive, given the generally positive trends over time for the other strategies. However, this becomes plausible when considering the different demands that each strategy entails. While all four strategies require the handling of equations in relation to their physical meaning, DiA relies substantially on declarative knowledge, particularly the recall of correct dimensions and units [50]. The SCALCA group did not receive exercises on DiA and thus had limited opportunities to apply it in the context of the course topics, which may have contributed to their “decrease” in performance. In contrast, CoA, SCA, and LCA require particularly procedural or strategic knowledge. As Nersessian argues, strategies like LCA reflect “nonalgorithmic” approaches that can be mastered in different ways and adapted to different scenarios [27]. Warren also highlights that such approaches are considered more effective for gaining deeper physical insights in comparison to the analysis of dimensions [16].

D. Constraints and limitations

One important limitation concerns the design of the used tasks. The strategies were practiced primarily using relatively simple equations and idealized scenarios, especially in test tasks. Moreover, students were only asked to apply strategies that were explicitly prescribed in the intervention. They were not encouraged to independently select or even combine strategies based on the given context. While this ensured accessibility and comparability in the intervention design, it does not reflect the complexity of physics problems encountered in later stages of physics studies or expertlike sensemaking in professional practice.

Additionally, the four examined strategies represent only a selection of the various approaches to sensemaking in physics problem solving, e.g., order-of-magnitude reasoning. This strategy was not included because our research focused on those approaches that are explicitly and directly tied to the formal structure of equations (e.g., DiA, CoA, SCA, and LCA). By contrast, order-of-magnitude reasoning engages with equations in a way that is simultaneously heuristic and critical. It can serve as a tool to generate estimates, but also to evaluate whether results obtained from equations make physical sense [51,52]. This implies increased practical demands for instruction. Unlike the other four strategies, order-of-magnitude reasoning

presupposes a wide range of declarative knowledge about typical values and physical scales. Moreover, suitable exercises are highly context-specific, which makes it difficult to establish them as a generalizable, domain-independent strategy. Thus, within the given practical constraints (limited teaching and learning time, increased instructional demands), we did not include order-of-magnitude reasoning, although it represents a relevant sense-making approach for evaluating arithmetical results [52]. Given its importance in both historical and educational contexts, we acknowledge that its omission represents a limitation of the present work and an interesting direction for future research.

The study was conducted in a field setting, which naturally introduces further limitations, as previously mentioned. A key constraint is the absence of a traditional control group that received no intervention. Instead, both groups were exposed to different strategy-based interventions. An alternative study design in which all students received instruction in the same set of strategies would have increased the overall sample size and reduced potential contamination. However, such designs would have limited the ability to examine strategy-specific effects on the target variables. These aspects should be taken into account when interpreting the reported results.

Finally, the unequal group sizes and the overall small sample size pose additional challenges to interpretation and limit the generalizability of the findings. Moreover, the cohort sizes in our undergraduate physics program are inherently limited, and the participating students in the present study are representative of the overall first-year cohort at our institution.

Despite these constraints and limitations, the findings of this study largely align with the hypothesized effects.

VI. CONCLUSION AND OUTLOOK

Given the central role of mathematics as the language of physics, this study aimed to integrate and evaluate physics-specific sensemaking strategies within undergraduate physics degree programs. In an introductory physics course, the analysis of dimensions, covariations, special and limiting cases was embedded in recitation sessions and homework assignments. These strategies are essential tools for meaningfully connecting mathematical expressions with physical concepts, fostering a deeper comprehension of physics equations.

The results emphasize the importance of explicitly teaching such strategies. Students recognized the relevance and usefulness of all four strategies and reported increased confidence in applying them. These gains in perceived competence were accompanied by improvements in performance, particularly for DiA and SCA. This indicates that the intervention approach can effectively promote sense-making strategies in physics.

Despite improved perceived competence, no significant performance gain was observed for LCA, suggesting that its high cognitive demands may require extended or more intensive practice. Future research should use a dose-response framework to investigate how the quantity of practice affects learning outcomes. Specifically, the question remains unanswered whether there is a threshold of exposure required for students to apply LCA effectively.

Another concern is the quality of students' written justifications, which were frequently superficial or minimal. This suggests the need for a more nuanced task design in future studies. For example, tasks should be structured to explicitly separate the application of strategies from the reasoning process by including subprompts that ask students to assess the physical plausibility based on the outcome.

While the present study focused on prompted application of prescribed strategies, future research should investigate students' ability to select appropriate strategies. Fostering the selection process represents a critical step toward developing expertlike sensemaking in physics.

Overall, the intervention positively impacted students' perceived competence. This is a favorable result, as an accurate self-assessment is associated with more goal-directed learning behaviors and academic success [53]. Therefore, enhancing students' confidence in their abilities may contribute to greater persistence and reduced dropout rates in physics degree programs.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of all students that participated and supported this study. Our gratitude is

extended to the student assistants who contributed to the data curation and interrater processes (Laura Pflügl, Stine Gerlach, Deniz Can Senel). Furthermore, we gratefully acknowledge the generous support of the *Wilfried and Ingrid Kuhn Foundation*, which made this research possible. Additional support was provided by the *Hochschulpakt Lehre*. Publication of this article was partly funded by the Open Access Publication Funds of the Göttingen University.

J. H.: conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft, writing—review and editing, visualization. P. K.: conceptualization, methodology, investigation, writing—review and editing, supervision, project administration, funding acquisition. A. M.: conceptualization, methodology, writing—review and editing, supervision. J. N.: conceptualization, methodology, investigation, writing—review and editing.

Participation in the study was voluntary and all data were obtained anonymously. In order to match data from different time points, the use of pseudonyms was employed to ensure that no inferences about the individual could be made.

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.

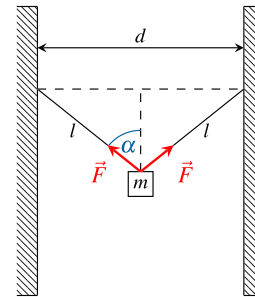
APPENDIX

Task *Street light*

A street light with mass m is suspended from two identical cables between two house walls at a distance of d . The cables from which the streetlight is suspended are both of length l and are attached to the walls at the same height. The angle α assumes values in the range $0^\circ \leq \alpha < 90^\circ$. Two equivalent equations that describe the force on one cable are as follows:

$$|\vec{F}| = \frac{mg}{2\cos\alpha},$$

$$|\vec{F}| = \frac{mgl}{2\sqrt{l^2 - \left(\frac{d}{2}\right)^2}}.$$



In the following, the physical plausibility of the equation is to be checked without deriving the equation independently. Choose **one equation** and proceed as follows for the check:

- Check the dimensions/units of the equation.
- Examine a relevant special case of the equation.
- Examine a relevant limiting case of the equation.
- In winter, both cables shorten by the same amount due to the cold. Investigate how the shortening affects a variable of your choice.
- Consider the following dependencies by marking the statements as true or false:

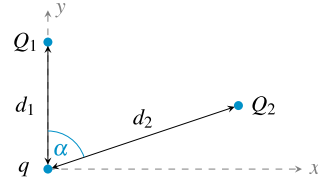
	true	false
The greater the distance between the walls, the lower the force on the cable.	☐	☐
The smaller the angle, the lower the force on the cable.	☐	☐
To keep the force on the cable constant, the mass must be reduced if the angle is increased.	☐	☐
To keep the force on the cable constant, the distance must be increased for longer cables.	☐	☐
The longer the cables, the smaller the angle (other quantities constant).	☐	☐

FIG. 7. Example task on force on cables of a hanging street light for the pretest (T0).

Task *Discrete Charge Distribution*

A test charge q is located in the electric field of two charges Q_1 and Q_2 (all charges are point charges). The distance between charge q and Q_1 is denoted by d_1 and between q and Q_2 by d_2 . The charge Q_1 is located on the y -axis and Q_2 is variable in the plane (see figure). The three charges Q_2 , q and Q_1 include the angle α . The electric field constant is $\epsilon_0 \approx 9 \cdot 10^{-12} \frac{\text{A}^2 \text{s}^4}{\text{kgm}^3}$. The force $\vec{F}_q$ acting on the test charge q can be described by the following equation:

$$\vec{F}_q = \frac{q}{4\pi\epsilon_0} \left(\frac{Q_2 \sin(\alpha)}{d_2^2} \vec{e}_\perp + \frac{Q_2 \cos(\alpha)}{d_2^2} + \frac{Q_1}{d_1^2} \right)$$



In the following, the physical plausibility of the equation is to be checked without deriving the equation independently. Proceed as follows for the check:

- Check the dimensions/units of the equation. Based on your analysis, is the equation physically plausible?
- Examine a relevant special case of the equation. Based on your analysis, is the equation physically plausible?
- Examine a relevant limiting case of the equation. Based on your analysis, is the equation physically plausible?
- Based on the equation, investigate how a change in a variable affects the force. Based on your analysis, is the equation physically plausible?
- Examine the following dependencies of the equation. Variables not explicitly mentioned are assumed to be constant. Indicate whether the statements are correct or incorrect:

	true	false
The smaller the charge Q_2 , the smaller the force.	☐	☐
The greater the distance d_1 , the greater the force.	☐	☐
The greater the angle, the greater the force.	☐	☐
To keep the force constant, all distances must be doubled when the test charge is quadrupled.	☐	☐
To keep the force constant, the charge Q_2 must be doubled if the distance d_2 is halved.	☐	☐

FIG. 8. Example task on a discrete charge distribution for the post-test (T1).

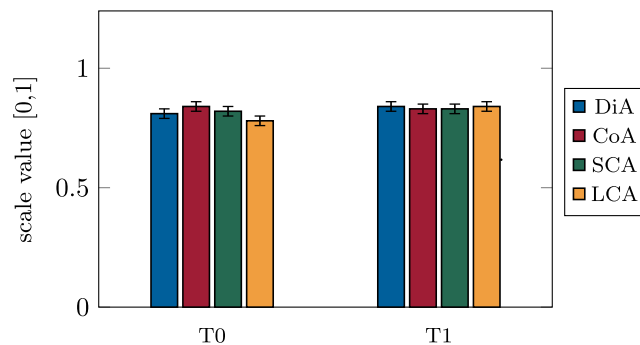


FIG. 9. Students' indicated usefulness of the four strategies (mean and standard error) for T0 and T1 for the core sample.

TABLE V. Translated items of the questionnaire with characteristic values (item-total correlation r_{it} , internal consistency coefficient Cronbach's α). As a measure of internal consistency, Cronbach's α was calculated and interpreted in accordance with the guidelines of Cronbach [54]. Item analyses were carried out with the initial sample of $N_0 = 116$ students at T0. The scales on perceived competence and usefulness are based on the Intrinsic Motivation Inventory [40].

Construct	Item	$r_{it,DiA}$	$r_{it,CoA}$	$r_{it,SCA}$	$r_{it,LCA}$
Perceived competence	I believe that I am not making any mistakes when applying the strategy.	0.82	0.86	0.83	0.87
	I am good at applying the strategy.	0.90	0.88	0.90	0.85
	I feel confident using the strategy.	0.89	0.91	0.90	0.90
	I know what to do when applying the strategy.	0.82	0.79	0.76	0.78
	Cronbach's α	0.89	0.92	0.90	0.90
Usefulness	I believe that mastering the strategy is important.	0.82	0.83	0.83	0.77
	I believe that using the strategy has a practical benefit for me.	0.80	0.80	0.77	0.88
	I believe that mastering the strategy brings a deeper physical insight.	0.76	0.84	0.55	0.81
	I will continue to use the strategy in the future because it has a practical benefit for me.	0.78	0.77	0.71	0.78
	Cronbach's α	0.80	0.86	0.72	0.86
Manipulation check	The strategy was explicitly subject of previous courses. ^a	0.83	0.91	0.90	0.90
	I was able to practice applying the strategy in previous courses. ^a	0.77	0.85	0.73	0.81
	The strategy was explicitly included in exercise sheets from previous courses. ^a	0.75	0.94	0.93	0.92
	Cronbach's α	0.72	0.92	0.85	0.90

^aWith the addition in the post-test (T1): “including this course.”

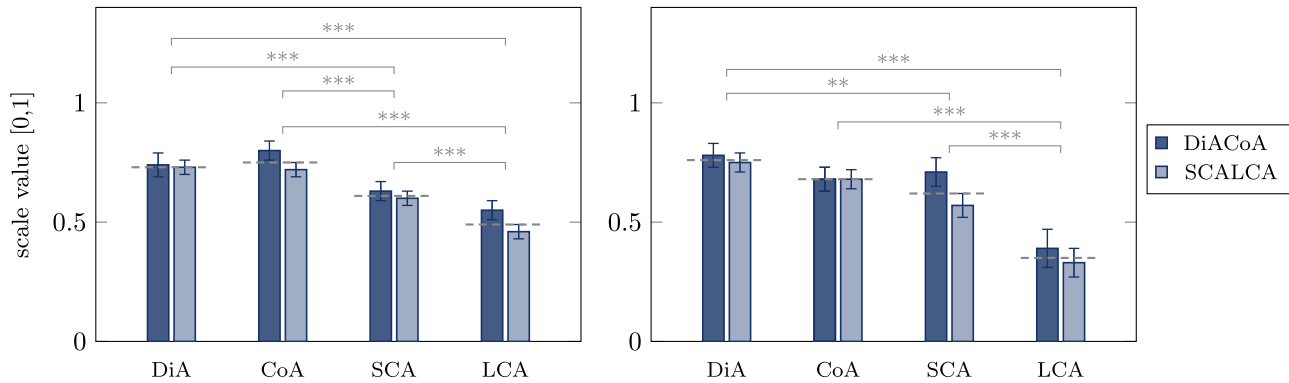


FIG. 10. Students' perceived competence (left) and performance (right) at T0 of the four strategies (mean and standard error, mean for each strategy in gray) in comparison of the two intervention groups.

TABLE VI. Coding categories for evaluating the pre- and post-test for the application of the analysis of dimensions (DiA), covariations (CoA), special (SCA), and limiting cases (LCA), based on the rubrics of the scientific abilities of Etkina *et al.* [32].

Category	Description	0	1	2	3
DiA	The person can analyze the dimensions/units of an equation to examine the plausibility.	Empty/No attempt was made to analyze the dimensions/units of an equation.	An attempt was made to identify the dimensions of variables involved, but the equation was not analyzed.	An analysis of the dimensions was carried out, but it is flawed. ^a	A dimension analysis was carried out without errors.
CoA	The person can analyze a relevant dependency of two variables of an equation to examine the plausibility.	Empty/No attempt was made to analyze a relevant dependency of two variables of an equation.	An analysis was attempted, but the selected dependency is not relevant or there are large gaps/many errors in the analysis.	An analysis of a dependency of two or more variables was performed, but it is flawed. ^a	A covariational analysis was carried out without errors.
SCA	The person can analyze a relevant special case of an equation to analyze the plausibility.	Empty/No attempt was made to analyze a relevant special case of an equation.	An analysis was attempted, but the selected case is not relevant or there are large gaps/many errors in the analysis.	An analysis of a relevant special case was performed, but it is flawed. ^a	A special case analysis was carried out without errors.
LCA	The person can analyze a relevant limiting case of an equation to analyze the plausibility.	Empty/No attempt was made to analyze a relevant limiting case of an equation.	An analysis was attempted, but the selected case is not relevant or there are large gaps/many errors in the analysis.	An analysis of a relevant limiting case was performed, but it is flawed. ^a	A limiting case analysis was carried out without errors.

^aFlaws: algebraically, mathematically, physically and notation.TABLE VII. Cohen's weighted kappa κ_w [46] as an indicator of the reliability of the rating according to the rubrics of the scientific abilities of Etkina *et al.* [32].

Strategy	First rating κ_w	Second rating κ_w
DiA	0.76	0.84
CoA	0.68	0.86
SCA	0.78	1.00
LCA	0.93	1.00
Total	0.81	0.95

TABLE VIII. Students' performance, perceived competence, usefulness, and manipulation check with characteristic values: mean M , standard deviation SD, adjusted mean M_{adj} (adjusted for prescore) at time points T0 and T1.

	T0				T1					
	DiACoA		SCALCA		DiACoA			SCALCA		
	$N = 25$		$N = 42$		$N = 25$			$N = 42$		
	M	SD	M	SD	M	SD	M_{adj}	M	SD	M_{adj}
Performance										
DiA	0.78	0.23	0.75	0.25	0.83	0.29	0.82	0.57	0.41	0.57
CoA	0.68	0.26	0.68	0.27	0.86	0.13	0.86	0.79	0.21	0.79
SCA	0.71	0.28	0.57	0.34	0.75	0.35	0.73	0.87	0.26	0.89
LCA	0.39	0.28	0.33	0.31	0.69	0.40	0.69	0.66	0.41	0.66
Perceived competence										
DiA	0.74	0.25	0.73	0.18	0.83	0.18	...	0.77	0.15	...
CoA	0.80	0.21	0.72	0.21	0.82	0.15	...	0.72	0.19	...
SCA	0.63	0.20	0.60	0.20	0.67	0.24	...	0.69	0.20	...
LCA	0.55	0.20	0.46	0.22	0.67	0.20	...	0.73	0.20	...
Usefulness										
DiA	0.82	0.15	0.80	0.20	0.86	0.13	...	0.82	0.18	...
CoA	0.89	0.12	0.80	0.18	0.86	0.15	...	0.82	0.20	...
SCA	0.83	0.15	0.80	0.15	0.86	0.16	...	0.82	0.19	...
LCA	0.83	0.20	0.75	0.18	0.87	0.14	...	0.83	0.17	...
Manipulation check										
DiA	0.56	0.26	0.61	0.26	0.89	0.17	...	0.69	0.25	...
CoA	0.62	0.25	0.49	0.28	0.71	0.26	...	0.64	0.28	...
SCA	0.54	0.27	0.44	0.26	0.61	0.29	...	0.78	0.27	...
LCA	0.56	0.31	0.32	0.26	0.68	0.29	...	0.86	0.22	...

- [1] D. J. Griffiths, *Introduction to Electrodynamics*, 4th ed. (Pearson, Boston, 2013).
- [2] K. T. Hahn, P. J. Emigh, M. Lenz, and E. Gire, Student sense-making on homework in a sophomore mechanics course, *presented at PER Conf. 2018, Cincinnati, OH*, pp. 160–163, [10.1119/perc.2017.pr.035](#).
- [3] M. Lenz, Sensemaking throughout the physics curriculum: Understanding expert and student ideas about sensemaking in a physics context, Ph.D. thesis, Oregon State University, 2020.
- [4] E. F. Redish and E. Kuo, Language of physics, language of math: Disciplinary culture and dynamic epistemology, *Sci. Educ.* **24**, 561 (2015).
- [5] B. R. Wilcox, M. D. Caballero, R. E. Pepper, and S. J. Pollock, Upper-division student understanding of Coulomb's law: Difficulties with continuous charge distributions, *AIP Conf. Proc.* **1513**, 418 (2013).
- [6] O. Uhden, R. Karam, M. Pietrocola, and G. Pospiech, Modelling mathematical reasoning in physics education, *Sci. Educ.* **21**, 485 (2012).
- [7] 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).
- [8] E. W. Burkholder, O. C. Miller, and L. F. Blackmon, Investigating the epistemology of physics students while reflecting on solutions, *Eur. J. Phys.* **44**, 045701 (2023).
- [9] G. White, T.-R. Sikorski, J. Landay, and M. Ahmed, Limiting case analysis in an electricity and magnetism course, *Phys. Rev. Phys. Educ. Res.* **19**, 010125 (2023).
- [10] M. Lenz, K. Hahn, P. Emigh, and E. Gire, Student perspective of and experience with sense-making: a case study, *presented at PER Conf. 2017, Cincinnati, OH*, pp. 240–243, [10.1119/perc.2017.pr.055](#).
- [11] 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).
- [12] T. O. B. Odden and R. S. Russ, Defining sensemaking: Bringing clarity to a fragmented theoretical construct, *Sci. Educ.* **103**, 187 (2019).

- [13] A. Karelina and E. Etkina, Acting like a physicist: Student approach study to experimental design, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020106 (2007).
- [14] M. E. Loverude, Mathematization and the “boas course”, presented at *PER Conf. 2017, Cincinnati, OH*, pp. 15–18, 10.1119/perc.2017.juried.004.
- [15] A. R. Akinyemi, An investigation of students’ use and understanding of evaluation strategies, Ph.D. thesis, University of Maine, 2021.
- [16] A. R. Warren, Impact of teaching students to use evaluation strategies, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020103 (2010).
- [17] M. E. Loverude, Quantitative reasoning skills in math methods, presented at *PER Conf. College Park, MD*, pp. 203–206, 10.1119/perc.2015.pr.046.
- [18] J. J. Clement, *Creative Model Construction in Scientists and Students: The Role of Imagery, Analogy, and Mental Simulation*, 1st ed. (Springer, Dordrecht, 2009).
- [19] E. F. Redish, Using math in physics: 2. Estimation, *Phys. Teach.* **59**, 525 (2021).
- [20] B. Sherin, The symbolic basis of physical intuition : A study of two symbol systems in physics instruction, Ph.D. thesis, University of California, 1996.
- [21] K. Plicht, *Ein Physikübungs-konzept Zur Förderung Der Problemlösekompetenz: Entwicklung Und Empirische Evaluation Eines Strategietrainings Auf Der Basis Von Expertisemerkmale* (Logos-Verlag Berlin, 2024).
- [22] M. Carlson, S. Jacobs, E. Coe, S. Larsen, and E. Hsu, Applying covariational reasoning while modeling dynamic events: A framework and a study, *J. Res. Math. Educ.* **33**, 352 (2002).
- [23] L. Saldanha and P. Thompson, Re-thinking co-variation from a quantitative perspective: Simultaneous continuous variation, in *Proceedings of the Annual Meeting of the Psychology of Mathematics Education—North America* (1998), pp. 298–304, https://www.academia.edu/110758966/Proceedings_of_the_Annual_Meeting_of_the_North_American_Chapter_of_the_International_Group_for_the_Psychology_of_Mathematics_Education_20th_Raleigh_NC_October_31_November_3_1998_Volume_1.
- [24] E. F. Redish, Using math in physics: Overview, *Phys. Teach.* **59**, 314 (2021).
- [25] S. White Brahmia, A. Olsho, T. I. Smith, A. Boudreaux, P. Eaton, and C. Zimmerman, Physics inventory of quantitative literacy: A tool for assessing mathematical reasoning in introductory physics, *Phys. Rev. Phys. Educ. Res.* **17**, 020129 (2021).
- [26] A. Zietsman and J. Clement, The role of extreme case reasoning in instruction for conceptual change, *J. Learn. Sci.* **6**, 61 (1997).
- [27] N. Nersessian, How do scientists think? Capturing the dynamics of conceptual change in science, in *Cognitive Models of Science* (University of Minnesota Press, Minneapolis, 1992), pp. 3–45.
- [28] M. Lenz, P. J. 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.
- [29] R. Robinett, Quantum fields from dimensional analysis: Applications in condensed matter systems, *Am. J. Phys.* **93**, 344 (2025).
- [30] K. Gifford, G. S. Ehrlich, E. Bumbacher, and E. Kuo, “I forgot the formula.” How students can use coherence to reconstruct a (partially) forgotten equation, [arXiv:2506.19641](https://arxiv.org/abs/2506.19641).
- [31] E. Buckingham, On physically similar systems; illustrations of the use of dimensional equations, *Phys. Rev.* **4**, 345 (1914).
- [32] E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. T. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, Scientific abilities and their assessment, *Phys. Rev. ST Phys. Educ. Res.* **2**, 020103 (2006).
- [33] T.-R. Sikorski, G. D. White, and J. Landay, Uptake of solution checks by undergraduate physics students, presented at *PER Conf. 2017, Cincinnati, OH*, pp. 368–371, 10.1119/perc.2017.pr.087.
- [34] W. Demtröder, *Experimentalphysik 2: Elektrizität Und Optik*, Springer-Lehrbuch (Springer, Berlin, Heidelberg, 2013).
- [35] Kultusministerkonferenz, Abiturnoten im Ländervergleich, <https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/abiturnoten.html> (2024) [accessed November 6, 2024].
- [36] A. Renkl, Toward an instructionally oriented theory of example-based learning, *Cognit. Sci.* **38**, 1 (2014).
- [37] J. Wittwer and A. Renkl, How effective are instructional explanations in example-based learning? A meta-analytic review, *Educ. Psychol. Rev.* **22**, 393 (2010).
- [38] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/mb7q-kf2s> for intervention tasks, performance tests, and questionnaire.
- [39] N. D. Finkelstein and S. J. Pollock, Replicating and understanding successful innovations: Implementing tutorials in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **1**, 010101 (2005).
- [40] Intrinsic motivation inventory (IMI), <https://selfdeterminationtheory.org/intrinsic-motivation-inventory/> (2025) [accessed November 2024, 6].
- [41] L. Ding and R. Beichner, Approaches to data analysis of multiple-choice questions, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020103 (2009).
- [42] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (L. Erlbaum Associates, Hillsdale, NJ, 1988).
- [43] S. B. Morris and R. P. DeShon, Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs., *Psychol. Methods* **7**, 105 (2002).
- [44] M. J. Blanca, R. Alarcón, J. Arnau, J. García-Castro, and R. Bono, How to proceed when normality and sphericity are violated in the repeated measures ANOVA, *An. Psicol.* **40**, 466 (2024).
- [45] P. Mayring and T. Fenzl, Qualitative inhaltsanalyse, in *Handbuch Methoden Der Empirischen Sozialforschung* (Springer Fachmedien, Wiesbaden, 2019), pp. 633–648.
- [46] M. Wirtz and F. Caspar, *Beurteilerübereinstimmung Und Beurteilerreliabilität: Methoden Zur Bestimmung Und Verbesserung Der Zuverlässigkeit Von Einschätzungen Mittels Kategoriensystemen Und Ratingskalen* (Hogrefe, Göttingen, 2002).
- [47] M. Lenz and E. Gire, Faculty views of and expectations for dimensional analysis, presented at *PER Conf. 2016, Sacramento, CA*, pp. 196–199, 10.1119/perc.2016.pr.044.

- [48] G. Greefrath, R. Oldenburg, H.-S. Siller, V. Ulm, and H.-G. Weigand, *Didaktik Der Analysis* (Springer, Berlin Heidelberg, 2016).
- [49] J. A. Fernández-Plaza and A. Simpson, Three concepts or one? Students' understanding of basic limit concepts, *Educ. Stud. Math.* **93**, 315 (2016).
- [50] T. de Jong and M. G. M. Ferguson-Hessler, Types and qualities of knowledge, *Educ. Psychol.* **31**, 105 (1996).
- [51] C. Loretan, A. Müller, M. Delaval, S. Roch, and L. Weiss, Understanding of size and scale and order-of-magnitude reasoning in secondary science: A teaching experiment with worked examples as educational scaffold, [arXiv: 2506.19641](https://arxiv.org/abs/2506.19641).
- [52] D. R. M. Edge and M. K. Dirks, Problem solving, Enrico Fermi and the bull moose, *School Sci. Math.* **83**, 601 (1983).
- [53] U. Heublein, J. Ebert, C. Hutzsch, S. Isleib, R. König, J. Richter, and A. Woisch, *Zwischen Studienerwartungen Und Studienwirklichkeit. Ursachen Des Studienabbruchs, Beruflicher Verbleib Der Studienabbrecherinnen Und Studienabbrecher Und Entwicklung Der Studienabbruchquote* (Forum Hochschule, Hannover, 2017).
- [54] L. J. Cronbach, Coefficient alpha and the internal structure of tests, *Psychometrika* **16**, 297 (1951).