

# Sensemaking tasks with limiting case analysis: A dose-effect relationship in undergraduate physics education

Julia Hofmann<sup>1,\*</sup>, Pascal Klein<sup>1</sup>, Andreas Müller<sup>2</sup>, and Josefine Neuhaus<sup>1</sup>

<sup>1</sup>*Faculty of Physics, Physics Education Research, University of Göttingen, Friedrich-Hund-Platz 1, 37077, Göttingen, Germany*

<sup>2</sup>*Faculty of Sciences, Department of Physics and Institute of Teacher Education, University of Geneva, Boulevard du Pont d'Arve 40, 1211, Genève, Switzerland*



(Received 15 May 2026; accepted 21 September 2026; published 30 October 2026)

Making sense of mathematical expressions is a central component of expertlike reasoning in physics. Previous research has shown that sensemaking strategies can be successfully integrated into university instruction and that students benefit from explicit support in using such strategies. However, it remains largely unexplored to what extent the amount of practice influences students' learning. In the present study, we investigate dose-effect relationships for sensemaking tasks with limiting case analysis. The study was conducted during a two-week preparatory course for incoming undergraduate physics students at a German university. The course focuses on the repetition of important mathematical concepts and prepares students for the demands of the first semester of physics studies. Limiting case analysis was systematically embedded in daily exercises that students solved in recitation sessions. A total of  $N = 103$  students participated in the study and were randomly assigned to one of five intervention groups, each with a different number of limiting case tasks (0–8 tasks). The study employed a pre-post design to assess perceived usefulness, instructional integration, competence, and performance regarding the analysis of limiting cases. The results show that students' perceived competence increased with the number of tasks, with effect sizes up to  $d = 0.88$ . Students' performance also showed an increase in effect size for a greater number of tasks with effect sizes up to  $d = 1.07$ . The observed patterns in the target variables are compatible with a sigmoidal dose-effect relationship. Overall, this proof-of-concept study provides the first empirical evidence for dose-effect relationships in the context of sensemaking strategies in physics and demonstrates the feasibility of dose-effect study designs in undergraduate physics education research. The study highlights the importance of explicitly integrating sensemaking strategies into undergraduate physics and indicates that increased and targeted practice can foster both students' confidence and their ability to apply such strategies.

DOI: [10.1103/gx8s-rldd](https://doi.org/10.1103/gx8s-rldd)

## I. INTRODUCTION

Tying meaning to equations is a key component of a physicist's professional practice, frequently involving the application of sensemaking strategies (e.g., the analysis of limiting cases). While there is strong evidence that continuous and deliberate practice is a key factor in developing high levels of expertise in any field [1], particularly in physics [2,3], little is known on the extent to which learning outcomes depend on the amount of practice with such sensemaking strategies. It is likely to say that increased practice results in improved learning

outcomes. However, the exact relationship between the dose (amount of practice) and the effect (learning outcome) remains unclear. Such dependencies are investigated in terms of “dose-effect relationships” [4] or “learning curves” [5]. They demonstrate how learning outcomes depend on the amount of practice, and they have already been investigated in other fields [6–8]. The present study investigates how students' perceptions and performance with respect to sensemaking strategies develop as a function of a systematically varied number of completed tasks.

## A. Sensemaking with limiting case analysis

Sensemaking refers to a dynamic process in which learners construct new knowledge or resolve gaps and inconsistencies in their existing understanding by actively “figuring something out” [9]. In physics, sensemaking is closely associated with coordinating mathematical

\*Contact author: [julia.hofmann@uni-goettingen.de](mailto: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.

structures with contextual information and physical interpretations [10–12]. A variety of strategies have been identified in the literature to support the process of making sense of physics equations [10,13–16]. Among these strategies, the analysis of limiting cases (LCA) is a particular important one. It requires the construction and dynamic manipulation of mental models, as well as the coordinated use of mathematical concepts like limits and infinity [12,17]. A variable is mathematically pushed toward an extreme value (e.g., infinity) which is expected to reveal characteristic properties of the physical situation. Consider, for instance, a point charge  $q$  with mass  $m$  moving with constant speed  $|\vec{v}|$  in a homogeneous magnetic field  $|\vec{B}|$ . If  $\alpha$  denotes the angle between the velocity and the magnetic field, the trajectory is in general a helix, whose radius of curvature is given by

$$r = \frac{m|\vec{v}|}{|q||\vec{B}|\sin\alpha}.$$

When taking the mathematical limit  $\alpha \rightarrow 0^\circ$ , the radius becomes infinitely large  $r \rightarrow \infty$ . This result is physically plausible because a charge moving parallel to the magnetic field does not experience the Lorentz force. Therefore, it is not deflected and moves in a straight line, which is presented as an infinitely large radius. In this way, limiting case analysis links the mathematical structure to qualitative expectations about the physical situation. Examples of structural elements that invite a limiting case analysis include the term  $\frac{1}{x}$  for very large or very small  $x$ , and terms in which the cosine of an angle appears in the denominator as the angle approaches  $\frac{\pi}{2}$  (see Fig. 3).

Previous research provides evidence that several sensemaking strategies can be successfully fostered through instruction and that students perceive them as beneficial for their learning [12,18,19]. In a prior intervention study conducted in a first-year university physics course on electricity and magnetism, we integrated fundamental sensemaking strategies (including limiting case analysis) into weekly exercise sheets [20]. Students showed increased confidence in applying these strategies and improved performance after a seven-week intervention period. However, limiting case analysis exhibited particularly strong effects on students' perceptions, while performance-related effects were less pronounced compared to other integrated strategies. A possible explanation may lie in the complexity, which may require students to engage more frequently with exercises.

Despite their central role in physicistlike reasoning, sensemaking strategies are rarely explicitly integrated into university courses. At the same time, many studies observe that students experience difficulties in meaningfully interpreting equations and assessing their plausibility [11,16,21,22]. Thus, supporting students in assigning meaning to mathematical expressions and evaluating their

physical plausibility remains an important aspect of physics instruction.

## B. Research on the relation of practice and learning outcomes (dose-effect relationships)

In undergraduate physics programs, students are often encouraged to solve many problems in order to succeed on exams [23]. While this assumption is widespread, there are only a few studies on the actual dose-effect relationship in physics education. Detailed knowledge of the relationship between the amount of practice and learning outcomes is valuable at the foundational and practical levels [7]. On a foundational level, this knowledge would deepen our understanding of practice-based learning and provide an empirical ground for evidence-based instructional approaches. On the practical level, the consequences of a poorly calibrated practice dose can be substantial in either direction. Too little practice may fail to produce meaningful learning gains, while too much practice wastes precious instructional time that could be invested more effectively elsewhere. In both cases, learning outcomes and student engagement may suffer due to a lack of success or a sense of futility.

An early study by Kim and Pak [22] investigated the relationship between the quantity of traditional textbook problem solving and students' conceptual understanding in introductory mechanics. The study involved 27 first-year physics education majors at a Korean university. Students self-reported the number of textbook problems they had solved in preparation for an exam (ranging from approximately 300 to 2900 problems per student). Despite this intense practice, little correlation between the number of solved problems and students' performance (mechanics and mathematics tests) was found. The authors concluded that increasing the quantity of traditional problem solving alone does not necessarily lead to improved conceptual understanding. However, the study provides only limited methodological detail and does not report a systematic statistical analysis of the relationship between the amount of practice and test performance.

In contrast, a more recent large-scale study suggests a positive association between practice volume and course performance. Black *et al.* investigated the relationship between students' voluntary use of an online practice system and final grades in introductory STEM courses [23]. The system provides a large pool of multiple-choice questions designed to support exam preparation. For introductory physics courses, they found that higher practice volume on the study system was associated with higher final course grades. After controlling for students' precollege mathematics test scores, the mean course grade increased approximately quadratically with the logarithm of the total number of questions answered over the semester. This indicates a nonlinear but monotonic relationship. Similar results were reported in an earlier study

using the same practice system [24]. However, it should be noted that the studies did not employ a pre-post design and participation in the online system was voluntary. Furthermore, the multiple-choice questions were closely aligned with the format, structure, and content of the actual course exam.

To investigate such relationships in more detail, Vogt and coauthors have employed a systematic dose-effect study design in lower German secondary physics classes in grades 9 and 10 [4,25]. They investigated the impact of context-based science education using advertisement story problems. The tasks included physics questions on real advertisements that were related to a scientific context (e.g., an advertisement for a portable heating system). In this study, the number of tasks integrated into regular physics lessons was systematically varied (0–17 tasks) within a pre-post intervention design. The number of tasks was found to be associated with changes in several affective variables, including students' motivation toward physics, self-concept, and perceived authenticity of the tasks. At low numbers of tasks, only small effects were observed. These were followed by a steep increase over a moderate range of task numbers, before the curve flattened again at higher numbers of tasks, indicating a saturation level. This relationship can be described by a sigmoidal (logistic) function of the form

$$d(x) = \frac{b}{1 + e^{\frac{x-x_0}{s}}}, \quad (1)$$

where  $x$  denotes the number of tasks,  $x_0$  the inflection point of the curve,  $b$  the saturation level, and  $s$  a slope parameter [25,26]. The function is characterized by a lower asymptote, a region of steep increase around the inflection point, and an upper asymptote.

In two studies by Kotsis and Panagou [27,28], common misconceptions of force and energy were investigated across different age groups (primary school students to in-service primary education teachers; ages 11–50). Using a self-developed multiple-choice questionnaire, the authors report an increase in correct answers with age, approximated by a sigmoidal curve. However, as the study relies on a cross-sectional rather than a longitudinal design, causal conclusions about learning development have to be drawn with caution.

As this short review shows, only limited research in physics education research has explicitly addressed dose-effect relationships in instructional interventions. Evidence from other disciplines of learning research, however, suggests that such relationships are typically nonlinear. Murre, for example, investigated the relationship between the number of learning trials and learning success in foreign-language vocabulary acquisition [5]. The results show that the learning curves follow an S-shaped curve rather than a linear increase and can be described with a

logistic or sigmoidal function. Learning success initially improves slowly, followed by a phase of rapid gains, and eventually approaches a plateau. This sigmoidal pattern is particularly pronounced for learners with lower learning rates and for more complex or unfamiliar learning material. In contrast, faster and more experienced learners tend to show more asymptotically increasing, exponential-like learning progress. Such exponential-like curves can be obtained as an approximation of the upper branch of the logistic curve when the inflection point lies far to the left of the observed range. A derivation of the relative deviation between the sigmoidal and exponential models is provided in the Appendix. The findings of Murre highlight that the effectiveness of increasing practice depends both on learner characteristics and on the complexity of the learning topic, and that equal increases in practice dose do not necessarily result in equal learning gains.

Several theoretical considerations also support the wide occurrence of nonlinear dose-effect relationships. First, Heathcote *et al.* [29] have shown that an exponential function often fits individual-level learning data better, and that the often-cited “power law of practice” is obtained from averaging across individuals. Second, Murre [5] has shown that for a compound task, composed of several independent subtasks, the dose-effect relationship is sigmoidal, even if the curves of the subtasks are power laws or exponential. Third, Pritchard *et al.* [30] derived logistic dose-effect relationships as one of several mathematical learning models based on general assumptions. Finally, Grebenc [26] identifies 30 different mathematical scenarios leading to sigmoid functions.

Overall, existing research indicates that the relationship between practice and learning outcomes is unlikely to be linear. Instead, threshold and saturation effects appear to be characteristic features of learning processes. Against this background, investigating dose-effect relationships for specific sensemaking strategies in physics constitutes a promising and still unexplored direction for physics education research. A deeper understanding of these relationships would be of particular practical value for undergraduate courses, where available practice time is inherently limited. This could enable educators to identify an effective dose of practice needed to achieve meaningful learning gains.

### C. Research questions

Building on theoretical considerations of sensemaking and the limited empirical evidence on effective amounts of practice, this study investigates dose-effect relationships in the context of sensemaking strategies in undergraduate physics. As a starting point, we first examine students' initial perceptions of a sensemaking strategy at the very beginning of their university studies, in particular their perceived usefulness, their perceived integration of the strategy into previous courses, and their perceived

TABLE I. Overview of topics addressed in the daily lectures and corresponding exercise sheets, including topics of intervention tasks and corresponding equations.

| Sheet no. | Topic of daily lecture and exercise sheet | Physical context                         | Equation                   |
|-----------|-------------------------------------------|------------------------------------------|----------------------------|
| 1         | Physical quantities and units             | Blood flow and constricted blood vessels | Hagen-Poiseuille equation  |
| 2         | Rational functions                        | Growth of an animal population           | Logistic equation          |
| 3         | Differential calculus                     | Reflection and refraction of light       | Snell's law                |
| 4         | Integrals                                 | Free fall of skydiver                    | Velocity equation          |
| 5         | Vector calculus, multivariable calculus   | Interaction of atoms                     | Lennard-Jones potential    |
| 6         | Coordinate systems                        | Planetary orbits                         | Gravitational force        |
| 7         | Differential operators, line integrals    | Rocket in Earth's gravitational field    | Escape velocity            |
| 8         | Surface and volume integrals              | Oscillating object in water              | Damped harmonic oscillator |

competence. These findings provide the foundation for the subsequent dose-effect analyses.

As outlined in the theoretical background, dose-effect relationships have rarely been investigated in physics education research, and no prior study has systematically examined how the extent of practice with sensemaking strategies affects students' learning outcomes. Addressing this gap, the present study focuses on limiting case analysis as a particularly challenging and important sensemaking strategy [20]. The present study builds on the initial baseline by investigating how students' perceptions and their performance develop as a function of practice. Therefore, the amount of practice is operationalized as the number of intervention tasks.

We address the following research questions:

What is the dose-effect relationship between the number of intervention tasks and...

RQ1: students' perceived usefulness of limiting case analysis?

RQ2: students' perceived competence in applying limiting case analysis?

RQ3: students' performance in applying limiting case analysis?

Perceived integration in previous and current courses was analyzed separately as a manipulation check, verifying that students perceived the experimentally varied dose as intended, and was therefore not formulated as an additional research question.

## II. METHODS

### A. Data collection and sample

The study was conducted within a preparatory course offered prior to the winter term 2024 of the undergraduate physics degree program at Göttingen University in Germany. The course spans two weeks and takes place immediately before the start of the first semester. It is specifically designed to prepare incoming undergraduates for their physics studies by repeating core mathematical concepts and by providing insights into topics typically encountered in the first year of the curriculum. Daily lectures (2 h each) cover a range of topics, including,

for example, partial fraction decomposition, differential calculus, integrals, vector calculus, and multivariable calculus. A complete overview of lecture topics is provided in Table I. Each lecture introduced a new topic, which students explored in greater depth through corresponding exercises on exercise sheets. During daily recitation sessions, students deepened their understanding of the lecture content, clarified questions, and worked through the exercises. Students were assigned to one of eleven recitation groups, which were supervised by a student tutor.

Demographic characteristics of all participants are summarized in Table II. At the beginning of the course,  $N_0 = 127$  students participated in the first lecture and also in the study. By the end of the course,  $N_1 = 109$  students were still attending the lectures and participating in the study. The majority of participants were enrolled in the undergraduate physics program, and a smaller proportion enrolled in the teacher training program with the intention of becoming secondary school physics teachers.

### B. Study design

The study was conducted over the two-week duration of the course. An overview of the study design is provided in Fig. 1. Data were collected at two time points: at the beginning of the course (T0, within the first lecture) and at

TABLE II. Sample characteristics of the initial sample at T0 and the core sample (mean and standard deviation).

|                                           | Sample T0      | Core sample    |
|-------------------------------------------|----------------|----------------|
| Sample size                               | 127            | 103            |
| Female                                    | 30             | 12             |
| Age (in years)                            | $19.8 \pm 2.4$ | $19.5 \pm 1.3$ |
| Grade of high school diploma <sup>a</sup> | $1.6 \pm 0.6$  | $1.6 \pm 0.5$  |
| First-time participants                   | 99%            | 99%            |
| Physics major                             | 90%            | 90%            |
| Physics teacher training                  | 7%             | 8%             |

<sup>a</sup>Students reported their grade. Range: 1.0 (best score) to 4.0. In 2024, the average high school diploma in Germany was: 2.33 [31].

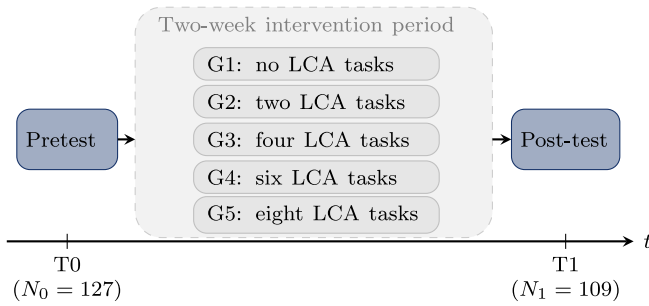


FIG. 1. Study design with pretest (T0) and post-test (T1) including questionnaires. During the two-week intervention period, students were assigned to different groups (G1–G5) that received systematically varied numbers of limiting case analysis tasks (0–8 tasks).

the end of the course (T1, within the last lecture). For the analyses,  $N = 103$  matched datasets from T0 and T1 were available, constituting the core sample of the study (see Table II for demographic characteristics).

The intervention followed a dose-effect study design and was implemented within the daily exercise sheets (eight sheets in total). The intervention dose varied based on a different number of limiting case analysis (LCA) tasks. Five intervention conditions were defined, ranging from no tasks on limiting case analysis (group G1) to limiting case analysis tasks on each of the eight exercise sheets (group G5). Groups G2–G4 received two, four, or six tasks. Each group consisted of at least two recitation groups. The distribution of all tasks is summarized in Table III. With respect to holidays and weekends, we attempted to distribute the intervention dose for all groups evenly across the two-week course. To compensate for the higher workload of the intervention dose, additional tasks were given to lower-dose groups, keeping the total time on task approximately constant among the five groups. These compensatory tasks focused on dimensional analysis as a sensemaking strategy that similarly involves verifying the plausibility of equations, but is conceptually distinct from LCA as it does not involve taking mathematical limits or reasoning about extreme physical scenarios.

Participation in the course was voluntary. Due to its preparatory nature, attendance at lecture and recitation

sessions was recommended but not mandatory. In accordance with our experiences and observations from previous years, the course is widely perceived as beneficial, which typically results in high participation rates among the first-year physics students. Participation in the accompanying study was also voluntary. Pseudonymous codes were used to match students' responses across two measurements, ensuring that no conclusions could be drawn about the individual.

## C. Materials and measures

All study materials, including performance tests, questionnaires, and intervention exercises, are included as translated versions (original language: German) in Supplemental Material [32].

### 1. Intervention tasks

The intervention was implemented across eight exercise sheets that students worked on during daily recitation sessions in small groups under the supervision of a tutor. Recitation sessions provide a particularly suitable setting for instructional interventions, as students actively engage with the learning material, work on exercises, and receive immediate support from a tutor if needed [33]. Each exercise sheet included up to one task related to the intervention.

The intervention tasks were designed to support students' sensemaking with limiting case analysis (see Fig. 2 for an example). Building on previous implementations of similar instructional approaches [10,12,18,20], the tasks followed a comparable structure. Each task was embedded in a physically meaningful but conceptually accessible context. After a set of general and context-specific questions, students were required to apply limiting case analysis twice within a task. First, they analyzed a specified extreme case to assess the physical plausibility of a correct equation and to justify that the equation is physically plausible. In a second assignment, students were presented with an incorrect equation within the same physical context and were asked to analyze a relevant limiting case in order to justify that the equation is not physically plausible. For a detailed overview of all equations and limiting cases that required consideration, please refer to Tables I and V.

To support students' engagement with the strategies, all participants had access to a guide that introduced the sensemaking strategy in a step-by-step manner, including worked examples [34].

### 2. Perception of limiting case analysis and the intervention

To assess students' perceptions of the sensemaking strategy, a questionnaire was deployed. The exact wording of all items is provided in Table VI in the Appendix.

TABLE III. Distribution of limiting case analysis tasks across intervention groups G1–G5, including group size, number of tasks, and the exercise sheets on which the tasks were assigned.

| Group | Students | LCA tasks | Sheet no.         |
|-------|----------|-----------|-------------------|
| G1    | 23       | 0         | ...               |
| G2    | 32       | 2         | 2, 5              |
| G3    | 24       | 4         | 2, 3, 4, 6        |
| G4    | 9        | 6         | 1, 2, 3, 4, 6, 7  |
| G5    | 15       | 8         | 1–8 (every sheet) |

**Task** *Hagen–Poiseuille equation* 

We are interested in the volume flow  $\Phi$  (unit:  $[\Phi] = \text{m}^3/\text{s}$ ) through a pipe with radius  $R$  and length  $L$ . The volume flow depends on the pressure drop  $\Delta P/L$ , pipe radius  $R$ , and viscosity  $\eta$  of the fluid (the more viscous a fluid, the greater  $\eta$ ). The pressure corresponds to the force per surface area. The unit of pressure is the Pascal  $[P] = \text{Pa} = \text{N}/\text{m}^2$  and the unit of viscosity is  $[\eta] = \text{Pa} \cdot \text{s}$ .

In a dimensional analysis, the following dependence of the volume flow on the above parameters was asserted:

$$\Phi = C \left( \frac{\Delta P}{L} \right)^\alpha R^\beta \eta^\gamma,$$

where  $C$  is a dimensionless constant.

- (a) The following exponents were determined by a dimensional analysis:  $\alpha = +1$ ,  $\beta = 4$  and  $\gamma = -1$ . Can the constant  $C$  also be calculated in a dimensional analysis?
- (b) How does the flow through the pipe change when the pipe diameter decreases by a factor of 2 due to corrosion?
- (c) Atherosclerosis refers to the pathological accumulation of cholesterol and other fats in the inner wall layer of arterial blood vessels. Assume that Hagen-Poiseuille's law also applies to blood.

Rewrite Hagen-Poiseuille's equation in terms of  $\Delta P$ . Verify the rewritten law by considering the limiting case of arteries that become narrower and narrower. Based on your analysis, justify that the equation is physically plausible.

*Note: You can read how to proceed when considering limiting cases at any time in the "Strategy Guide".*

- (d) As part of an exercise, the following *incorrect* equation for  $\Delta P$  was derived:

$$\Delta P = \frac{\Phi L R^4}{C \eta}.$$

Check the equation by selecting and considering a relevant limiting case. Based on your analysis, justify that the equation is not physically plausible.

FIG. 2. Example of an intervention task (translated from the German original) embedded in the first exercise sheet (overall topic: physical quantities/dimensions). After some general questions, students were asked to analyze a relevant limiting case of (c) a correct equation and (d) an incorrect equation and justify that the equation is (not) physically plausible.

All items were rated on a four-point Likert-type scale ranging from 1 ("strongly disagree") to 4 ("strongly agree"). The questionnaire comprised 11 items in total in order to assess students' initial perceptions of limiting case analysis at the beginning of their studies.

Four items measured students' perceived competence in applying each strategy (e.g., "*I feel confident using the strategy.*"), and four items assessed their perceived usefulness (e.g., "*I believe that using the strategy has a practical benefit for me.*"). These eight items were adapted from corresponding scales of the *Intrinsic Motivation Inventory* [35,36]. Additionally, three items referred to students' perception of the extent to which the analysis of limiting cases has been integrated into previous courses

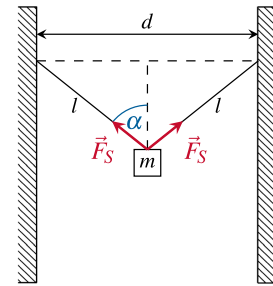
(e.g., "*I was able to practice applying the strategy in previous courses.*"), with the post-test version explicitly including the present one ("*including this course*"). These three items served as a manipulation check rather than as an outcome associated with the research questions. Specifically, they were used to verify that students in the higher-dose groups perceived LCA as being more strongly integrated and practiced during the course, thereby supporting the validity of the experimentally varied intervention dose.

In a previous study that employed the same questionnaire, its reliability was good, and no additional modifications were necessary [20]. In the present study, analyses of the data at time point T0 likewise indicated good to

### 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 street light is suspended are both of length  $l$  and are attached to the walls at the same height. The angle  $\alpha$  assumes values in the range  $0^\circ \leq \alpha < 90^\circ$ . The force on one cable  $|\vec{F}_S|$  can be described as follows:

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



In the following, a limiting case of the equation will be examined in order to assess the physical plausibility\* of the equation without deriving the equation itself.

\* An equation is considered physically plausible if, when tested under typical conditions (e.g., limiting cases), it yields results that are consistent with known facts or with experience and intuition.

- (a) Examine a **relevant limiting case** to assess the physical plausibility of the equation.
- (b) Based on your analysis, justify whether the equation is physically plausible.

FIG. 3. Example of a task used in the performance test (translated from the German original). After an introduction of the physical context and the corresponding equation, students were asked to (a) analyze a relevant limiting case and (b) justify the physical plausibility of the equation. A similar task was used in a previous study [20].

excellent internal consistency, as reflected by high values of Cronbach's  $\alpha$  [37]. A detailed overview of psychometric properties, including Cronbach's  $\alpha$  and item-total correlations is provided in Table VI in the Appendix.

### 3. Performance tests

In accordance with traditional pre-post study designs, performance tests were administered alongside the questionnaires to evaluate students' ability to perform limiting case analysis.

At both time points, the performance test consisted of the same five tasks. The wording and structure of these tasks were influenced by previous studies [10,12,18,20]. Example tasks are shown in Figs. 3 and 4. In three tasks, students were presented with a correct physical equation and asked to select and analyze a relevant limiting case (*"Examine a relevant limiting case to obtain clues as to the physical plausibility of the equation."*). In two additional tasks, students were given both correct and incorrect equations. They were required to analyze an appropriate limiting case for each equation (*"For each of the three equations, examine a relevant limiting case to obtain clues as to the physical plausibility of the equation."*).

Additionally, students were asked to justify whether the equation under consideration was physically plausible.

### D. Analysis methods and data processing

Data analyses were conducted using standard statistical methods, and the assumptions required for each analysis were verified beforehand. For comparisons of means, effect sizes were quantified using Cohen's  $d$  and interpreted following conventional criteria [38]. For analyses involving variances, total eta squared ( $\eta^2$ ) was reported as a measure of effect size [38]. When the assumption of sphericity was violated, as indicated by Mauchly's sphericity test, appropriate corrections (Huynh-Feldt or Greenhouse-Geisser) were applied depending on the value of  $\epsilon$  [39].

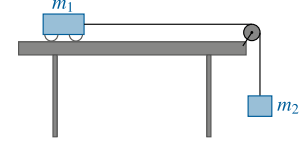
Analyses of the scales of the questionnaire were conducted, and the internal consistency was evaluated using Cronbach's  $\alpha$ . The interpretation was performed in accordance with established guidelines [40].

The normalized Hake gain  $g_H$  was calculated to measure students' learning gains [41,42]. It is defined as the ratio of students' absolute improvement to the maximum possible improvement.

### Task – Two chained masses

A mass  $m_2$  is connected to a mass  $m_1$  via a rope and a pulley. We assume that the system is frictionless. The acceleration due to gravity is  $g$ . After being released, mass  $m_2$  starts to accelerate until mass  $m_1$  reaches the pulley. As part of an exercise, the following equations for the acceleration  $a$  were derived:

$$1. a = \frac{(m_1 + m_2)g}{m_1}, \quad 2. a = \frac{m_2 g}{m_1 + m_2}, \quad 3. a = \frac{m_1 g}{m_1 + m_2}.$$



- (a) For each of the three equations, examine **relevant limiting cases** to assess the physical plausibility of the equation.
- (b) Based on your analysis, justify whether each of the three equations are physically plausible.

FIG. 4. Example of a task used in the performance test with correct and incorrect equations (translated from the German original). After an introduction of the physical context and the three equations, students were asked to (a) analyze relevant limiting cases and (b) justify the physical plausibility of each equation.

To analyze students' performance, a category system was applied to capture both the application and the reasoning processes (see Table VII in the Appendix). The coding framework for evaluating the application of limiting case analysis was adapted from Etkina *et al.*'s scientific abilities framework [43] and has been successfully deployed in an earlier study [20]. Student responses were rated using four categories: 0 ("no analysis"), 1 ("analysis attempted"), 2 ("flawed analysis"), and 3 ("flawless analysis"). Interrater reliability was determined using the Weighted Cohen's Kappa  $\kappa_w$  [44]. Two independent raters each coded approximately 10% of the data in two successive rounds. Interrater agreement improved from  $\kappa_w = 0.86$  in the first round to  $\kappa_w = 0.92$  in the second round, indicating excellent reliability.

The evaluation of students' explanations of the plausibility of an equation followed a structure similar to that used for rating strategy application. A separate category system was developed based on the scientific abilities framework developed by Etkina *et al.* [43], with a particular focus on explanation processes. After an initial rating ( $\kappa_w = 0.50$ ) and modification of the category system, interrater agreement improved to a good level in a subsequent rating round ( $\kappa_w = 0.81$ ).

To ensure comparability across measures, data from all sources were linearly transformed to the interval [0,1], with higher values indicating a higher scale value (e.g., high performance).

### III. RESULTS

The core sample of  $N = 103$  students is representative of the initial sample in terms of demographic data.

A significant difference was observed in the grade of students' high school diploma between students who took part in both measurements and those who dropped out. Students who did not take part in the second measurement had, on average, lower high school grades than the students who were in the core sample [ $t(123) = -61.63, p < 0.001, d = 0.75$ ]. No significant differences were found with respect to performance scores in the pretest [ $t(125) = -0.25, p = 0.81$ ].

Central descriptive statistics of affective and performance measures are reported separately for each intervention group in Table VIII in the Appendix. Group G4 was excluded from regression analyses due to its comparatively small sample size (see Table III).

Changes in students' perceptions and performance during the intervention period were analyzed with respect to the different intervention groups (G1–G5) and the time point (T0, T1). The results of the two-way analyses of variances are summarized in Table IV.

Before addressing the research questions, we first report the manipulation check to verify that the experimentally varied task doses were perceived as intended.

#### A. Perceived integration (manipulation check)

Students' initial perception (T0) of the integration in past courses was moderate ( $0.43 \pm 0.3$ ). No intervention group dependent differences were found [ $F(3, 90) = 1.21, p = 0.31, \eta^2 = 0.04$ ].

Students' perception of the integration of limiting case analysis in past and current courses exhibited a group-dependent change over time (see Fig. 5). A significant interaction between the intervention group and time was

TABLE IV. Results of the two-way analyses of variance (group  $\times$  time) regarding limiting case analysis for the dependent variables considered in this study. When a significant interaction was observed, the corresponding main effects are not reported. The test statistics provided include degrees of freedom  $df$  and effect sizes  $\eta^2$ .

| Dependent variable    | Interaction/factors | $F$   | $df$ | $p$    | $\eta^2$ |
|-----------------------|---------------------|-------|------|--------|----------|
| Usefulness            | Group $\times$ time | ...   | ...  | ...    | ...      |
|                       | Group               | ...   | ...  | ...    | ...      |
|                       | Time                | 10.96 | 87   | 0.001  | 0.10     |
| Perceived integration | Group $\times$ time | 11.33 | 87   | <0.001 | 0.20     |
|                       | Group               | ...   | ...  | ...    | ...      |
|                       | Time                | ...   | ...  | ...    | ...      |
| Perceived competence  | Group $\times$ time | 3.49  | 86   | 0.02   | 0.09     |
|                       | Group               | ...   | ...  | ...    | ...      |
|                       | Time                | ...   | ...  | ...    | ...      |
| Performance           | Group $\times$ time | 2.02  | 90   | 0.01   | 0.07     |
|                       | Group               | ...   | ...  | ...    | ...      |
|                       | Time                | ...   | ...  | ...    | ...      |
| Explanation           | Group $\times$ time | ...   | ...  | ...    | ...      |
|                       | Group               | ...   | ...  | ...    | ...      |
|                       | Time                | 53.81 | 90   | <0.001 | 0.34     |

observed. An examination of the pairwise comparisons over time reveals an increase in effect size (Cohen's  $d$ ) with the number of tasks completed within the intervention period. For all groups except G1, medium- to large-sized effects between T0 and T1 were found [all:  $p < 0.001$ ; G2:  $d = 0.75$ ; G3:  $d = 1.22$ ; G5:  $d = 1.75$ ]. As the number of tasks increases, the effect size rises but shows clear signs of saturation. While the increase from two to four tasks (G2 to G3) yields a substantial difference in effect size ( $\Delta d = 0.47$ ), further increases in number of tasks lead to comparatively smaller gains relative to the additional practice (dose) provided. This pattern is consistent with the characteristic flattening of a sigmoidal dose-effect curve.

Figure 6 shows effect sizes (effect) as a function of the number of tasks (dose). The following observations can be made: First, the relationship is monotonic. Second, a

medium effect size becomes evident once students complete more than two tasks. Third, the dose-effect relationship is compatible with a logistic curve, with an excellent goodness of fit (Pseudo- $R^2 = 0.99$ ). Because group G1 exhibited a negative effect size due to decreased perceptions in the absence of intervention tasks, an additional parameter resembling a shift in  $y$  direction was introduced to account for a nonzero baseline. All regression parameters are summarized in Table IX in the Appendix. An exponential model was additionally fitted to all variables and the corresponding plots and regression parameters are provided in Fig. 8 and Table X in the Appendix.

### B. Perceived usefulness

At the beginning of the preparatory course (T0), students perceived the usefulness of LCA as being very

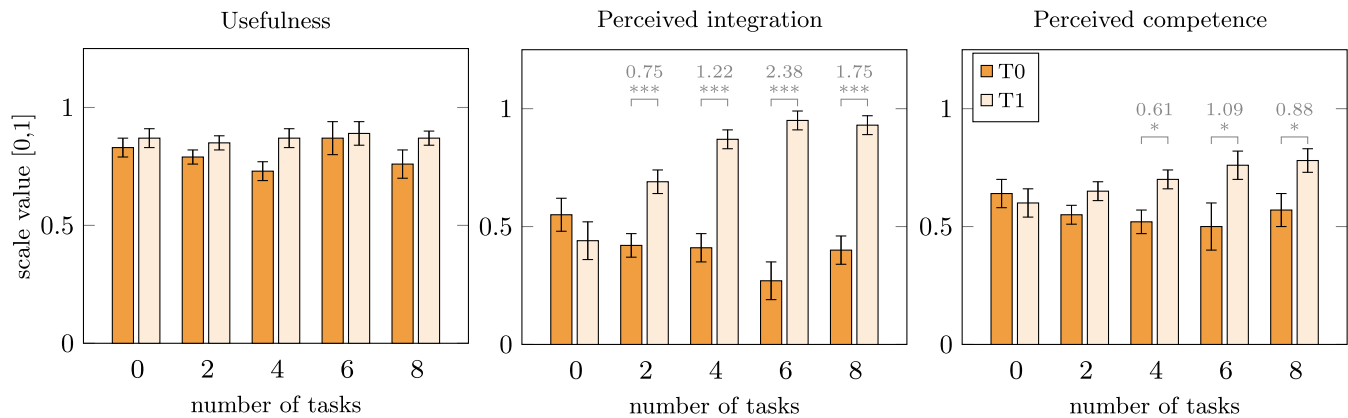


FIG. 5. Students' perceptions of the usefulness (left), integration into previous and current courses (center), and competence (right) of limiting case analysis (mean and standard error). Dark shading indicates T0 and light shading indicates T1. Values indicate Cohen's  $d$ . Varying intervention in terms of number of tasks: 0 tasks (G1), 2 tasks (G2), 4 tasks (G3), 6 tasks (G4), 8 tasks (G5).

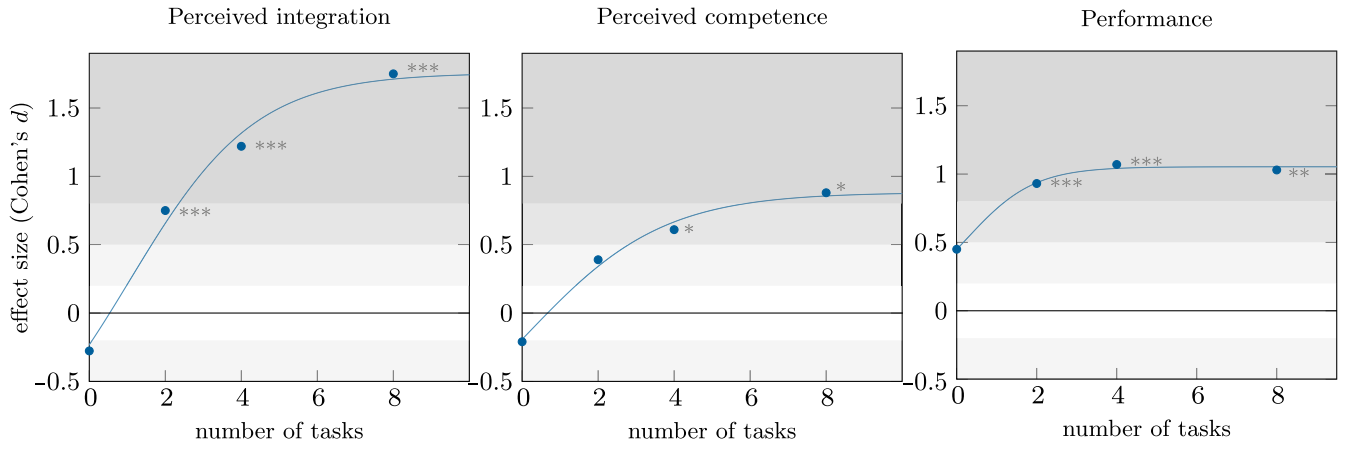


FIG. 6. Effect sizes (Cohen's  $d$ ) as a function of the number of limiting case analysis tasks for perceived integration in previous and current courses (left), perceived competence (center), and performance (right). Points represent observed effects, and solid lines indicate fitted logistic regression models illustrating the dose-effect relationship. The goodness of fit was very high in each case, and the corresponding model specifications are reported in Table IX in the Appendix. Data points for group G4 were excluded due to small sample size (see text). Shaded regions denote small ( $0.2 < |d| < 0.5$ ), medium ( $0.5 < |d| < 0.8$ ), and large ( $|d| > 0.8$ ) effect sizes according to Cohen [38]. Significance levels are indicated as  $*p < 0.05$ ,  $**p < 0.01$ , and  $***p < 0.001$ .

high ( $0.78 \pm 0.21$ ), with no intervention group dependent difference [ $F(3, 90) = 1.04$ ,  $p = 0.38$ ,  $\eta^2 = 0.03$ ].

Throughout the intervention period, there was no substantial change in this measure across groups. Only a medium-sized temporal main effect was observed, indicating an increase in all students' perceptions, regardless of the intervention group (see Fig. 5).

### C. Perceived competence

At the beginning of the course (T0), students felt moderately confident in analyzing limiting cases ( $0.56 \pm 0.27$ ), and no intervention group dependent differences were found [ $F(3, 89) = 0.92$ ,  $p = 0.43$ ,  $\eta^2 = 0.03$ ].

Students' perceived competence in applying limiting case analysis increased over the intervention period (see Fig. 5), as reflected by a medium-sized interaction effect between group and time. A significant increase over time was observed for groups G3 and G5 with medium- and large-sized effects [G3:  $p = 0.01$ ,  $d = 0.61$ ; G5:  $p = 0.02$ ,  $d = 0.88$ ].

The dose-effect pattern for students' perceived competence closely mirrors the results obtained for perceived integration (see Fig. 6). When effect sizes are related to the number of tasks completed within the intervention period, a monotonic increase with a tendency toward a saturation at higher numbers of tasks becomes apparent. The data are again compatible with a logistic model, demonstrating an excellent goodness of fit ( $\text{Pseudo-}R^2 = 0.99$ ).

### D. Performance

Performance regarding limiting case analysis was evaluated by considering the application of the strategy and the subsequent explanation of the physical plausibility across

all test items (see example task in Fig. 3). A summary of both measures is presented in Fig. 7.

Students' initial performance before the intervention (T0) was very low ( $0.17 \pm 0.22$ ). No differences between the groups were observed for either the application or explanation process [application:  $F(3, 90) = 0.90$ ,  $p = 0.44$ ,  $\eta^2 = 0.03$ ; explanation:  $F(3, 90) = 0.64$ ,  $p = 0.60$ ,  $\eta^2 = 0.02$ ].

After the intervention period, an interaction between intervention group and time was observed for the application of limiting case analysis, indicating a group-dependent change over time. With the exception of group G1, all groups showed a significant increase in performance from T0 to T1 with medium to large effect sizes [G2:  $p < 0.001$ ,  $d = 0.93$ ; G3:  $p < 0.001$ ,  $d = 1.07$ ; G5:  $p = 0.001$ ,  $d = 1.03$ ]. When plotting the effect size against the number of tasks, a steep monotonic increase with a subsequent plateau can be observed (see Fig. 6). In accordance with students' perceptions, a logistic fit was employed to model the dose-effect relationship, which demonstrated a good fit ( $\text{Pseudo-}R^2 = 0.99$ ).

The normalized Hake gain (see Fig. 7) showed increasing values across groups [G1:  $g_H = 0.14$ ; G2:  $g_H = 0.23$ ; G3:  $g_H = 0.31$ ; G5:  $g_H = 0.32$ ].

Students' explanation scores during the intervention showed no interaction between the factors group and time. Only a significant temporal main effect with a large effect size was found. All groups improved their explanation score, independent of the intervention group.

## IV. DISCUSSION

Before examining the dose-dependent effects during the intervention, we first discuss students' initial perceptions

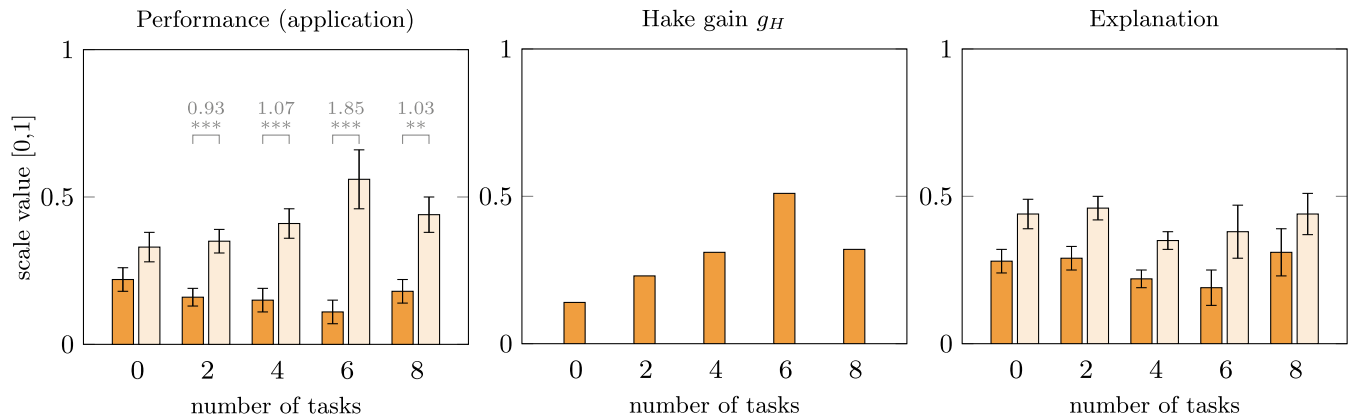


FIG. 7. Students' performance (mean and standard error) of limiting case analysis for strategy application (left), explanation (right), and normalized Hake gain  $g_H$  of students' performance (center). Dark shading indicates T0 and light shading indicates T1. Values indicate Cohen's  $d$ .

of limiting case analysis at the beginning of the preparatory course. Overall, students rated limiting case analysis as highly useful. Students' perception of the usefulness did not show differences in comparison with the results of a prior study that employed the same questionnaire but with second-semester students [20]. Both cohorts recognize the importance of the sensemaking strategy on a similar level, though even the second-semester students may not feel prepared to use it. After one semester of studying physics, students feel less confident in applying limiting case analysis as compared to the other cohort with students at the beginning of their studies. Overall, students show a low level of confidence in applying limiting case analysis in both studies. This is in line with the reports of both cohorts that the strategy was rarely addressed, both before and during the first semester.

Taken together, the low perceived competence and limited reported exposure to limiting case analysis in both studies suggest that students are not receiving sufficient training on limiting case analysis. Similarly, other studies report that such strategies are rarely practiced in undergraduate settings [16,45]. This further highlights the importance of explicitly and repeatedly integrating limiting case analysis into undergraduate courses.

#### A. Perceived integration (manipulation check)

Students' perceived integration of limiting case analysis increased systematically with the number of completed tasks. Groups that worked on more tasks reported higher perceptions of integration, whereas group G1, which did not receive any corresponding intervention tasks, showed a decrease. When effect sizes are plotted as a function of number of tasks, the resulting pattern is consistent with a sigmoidal dose-effect relationship. Such learning curves have been reported in prior research for affective and performance variables and are interpreted as reflecting an initial threshold, a phase of steep increase, and a

subsequent saturation [4,5,7,25]. The present data appear to capture primarily the upper branch of this curve. The present data show a medium-sized effect after two tasks and a large effect after four tasks. Since this variable served as a manipulation check, the intervention-sensitive dose-effect pattern provides evidence that students largely perceived the varying amounts of practice as intended.

#### B. Perceived usefulness

Students evaluated limiting case analysis as highly useful, with a slight increase observed during the intervention that did not vary systematically with the number of tasks. Given the high initial values, a ceiling effect seems to be a plausible explanation. Students appear to acknowledge the relevance of this strategy for learning and doing physics, leaving little room for further growth. This finding is consistent with prior research showing that students generally recognize the value of such sensemaking strategies [16,18,20].

Overall, perceived usefulness of limiting case analysis appears to be insensitive to the instructional dose and seems to reflect a stable, general valuation of the strategy rather than a response to practice. The results suggest that students' perception of usefulness may be less affected by short-term interventions than, for example, their perception of competence.

#### C. Perceived competence

A similar pattern was found for students' perceived competence in applying limiting case analysis, with the data again compatible with a sigmoidal regression model. This is consistent with Vogt, who reported an analogous sigmoidal relationship for students' self-concept [25]. A medium effect size becomes apparent from four tasks onward, and a large effect is observed at the highest task count (eight tasks). As with perceived integration, students' perceived competence is also sensitive to

increases in practice, with meaningful gains emerging after only a small number of tasks.

Notably, group G1 shows negative effect sizes for both perceived competence and perceived integration, indicating a slight decline in these perceptions during the intervention period. This finding is not captured by a standard sigmoidal model that assumes a lower bound of zero [see Eq. (1)]. A likely explanation is a recalibration of students' self-assessment. As students engage with increasingly demanding content, they may become more aware of the limits of their current skills. For group G1, which received no intervention tasks, the absence of explicit practice with limiting case analysis may have further reinforced this awareness. It should also be noted that the items measuring perceived integration included the phrase "including this course" after the intervention, which may have additionally influenced students' responses. Importantly, the results highlight the value of explicitly integrating limiting case analysis into instruction to counteract the observed decline in students' confidence in applying the strategy.

#### D. Performance

Students initially performed poorly on limiting case analysis, further highlighting the demanding nature of this strategy [12,17,20]. During the intervention, the quality of students' explanations of the plausibility of an equation improved across all groups. Given that participants were incoming first-year students with limited prior experience in scientific reasoning, this overall improvement is likely attributable to multiple factors. The course introduced a wide range of concepts, and students engaged in intensive discussions during recitation sessions. In particular, these sessions are learning opportunities that may have fostered general reasoning skills independently of the specific intervention tasks.

The normalized Hake gain of students' performance increased from group G1 to G5, suggesting an intervention-sensitive learning gain. The gains observed in groups G3 and G5 exceed the threshold of  $g_H = 0.3$  separating low from medium normalized gains. This level is typically associated with engaging and innovative instructional approaches [41,42]. Gains for groups G1 and G2 remain closer to values reported for traditional instruction ( $g_H \approx 0.2$ ).

In contrast to our previous study [20], in which no clear intervention effect on performance was observed, the present study provides evidence for performance gains associated with increasing practice. It should be noted that the two studies differ substantially in terms of design, sample, and, notably, also in the density and distribution of intervention tasks (intervention intensity). The previous intervention study implemented fewer tasks within a longer time period, which may have been insufficient to produce detectable performance effects.

As for affective variables, students' performance is compatible with a sigmoidal model. With an increasing number of intervention tasks, students showed systematically higher performance gains. However, the data suggest that the observed groups for the target variables fall in the upper branch of the sigmoid, where the curve begins to flatten. Although group G5 completed twice as many tasks as group G3, both groups exhibit similarly large effect sizes and comparable learning gains, indicating a plateau at higher task counts. Additional practice therefore did not translate into additional performance gains. This saturation region is of particular theoretical and practical interest, as it marks the point beyond which further increases in practice yield only marginal improvements. Future work should examine this region more precisely.

At the same time, post-test performance levels remain moderate and do not exceed an average score of 56%, independent of the intervention group. Although students benefited from increased practice, the implemented doses did not yield high absolute performance scores. One possible explanation is that the short intervention period coincided with students adapting to a new university environment while simultaneously processing large amounts of new content, which may have contributed to cognitive overload. Additionally, we assumed the intervention tasks to be equally effective, even though they involved different physical contexts (see Table I). This assumption of context-independence may not fully hold, as sensemaking strategies rely on connecting physical meaning to mathematical expressions, a process that is inherently context-dependent. Whether competence developed in one context readily transfers to others therefore remains an open question.

The results suggest that increasing practice alone is not sufficient to gain very high performance scores in applying limiting case analysis. Given the conceptual complexity of the strategy [17], students may require more extensive practice. This points to the value of strongly guided instructional approaches for beginning physics students. For instance, structured reflection opportunities in recitation sessions and lectures, or recurrent engagement with worked examples that include explicit self-explanation phases to support active sensemaking [34].

As observed for affective variables, the dose-effect curve for performance does not originate at zero: Group G1, which completed no intervention tasks, already showed a small performance gain and a small effect size. At the same time, affective perceptions in this group decreased slightly. The positive learning gain in the absence of targeted practice indicates that performance improvements do not solely originate from the intervention but likely reflect general learning processes during the two-week course, e.g., increased familiarity with problem-solving formats and the overall development of scientific reasoning skills.

### E. Constraints and limitations

The present study provides initial evidence for a dose-dependent relationship between the number of tasks and students' perceptions and performance in limiting case analysis. As a proof-of-concept study, it demonstrates both the feasibility of dose-effect designs in authentic undergraduate settings and the empirical potential of such approaches. Nevertheless, the findings must be interpreted with caution in light of methodological and practical constraints inherent to the field setting.

A central limitation concerns the small number of intervention groups and the limited overall sample size, which was unevenly distributed across intervention conditions. Although students were randomly assigned, group sizes differed substantially, and one group had to be excluded from the dose-effect analyses due to insufficient participant numbers. As a consequence, only four data points were available for regression modeling, and the fitted dose-effect curves must therefore be regarded as exploratory rather than confirmatory. It should be noted that the present study does not seek to compare competing learning models. Rather, the sigmoidal model was selected on theoretical grounds, with the analyses intended to demonstrate the compatibility of the observed data with this type of learning curve. Future studies should employ a larger sample with a more fine-grained distribution of task numbers to examine dose-effect patterns more reliably and to better constrain the curve's trajectory.

The intervention was embedded in a natural and intensive two-week preparatory course, which constitutes both a strength and a limitation of the study. The authentic setting reflects the conditions under which such interventions would naturally be embedded. At the same time, the short duration limits conclusions about the persistence of observed effects. Future studies should examine whether the dose-effect relationships observed here replicate over longer instructional periods.

A further limitation concerns the intensity of the intervention (number of tasks per time). The intervention spanned two weeks, and it remains an open question whether the same number of tasks spread over a longer period would produce comparable effects.

### V. CONCLUSION AND OUTLOOK

Limiting case analysis is a central component of expert-like reasoning in physics, as it supports the meaningful connection between mathematical expressions and their underlying physical context. The present results reinforce the finding that this strategy is typically underrepresented in early undergraduate physics instruction, and demonstrate that explicitly embedding limiting case analysis into courses can meaningfully enhance students' sensemaking skills and confidence even within a short intervention period.

This study contributes to research on sensemaking in physics by systematically examining how the amount of practice affects students' learning outcomes in limiting case analysis. To the best of our knowledge, the present study is the first to examine dose-effect relationships in the context of sensemaking of physics equations in undergraduate physics education. Beyond the empirical findings, this study demonstrates the feasibility of implementing dose-effect designs in authentic undergraduate settings. The systematic variation of the number of intervention tasks embedded in regular recitation sessions could serve as a methodological blueprint for future research in physics education.

Overall, the findings indicate that students perceived the intended differences in instructional dose, and that increasing the number of tasks was associated with greater perceived competence and improved performance. The observed effect sizes suggest that even a moderate number of tasks can produce meaningful improvements in students' confidence. Although performance gains generally increased with the number of tasks, students' absolute post-test performance remained at a moderate level, suggesting that practice volume alone is not sufficient to achieve high performance scores. More intensive instructional support may be necessary to complement practice and enable more robust skill development.

Future research should investigate such dose-effect relationships using a finer-grained and broader range of intervention doses, in order to identify threshold effects and saturation regions with greater precision and resolution. Beyond limiting case analysis, extending dose-effect frameworks to other sensemaking strategies, such as the consideration of covariations or special cases, would provide a more comprehensive understanding of how practice supports the development of expertlike sensemaking in undergraduate physics. Finally, follow-up studies are needed to examine whether the gains observed here persist beyond the intervention period and transfer to more complex contexts and mathematical expressions.

### VI. ETHICAL STATEMENT

Participation was voluntary, and data were collected in fully anonymous form. To allow matching across the two measurement time points, participants created self-generated pseudonyms, preventing any identification of individual respondents.

### ACKNOWLEDGMENTS

We want to thank all students who took part in the study and supported our research. We gratefully acknowledge the work of our student assistants, especially Johannes Behling, who provided invaluable support with the data curation and interrater processes. We also gratefully acknowledge the support of the *Wilfried and Ingrid*

*Kuhn Foundation.* This article was partially funded through the Open Access Publication Funds of the University of Göttingen.

J. H. contributed to the conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft, writing—review and editing, and visualization. P. K. contributed to the conceptualization, methodology, investigation, writing—review and editing, supervision, project administration, and funding

acquisition. A. M. contributed to the conceptualization, methodology, writing—review and editing, and supervision. J. N. contributed to the conceptualization, methodology, investigation, and writing—review and editing.

### DATA AVAILABILITY

There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

### APPENDIX: APPROXIMATION

The exponential function is an approximation for the upper branch of the logistic function, i.e., for higher dosage. To quantify this approximation, we investigate the relative deviation  $\delta$  between the two models, where  $d(x)$  denotes the logistic model and  $d_{\text{exp}}(x)$  the exponential model:

$$\delta = \left| \frac{d_{\text{exp}}(x) - d(x)}{d_{\text{exp}}(x)} \right| = \left| \frac{b(1 - e^{-ax}) - \frac{b}{1+e^{-ax}}}{b(1 - e^{-ax})} \right| = \left| \frac{(1 - e^{-ax})(1 + e^{-ax}) - 1}{(1 - e^{-ax})(1 + e^{-ax})} \right| = \left| \frac{-e^{-2ax}}{1 - e^{-2ax}} \right| = \left| \frac{1}{1 - e^{2ax}} \right|.$$

In analogy to the sensemaking strategy discussed in this paper, we consider the limiting case  $x \rightarrow \infty$ : Since  $e^{2ax} \rightarrow \infty$ , it follows that  $\delta \rightarrow 0$ . More precisely, for large  $x$  the deviation decays as  $\delta \approx e^{-2ax}$ , i.e., exponentially with a decay constant of  $2a$ ; twice the decay constant of the exponential model itself. Looking at  $x = 1/a$  for this decay, the relative deviation is  $\delta \approx e^{-2} = 0.14$ . For  $x > 1/a$ , the relative deviation falls below 0.1, so that the exponential model provides a good approximation of the logistic model beyond this point.

This shows that the relative deviation between both models decreases as  $x$  increases, and that the exponential model becomes an increasingly accurate approximation of the logistic model at higher dose values.

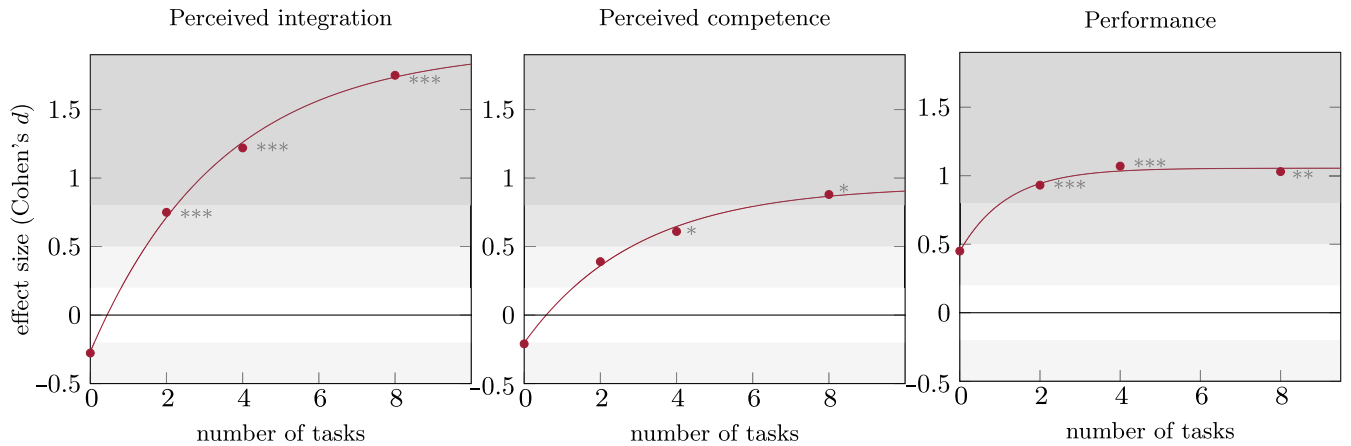


FIG. 8. Effect sizes (Cohen's  $d$ ) as a function of the number of limiting case analysis tasks for perceived integration in previous and current courses (left), perceived competence (center), and performance (right). Points represent observed effects, and solid lines indicate fitted exponential regression models illustrating the dose-effect relationship. Data points for group G4 were excluded from the sigmoidal model (see text). Shaded regions denote small ( $0.2 < |d| < 0.5$ ), medium ( $0.5 < |d| < 0.8$ ), and large ( $|d| > 0.8$ ) effect sizes according to Cohen [38]. Significance levels are indicated as \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$ .

TABLE V. A detailed overview of the equations and limiting cases to be addressed in each intervention task throughout the eight exercise sheets. For an overview of the topics in each lecture, see Table I.

| Sheet no. | Equation                                                                                                                | Limiting case            |
|-----------|-------------------------------------------------------------------------------------------------------------------------|--------------------------|
| 1         | Hagen-Poiseuille<br>$\Delta P = \frac{\Phi L \eta}{C \times R^4}$                                                       | $R \rightarrow 0$        |
| 2         | Logistic growth<br>$N(t) = \frac{S}{1 + (S-1)e^{-St}}$                                                                  | $t \rightarrow \infty$   |
| 3         | Reflection of light<br>$I_R = \left(\frac{n_2 - n_1}{n_2 + n_1}\right)^2$                                               | $n_2 \rightarrow \infty$ |
| 4         | Velocity of free fall<br>$v(t) = g\tau + (v_0 - g\tau)e^{-t/\tau}$                                                      | $t \rightarrow \infty$   |
| 5         | Lennard-Jones potential<br>$F(x) = 4\epsilon[-12\left(\frac{\sigma}{x}\right)^{12} + 6\left(\frac{\sigma}{x}\right)^6]$ | $x \rightarrow \infty$   |
| 6         | Gravitational force<br>$ \vec{F}  = \frac{GmM}{r^2}$                                                                    | $r \rightarrow \infty$   |
| 7         | Escape velocity of rocket<br>$v_f = \sqrt{\frac{2GM}{r}}$                                                               | $r \rightarrow \infty$   |
| 8         | Damped harmonic oscillator<br>$e^{-\delta t}[2u \cos(\omega t) - 2v \sin(\omega t)]$                                    | $t \rightarrow \infty$   |

TABLE VI. Items of the questionnaire regarding the analysis of limiting cases with item-total correlation  $r_{it}$  and internal consistency (Cronbach's  $\alpha$  [40]). The analyses were conducted using the initial sample  $N_0 = 127$ . The items of the perceived competence and usefulness were derived from the Intrinsic Motivation Inventory [35].

| Construct             | Item                                                                                         | $r_{it}$ |
|-----------------------|----------------------------------------------------------------------------------------------|----------|
| Usefulness            | I believe that mastering the strategy is important.                                          | 0.73     |
|                       | I believe that using the strategy has a practical benefit for me.                            | 0.67     |
|                       | I believe that mastering the strategy brings a deeper physical insight.                      | 0.58     |
|                       | I will continue to use the strategy in the future because it has a practical benefit for me. | 0.66     |
|                       | <b>Cronbach's <math>\alpha</math></b>                                                        | 0.83     |
| Perceived integration | The strategy was explicitly subject of previous courses. <sup>a</sup>                        | 0.83     |
|                       | I was able to practice applying the strategy in previous courses. <sup>a</sup>               | 0.83     |
|                       | The strategy was explicitly included in exercise sheets from previous courses. <sup>a</sup>  | 0.76     |
|                       | <b>Cronbach's <math>\alpha</math></b>                                                        | 0.90     |
| Perceived competence  | I believe that I am not making any mistakes when applying the strategy.                      | 0.78     |
|                       | I am good at applying the strategy.                                                          | 0.84     |
|                       | I feel confident using the strategy.                                                         | 0.83     |
|                       | I know what to do when applying the strategy.                                                | 0.69     |
|                       | <b>Cronbach's <math>\alpha</math></b>                                                        | 0.90     |

<sup>a</sup>Additionally in the post-test (T1): "including this course."

TABLE VII. Coding scheme of the evaluation of the application of LCA and the subsequent explanation of the physical plausibility of the equation. The categories were developed based on the rubrics of the scientific abilities of Etkina *et al.* [43].

| Category    | Description                                                                                  | 0                                                                             | 1                                                                                                                     | 2                                                                                     | 3                                                        |
|-------------|----------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|----------------------------------------------------------|
| 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. <sup>a</sup> | A limiting case analysis was carried out without errors. |
| Explanation | The person can give an explanation to determine whether an equation is physically plausible. | Empty/No attempt was made to provide an explanation.                          | A conclusion about the physical plausibility of the equation was made, but no explanation was provided.               | An explanation was provided, but it is flawed.                                        | An explanation was provided without errors.              |

<sup>a</sup>Flaws: algebraic, mathematical, physical, and notational.

TABLE VIII. Descriptive statistics (mean and standard deviation) of all variables regarding limiting case analysis for all intervention groups before (T0) and after the intervention (T1). See Table III for corresponding group sizes. Varying intervention in terms of number of tasks: 0 tasks (G1), 2 tasks (G2), 4 tasks (G3), 6 tasks (G4), and 8 tasks (G5).

|                       | G1              | G2              | G3              | G4              | G5              |
|-----------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Usefulness            |                 |                 |                 |                 |                 |
| T0                    | $0.83 \pm 0.20$ | $0.79 \pm 0.18$ | $0.73 \pm 0.22$ | $0.87 \pm 0.20$ | $0.76 \pm 0.25$ |
| T1                    | $0.87 \pm 0.19$ | $0.85 \pm 0.19$ | $0.87 \pm 0.18$ | $0.89 \pm 0.15$ | $0.87 \pm 0.13$ |
| Perceived integration |                 |                 |                 |                 |                 |
| T0                    | $0.55 \pm 0.34$ | $0.42 \pm 0.29$ | $0.41 \pm 0.29$ | $0.27 \pm 0.23$ | $0.40 \pm 0.24$ |
| T1                    | $0.44 \pm 0.37$ | $0.69 \pm 0.29$ | $0.87 \pm 0.20$ | $0.95 \pm 0.11$ | $0.93 \pm 0.14$ |
| Perceived competence  |                 |                 |                 |                 |                 |
| T0                    | $0.64 \pm 0.29$ | $0.55 \pm 0.24$ | $0.52 \pm 0.26$ | $0.50 \pm 0.29$ | $0.57 \pm 0.28$ |
| T1                    | $0.60 \pm 0.30$ | $0.65 \pm 0.24$ | $0.70 \pm 0.21$ | $0.76 \pm 0.17$ | $0.78 \pm 0.20$ |
| Performance LCA       |                 |                 |                 |                 |                 |
| T0                    | $0.22 \pm 0.20$ | $0.16 \pm 0.15$ | $0.15 \pm 0.18$ | $0.11 \pm 0.12$ | $0.18 \pm 0.15$ |
| T1                    | $0.33 \pm 0.24$ | $0.35 \pm 0.24$ | $0.41 \pm 0.26$ | $0.56 \pm 0.30$ | $0.44 \pm 0.25$ |
| Explanation           |                 |                 |                 |                 |                 |
| T0                    | $0.28 \pm 0.18$ | $0.29 \pm 0.22$ | $0.22 \pm 0.17$ | $0.19 \pm 0.18$ | $0.31 \pm 0.30$ |
| T1                    | $0.44 \pm 0.24$ | $0.46 \pm 0.22$ | $0.35 \pm 0.15$ | $0.38 \pm 0.27$ | $0.44 \pm 0.27$ |

TABLE IX. Sigmoid regression of students' perceived integration, competence, and performance of limiting case analysis. The offset parameter was introduced for perceived integration and competence due to a negative effect observed in group G1 (0 tasks).

| Dependent variable    | Sigmoid regression                                      | Goodness of fit (Pseudo- $R^2$ ) |
|-----------------------|---------------------------------------------------------|----------------------------------|
| Perceived integration | $d(x) = \frac{3.03}{1+e^{-\frac{x-1.07}{1.66}}} - 1.28$ | 0.9904                           |
| Perceived competence  | $d(x) = \frac{2.09}{1+e^{-\frac{x-0.10}{1.80}}} - 1.21$ | 0.9901                           |
| Performance           | $d(x) = \frac{1.05}{1+e^{-\frac{x-0.25}{0.84}}}$        | 0.9943                           |

TABLE X. Exponential regression of students' perceived integration, competence, and performance of limiting case analysis.

| Dependent variable    | Exponential regression                       | Goodness of fit (Pseudo- $R^2$ ) |
|-----------------------|----------------------------------------------|----------------------------------|
| Perceived integration | $d(x) = 1.95 (1 - e^{-\frac{x-0.44}{3.42}})$ | 0.9985                           |
| Perceived competence  | $d(x) = 0.94 (1 - e^{-\frac{x-0.58}{2.97}})$ | 0.9962                           |
| Performance           | $d(x) = 1.06 (1 - e^{-\frac{x+0.66}{1.19}})$ | 0.9917                           |

- |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>[1] K. A. Ericsson, R. T. Krampe, and C. Tesch-Römer, The role of deliberate practice in the acquisition of expert performance, <i>Psychol. Rev.</i> <b>100</b>, 363 (1993).</p> <p>[2] E. M. Smith and N. G. Holmes, Evaluating instructional labs' use of deliberate practice to teach critical thinking skills, <i>Phys. Rev. Phys. Educ. Res.</i> <b>16</b>, 020150 (2020).</p> <p>[3] C. Wieman, How to become a successful physicist, <i>Phys. Today</i> <b>75</b>, 46 (2022).</p> <p>[4] A. Müller and P. Vogt, Context-based science education by advertisement story problems: A dose-effect relationship</p> | <p>for motivation, in <i>Science education research for evidence-based teaching and coherence in learning: Proceedings of the ESERA 2013 conference. Part 2</i>, edited by C. P. Constantinou, N. Papadouris, and A. Hadjigeorgiou (European Science Education Research Association, 2014).</p> <p>[5] J. M. J. Murre, S-shaped learning curves, <i>Psychon. Bull. Rev.</i> <b>21</b>, 344 (2014).</p> <p>[6] M. V. Pusic, K. Boutis, M. R. Pecaric, O. Savenkov, J. W. Beckstead, and M. Y. Jaber, A primer on the statistical</p> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

- modelling of learning curves in health professions education, *Adv. Health Sci. Educ.* **22**, 741 (2017).
- [7] G. J. Duhon, B. C. Poncy, C. F. Krawiec, R. E. Davis, N. Ellis-Hervey, and C. H. Skinner, Toward a more comprehensive evaluation of interventions: A dose-response curve analysis of an explicit timing intervention, *Sch. Psychol. Rev.* **51**, 84 (2022).
- [8] A. H. Miller, D. R. Espinas, D. McNeish, and M. A. Barnes, Dosage response in intensive mathematics interventions for early elementary students with or at-risk for mathematics learning disability, *Educ. Psychol. Rev.* **37**, 91 (2025).
- [9] T. O. B. Odden and R. S. Russ, Defining sensemaking: Bringing clarity to a fragmented theoretical construct, *Sci. Educ.* **103**, 187 (2019).
- [10] 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.
- [11] E. F. Redish and E. Kuo, Language of physics, language of math: Disciplinary culture and dynamic epistemology, *Sci. Educ.* **24**, 561 (2015).
- [12] 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).
- [13] A. R. Akinyemi, An investigation of students' use and understanding of evaluation strategies, Ph.D. thesis, University of Maine, 2021.
- [14] 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).
- [15] E. F. Redish, Using math in physics: 2. Estimation, *Phys. Teach.* **59**, 525 (2021).
- [16] M. Lenz, K. T. Hahn, P. J. Emigh, and E. Gire, Student perspective of and experience with sense-making: A case study, in *2017 Physics Education Research Conference Proceedings* (2018), pp. 240–243.
- [17] N. J. Nersessian, How do scientists think? Capturing the dynamics of conceptual change in science, in *Cognitive Models of Science*, edited by R. N. Giere (University of Minnesota Press, Minneapolis, 1992), pp. 3–44.
- [18] A. R. Warren, Impact of teaching students to use evaluation strategies, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020103 (2010).
- [19] K. T. Hahn, P. J. Emigh, M. Lenz, and E. Gire, Student sense-making on homework in a sophomore mechanics course, in *2017 Physics Education Research Conference Proceedings* (2018), pp. 160–163.
- [20] J. Hofmann, P. Klein, A. Müller, and J. Neuhaus, Supporting sensemaking of physics equations: An intervention study in undergraduate courses, *Phys. Rev. Phys. Educ. Res.* **22**, 010144 (2026).
- [21] 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. Proceed.* **1513**, 418 (2013).
- [22] E. Kim and S.-J. Pak, Students do not overcome conceptual difficulties after solving 1000 traditional problems, *Am. J. Phys.* **70**, 759 (2002).
- [23] W. K. Black, R. L. Matz, M. Mills, and A. E. Evrard, Quantifying grade benefits of study: Practice makes better, *Phys. Rev. Phys. Educ. Res.* **19**, 010139 (2023).
- [24] A. E. Evrard, M. Mills, D. Winn, K. Jones, J. Tritz, and T. A. McKay, Problem roulette: Studying introductory physics in the cloud, *Am. J. Phys.* **83**, 76 (2015).
- [25] P. Vogt, Werbeaufgaben im physikunterricht: Motivations- und lernwirksamkeit authentischer texte, Ph.D. thesis, Vieweg+Teubner, Wiesbaden, 2010.
- [26] A. Grebenc, Foundations for a general theory of sigmoid functions: Modelling with 30 methods, *Intelligent Systems Conference* (Springer, New York, 2025), pp. 481–511.
- [27] K. T. Kotsis and D. Panagou, Using alternative ideas for determining the learning curve on the concept of force, *Eur. J. Sci. Math. Educ.* **10**, 495 (2022).
- [28] K. T. Kotsis and D. Panagou, The determination of the learning curve on the concept of energy using the alternatives ideas, *Contemp. Math. Sci. Educ.* **4**, ep23011 (2023).
- [29] A. Heathcote, S. Brown, and D. J. K. Mewhort, The power law repealed: The case for an exponential law of practice, *Psychon. Bull. Rev.* **7**, 185 (2000).
- [30] D. E. Pritchard, Y.-J. Lee, and L. Bao, Mathematical learning models that depend on prior knowledge and instructional strategies, *Phys. Rev. ST Phys. Educ. Res.* **4**, 010109 (2008).
- [31] Kultusministerkonferenz, Abiturnoten im Ländervergleich, <https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/abiturnoten.html> (2024), accessed: 08-04-2026.
- [32] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/gx8s-rldd> for questionnaire, intervention tasks, and performance tests (2026).
- [33] 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).
- [34] A. Renkl, Toward an instructionally oriented theory of example-based learning, *Cogn. Sci.* **38**, 1 (2014).
- [35] Center for Self-Determination Theory, Intrinsic Motivation Inventory (IMI), <https://selfdeterminationtheory.org/intrinsic-motivation-inventory/>, accessed: 17-09-2026.
- [36] R. M. Ryan, Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory, *J. Pers. Soc. Psychol.* **43**, 450 (1982).
- [37] L. Ding and R. Beichner, Approaches to data analysis of multiple-choice questions, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020103 (2009).
- [38] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (L. Erlbaum Associates, Hillsdale, N.J., 1988).
- [39] 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, *Anales Psicol.* **40**, 466 (2024).
- [40] L. J. Cronbach, Coefficient alpha and the internal structure of tests, *Psychometrika* **16**, 297 (1951).
- [41] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics

- test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
- [42] V. P. Coletta, J. A. Phillips, and J. J. Steinert, Interpreting force concept inventory scores: Normalized gain and SAT scores, *Phys. Rev. ST Phys. Educ. Res.* **3**, 010106 (2007).
- [43] 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).
- [44] M. Wirtz and F. Caspar, *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen* (Hogrefe, Göttingen, 2002).
- [45] 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).