



Erik Teichmann | Heather J. Lewandowski | Micol Alemani

# Investigating students' views of experimental physics in German laboratory classes

**Suggested citation referring to the original publication:**

Physical Review Physics Education Research 18 (2022), pp. 010135-1 - 010135-17

DOI <https://doi.org/10.1103/PhysRevPhysEducRes.18.010135>

ISSN 1554-9178

**Journal article | Version of record**

Secondary publication archived on the Publication Server of the University of Potsdam:

Zweitveröffentlichungen der Universität Potsdam : Mathematisch-Naturwissenschaftliche Reihe 1263

ISSN: 1866-8372

<https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-560429>

DOI: <https://doi.org/10.25932/publishup-56042>

**Terms of use:**

This work is licensed under a Creative Commons License. This does not apply to quoted content from other authors. To view a copy of this license visit

<https://creativecommons.org/licenses/by/4.0/>.



## Investigating students' views of experimental physics in German laboratory classes

E. Teichmann,<sup>1</sup> H. J. Lewandowski<sup>2,3</sup> and M. Alemani<sup>1,\*</sup>

<sup>1</sup>*Universität Potsdam, Institut für Physik und Astronomie, Potsdam 14476, Deutschland*

<sup>2</sup>*Department of Physics, University of Colorado, Boulder, Colorado 80309, USA*

<sup>3</sup>*JILA, National Institute of Standards and Technology and University of Colorado, Boulder, Colorado 80309, USA*



(Received 28 January 2022; accepted 16 March 2022; published 27 April 2022)

There is a large variety of goals instructors have for laboratory courses, with different courses focusing on different subsets of goals. An often implicit, but crucial, goal is to develop students' attitudes, views, and expectations about experimental physics to align with practicing experimental physicists. The assessment of laboratory courses upon this one dimension of learning has been intensively studied in U.S. institutions using the Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS). However, there is no such an instrument available to use in Germany, and the influence of laboratory courses on students views about the nature of experimental physics is still unexplored at German-speaking institutions. Motivated by the lack of an assessment tool to investigate this goal in laboratory courses at German-speaking institutions, we present a translated version of the E-CLASS adapted to the context at German-speaking institutions. We call the German version of the E-CLASS, the GE-CLASS. We describe the translation process and the creation of an automated web-based system for instructors to assess their laboratory courses. We also present first results using GE-CLASS obtained at the University of Potsdam. A first comparison between E-CLASS and GE-CLASS results shows clear differences between University of Potsdam and U.S. students' views and beliefs about experimental physics.

DOI: [10.1103/PhysRevPhysEducRes.18.010135](https://doi.org/10.1103/PhysRevPhysEducRes.18.010135)

### I. INTRODUCTION

Laboratory courses are an important part of the physics education curriculum, as they offer the opportunity for students to engage with, and become proficient at, the processes of experimental physics [1]. There is a large variety of teaching goals pursued in laboratory courses. Depending on the institution, one finds courses focused on reinforcing physics concepts, the acquisition of experimental skills, or both of these aspects [2–5]. In addition to these broad categories of goals, an often implicit, but crucial, goal of laboratory classes, is to have students' views and expectations of experimental physics align with those of practicing experimental physicists.

Studies of the degree to which students meet these varied goals have been conducted only recently. These studies have opened up a discussion about the effectiveness of traditional prescriptive laboratory courses to improve students' physics content knowledge [6,7], experimental skills [8–10], and the understanding of the nature of experimental

physics [11–13]. Moreover, it has been shown that having students understand the nature of experimental physics and promoting “expertlike” attitudes can enhance student motivation and performance in the laboratory [14–17]. New teaching approaches have been developed, shifting the focus from physics content reinforcement to the acquisition of experimental skills, while engaging students in authentic experimental physics practices [18–24]. In order to gain insight into students' learning in laboratory classes and evaluate the obtainment of the learning goals, instructors need ways to assess their laboratory classes in an easy and standardized way. The data from these assessments can then be used to guide course transformations [11]. Several research-based assessment tools for laboratory courses have been developed by the physics education research (PER) and are given in Refs. [12,25–30]. One important example is the Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS). It is a research-based survey focused on the nature of experimental physics. It was developed and validated at the University of Colorado Boulder and has been used extensively in the U.S. to assess students views and expectations about experimental physics [12,13,31].

In contrast, one can find only a few research-based surveys for assessing laboratory courses in German-speaking countries [28–30] and none of these tools focus on students' views of the nature of experimental physics.

\*alemani@uni-potsdam.de

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

Some recent studies concern students' views about the nature of science in general, but none of those studies focused on experimental physics [32–34]. The E-CLASS and its current automated administration system [35] is not able to be used in German-speaking countries because of the language barrier and the strict European data-protection laws present in the European Union.

Motivated by the lack of assessment tools and studies about students' views about experimental physics in German laboratory courses, we adapted the E-CLASS to the German context. We call the resulting German version of the E-CLASS the GE-CLASS (where G stands here for German).

The main goals of this work are to present the process of this adaptation, the resulting survey, and some initial results from the GE-CLASS from one institution, including a comparison to the U.S. results of E-CLASS. We hope this will motivate German instructors broadly to begin assessing this one dimension of learning in their laboratory classes using the GE-CLASS. Thus, instructors in German-speaking institutions will have data on student learning to help inform their instructional efforts and laboratory transformations. Moreover, by having many instructors use the GE-CLASS, we will gain insight into strengths and weaknesses of laboratory instruction in German-speaking institutions, which could allow for coordinated efforts to improve laboratory instruction. To this end, we present in this paper (i) how we translated the E-CLASS into German and validated the translation through student interviews, (ii) how we created a centralized survey administrations system for broad dissemination of the GE-CLASS in German-speaking institutions taking into account the constraints of privacy laws in Europe, and (iii) the results from a first exemplary study using the GE-CLASS system at the University of Potsdam.

## II. THE E-CLASS

The E-CLASS consists of 30 core statements, which probe ideas around the nature of experimental physics. Examples include the following: *The primary purpose of doing a physics experiment is to confirm previously known results*; *When I am doing an experiment, I try to make predictions to see if my results are reasonable*. For each core statement, students answer two types of questions. One is related to their personal view (*What do YOU think when doing experiments for class?* referred to as YOU questions). The other one is related to their perspective of experts views (*What would experimental physicists say about their research?* referred as EXPERT questions). Students answer on a five-point Likert scale from strongly disagree to strongly agree. Students' answers are compared to the answers given by 23 practicing physicists, referred to here as the EXPERT reference (ER). Since students complete the E-CLASS before and after the laboratory courses in a pre- or post-test mode, one can study if

students personal, as well as their perspective of experts, shift upon instruction to more expertlike or not (i.e., if they agree with the ER or not). An additional set of 23 questions probe students' postsurvey perception of the grading practice in the laboratory course. For example, students are asked *How important for earning a good grade in this class was calculating uncertainties to better understand my results?* Instructors can therefore verify if their grading practice is perceived by students in the way they intended.

The E-CLASS probes a wide variety of aspects of experimentation, for example, the importance of measurement uncertainties and systematic errors, self-confidence, and the relevance of experiments in physics as a discipline. This large variety of aspects makes the E-CLASS applicable in diverse instructional laboratory environments with different underlying goals, as instructors are provided with aggregate student responses to each individual statement and are encouraged to focus on only the statements that align with their particular goals for the class. Many research studies have been performed using the E-CLASS within recent years [12,13,35–38], with 95% of the E-CLASS dataset coming from U.S. institutions. The entire deidentified dataset has been recently made publicly available [39]. Some of the main results of those studies are the following:

1. Even when students have novice personal views about experimental physics, they know how an expert would respond to the statements [12];
2. In traditional, guided, first-year laboratory courses, a decrease in the mean overall E-CLASS score from the pretest to the post-test is observed [13];
3. Courses that aim to primarily improve experimental skills score higher on the E-CLASS than courses that aim to primarily reinforce concepts [37];
4. Partially open-ended labs show higher E-CLASS scores than fully guided laboratory courses [36]; and
5. Perception of the grading scheme is correlated with E-CLASS performance [40].

A more complete review of the results from a large selection of studies done with the E-CLASS can be found in Ref. [13].

## III. TRANSLATION OF THE E-CLASS TO GERMAN

In order to adapt the E-CLASS to the German-speaking context, we performed a multistep iterative translation accompanied by think-aloud student interviews. The process started by having two German-native-speaking experimental physicists independently translate the E-CLASS into German. They then discussed the discrepancies until they reached 100% consensus. This German translation was translated back to English by an English-native-speaking physicist. These three translators are not authors of this paper. The back translation was then compared to the original E-CLASS. The discrepancies between the two versions were analyzed and discussed by the translation

team and one of the authors (M. A.) until a consensus was achieved. The resulting German translation was validated through think-aloud interviews with students. A total of 12 think-aloud interviews with physics major students were conducted by one of the authors. Among the students interviewed, four were women, eight were men; six were first-year students and six were second-year students. All of the students volunteered to take part in the interview process by responding to an announcement after laboratory classes at the University of Potsdam. In the interviews, students were asked to read in order the items of the survey and for each item explain their understanding of that item. We did not interrupt the students unless it was not completely clear to us what students meant. In this case, we asked for clarification or to give an example of what they meant. The interviews were audio recorded for the data analysis. M. A. and H. L. (original E-CLASS survey developer) discussed the results of the translations and interviews to assure that students' interpretations of the survey items had the same meaning as the original English version. In case of discrepancies between the German and English versions, we refined the translation together with the translators before further interviews were performed. We note that H. L. is a native English speaker with only a basic level knowledge of German. M. A. is not a native German speaker, but is fluent, and has taught laboratory courses, in both German and English. After the last set of changes, students were interpreting the statements consistently and as we intended. A complete list of the translated items on GE-CLASS can be found in the Appendix.

In the following, we report some examples of the minor changes we made during the entire iterative translation process to obtain the final German survey version. Those changes arose for different types of reasons. In some cases, we found mistakes or not literal translations. For example, for question 8 "*When doing an experiment, I try to understand the relevant equations*" the word "equations" was first incorrectly translated as "questions" and for question 4 "*If I am communicating results from an experiment, my main goal is to create a report with the correct sections and formatting*" the word "sections" was first translated as "outline" (German "*Gliederung*") instead of sections (German "*Abschnitten*"). For some items, the two German translators used very slightly different translation options. In such cases, we decided on one of the two versions, but we kept the other version ready during the interviews in case students encountered difficulties with the chosen version. A third reason for changes was when students had comprehension problems. For the majority of the survey items, students were able to interpret as intended the German version during the interviews. However, difficulties arose with items Q14, Q15, Q22, Q24, and Q26. Below, we explain the issues with those items and any resulting change:

**Q14:** (*When doing an experiment I usually think up my own questions to investigate.*) In the first interviews, this question was misinterpreted and the word questions was interpreted as "a comprehension question students have when they read the lab manuals." We had to change the wording to clarify (by using the word "erforschen") that the questions are open (research) questions one can answer by doing an experiment.

**Q15:** (*Designing and building things is an important part of doing physics experiments.*) Students were sometimes interpreting our translation as meaning "essential part" instead of "important part." We changed the translation of "important" from "wesentlicher" into to "wichtiger" to address this slight difference in meaning.

**Q22:** (*If I am communicating results from an experiment, my main goal is to make conclusions based on my data using scientific reasoning.*) This item was often hard to understand for students initially without extra time for reflection. To help students understand this item easily, we created three slightly different versions of the statement. We asked students (starting from the second interview) each one in turn and had them discuss which one was easiest for them to understand. We kept the version with the highest rating (we had 10 students out of 11 choosing the final version).

**Q24:** (*Nearly all students are capable of doing a physics experiment if they work at it.*) Two interviewees out of 12 were unsure if this statement was meant to check if the lab manuals or equipment in the laboratory were good enough so that students could succeed in the experiment or (as actually intended by the item) if the statement was referring to students' skills in experimentation. The other 10 students understood this sentence immediately as intended as if students have the skills to perform physics experiments. We kept the translation as it was, as the other meaning understood by those two students cannot be ruled out even in the English version of this item and both students mentioned that they were unsure about the meaning of the item but reported both possibilities.

**Q26:** (*It is helpful to understand the assumptions that go into making predictions.*) similarly to Q22, many interviewees needed to think longer on this item, but all understood it as intended. The difficulty in the comprehension was due to the fact that students had to think about the meaning of the words "assumption" and "prediction." To facilitate the understanding of the word assumption (which was translated with "Annahmen") we added the adjective "assumed" ("vorausgesetzte") to it. The rest of the items were easily interpreted as intended by all students interviewed and were not changed for the final version shown in the Appendix.

#### IV. CENTRALIZED AUTOMATED SYSTEM

We have created an online, centrally administrated and automated system for instructors to use the GE-CLASS,



with the goal of encouraging and facilitating its use in German-speaking institutions. Centralized, automated assessments systems have two main advantages over paper and pencil assessments [35,41]. First, automated systems reduce the effort required of instructors using the assessments in their courses by collecting the data and performing the analysis. Second, centrally administered systems can provide additional information to instructors, i.e., the comparison between their course results and the results from similar courses at other institutions [13]. With this in mind, we have created a web-page front end for instructors to register their courses with a back end in which an automated data analysis system is integrated. The link to the webpage can be found in Ref. [42]. The web server, as well as the analysis for the report, are programmed in PYTHON3 [43] using flask (for the web server) and the SciPy [44–46] stack for the analysis. When setting up a new course, instructors are asked to answer a series of questions about their classes and instruction methods. We collect information about the course, such as type of institution, how many weeks students work in the laboratory, how many different experiments they perform, how many projects they do, goals of laboratory classes, and what kind of activities students do in the laboratory course.

Once a course is registered, the automated system assigns an identification code to the course that instructors have to provide to their students together with the web link for the student survey. It is important to note here that the way we constructed the GE-CLASS system was discussed beforehand with the ethics commission and the data-protection office of the University of Potsdam in order to respect the ethic codes of the University of Potsdam and the European Union General Data Protection Regulation (EU-GDPR). One result of this process was the obligation to assure student anonymity. When taking the survey, students have to create a self-generated anonymous code (composed by 3 letters and 2 numbers) by answering to 3 questions:

1. Last two letters of the name of your father (**Harald**);
2. The second letter of the place in which you were born (**Berlin**); and
3. The day of your mothers birthday (**26.12.1978**).

The code for the example above would be **lde26**.

For each student, we assign two identification codes. One ID of the course and one (anonymous) for the person.

We often found that students codes in the pre- and post-survey differed by 1 number or letter, suggesting that some students produced an erroneous code either in the pre- or postsurvey. Those data are unfortunately lost because they cannot be matched and used for the data analysis (see Sec. V C). We are now thinking about ways to avoid this problem, by either changing the questions to generate the code or suggesting students to write their code on their laboratory notebooks to remember. We also note that the questions asked above have to be considered in a cultural

context of Germany. We acknowledge that these questions would not necessarily be appropriate in a U.S. context.

Two weeks after the post-test start date, instructors can log on the GE-CLASS web page and find the analysis of their course data in form of a report. In the report, an overview of the results of the course is provided at the beginning. For example, one can find in the overview how many students did the pre- and postsurvey, how many students' surveys were considered valid for the data analysis (see Sec. V C) and how many surveys were used to provide a comparison with similar courses. The first graph of the report shows the overall agreement with experts averaged over all items and all students. Note that in the report, we do not provide a comparison with U.S. results, but for each graph we provide the comparison with the dataset of similar courses that we collected in the GE-CLASS system. In the second and third graphs, one can see the agreement with experts for each item for the YOU questions and EXPERT questions, respectively. In the fourth graph, we report for each item the comparison between results of the YOU and EXPERT questions. In the last graph, the results about students' grading perception for each survey item are shown. An example report can be found in Ref. [47].

## V. FIRST STUDY USING THE GE-CLASS SYSTEM

Using the above described GE-CLASS system, we have obtained the first results studying the physics laboratory classes at the University of Potsdam. The data have been collected for three consecutive semesters: Winter 2019, Summer 2020, and Winter 2021.

### A. Course contexts

We collected data from nine different instances of laboratory courses (see Table I). All of the instances of courses are part of the introductory physics laboratory course sequence for physics major students, either studying in the bachelor of science program (B.Sc.) or in the bachelor of education program (B.Ed.). Physics students at University of Potsdam must take four courses belonging to the introductory laboratory course sequence during the first four semesters, therefore during their first and second year of the university. We have indicated in the Table I which course corresponds to which semester. For each semester course, we have a set of different learning goals and have structured the courses differently. Note that in each semester, courses for physics students in the B.Sc. and B.Ed. have the same goals and content, but B.Ed. students have fewer weeks of labs. We refer to courses with the same goals and structure being a type of laboratory course. Details about the learning goals and structure of each type of laboratory course can be found below. We have also reported in Table I in which modality the courses were conducted (online, in-person, or hybrid) and have indicated

TABLE I. Description of the nine instances of the courses from which data have been obtained. We distinguish between courses offered for students in the first year and beyond the first year. We indicated the semester in which students take the courses, the study program [Bachelor of Science (B.Sc.) and Bachelor of Education (B.Ed.)], and course goals and structure (referred as course type). A detailed description of the course types can be found in the text. In the table, we also show, for each course, how many weeks the laboratory course lasted, the course modality (in-person, online, hybrid) and give details about how many GE-CLASS valid responses have been obtained for each course and the response rate.

	Year	Semester	Course type	Study program	Weeks	Modality	Matched answers	Participation rate
Course 1	FY	1st	A	B.Sc.	6	In-person	37	74% **
Course 2	FY	1st	A	B.Sc.	6	Hybrid*	25	48% **
Course 3	FY	1st	A	B.Ed.	4	Online*	21	51% **
Course 5	FY	1st	A	B.Ed.	4	In-person	5	33% **
Course 9	FY	2nd	B	B.Sc.	9	Hybrid*	1	3%
Course 4	BFY	3th	B	B.Sc.	7	In-person*	12	30%
Course 8	BFY	3th	B	B.Ed.	4	In-person*	2	20%
Course 7	BFY	4th	C	B.Sc.	8	Hybrid*	3	14%
Course 6	BFY	4th	C	B.Ed.	6	Online*	5	33%

with a asterisk if the courses were affected by the COVID-19 pandemic. The main aspect that changed during the COVID-19 pandemic was that students conducted the experiments alone in the lab and/or at home (when the restrictions occurred) and not in groups as usual. The course type C was created during the COVID-19 pandemic as a solution for letting students conduct the experiments at home, and because of the positive student response, we kept this type of course in our program. In the last two columns of Table I, we show for each course (i) how many matched pre- or post-test student responses we collected for the data analysis and (ii) what percentage of the class this number correspond to. Because of the data privacy rules, we are not able to give course credit for completion of the survey, which would increase participation rate as it was shown for the E-CLASS [12]. This fact explains, in part, the low percentages in the last column. Another cause for the low matching rate is, as stated above, that students ID codes often do not match. To have higher participation rate, we recommend allocating extra time during class to complete the surveys. In fact, courses in which we gave in-class time (indicated with a symbol \*\*) show higher participation rates.

Here, we describe briefly the general learning goals of our laboratory course, as well as the specific learning goals of each of the four parts of our introductory laboratory course curriculum and the course structure.

### 1. Description of type of courses

In 2016, we started a process of restructuring our laboratory courses at University of Potsdam, focusing on teaching students experimental skills and offering an authentic experience of the processes of experimental physics. Examples of these experimental skills are modeling and design, as well as critical thinking and problem solving. Furthermore, we want students to understand the nature of experimental physics and learn to think like a

scientist. (Note that the reconstruction of our courses is a “work in progress” and some experiments offered to students in course type B have not been transformed yet. We are using the results presented in this work, to guide further our course transformation.) Here, we describe briefly the specific learning goals and structures of the different types of courses investigated in this study.

**Course type A:** We had three primary specific goals for the first semester courses. We wanted students to

1. be able to apply the basic concepts of measurement uncertainties and systematic errors;
2. be able to write a good laboratory notebook and explain why physicists use laboratory notebooks; and
3. be able to recognize what components make up a good graph, be able to create a good graph, and be able to use graphical analysis methods (for example, least squares fitting and linearization).

This type of course included both a seminar component with exercises, group work, quizzes, peer instruction, and homework problems and a laboratory component. In the lab, students performed experiments that had them practice working with measurement uncertainties, systematic errors, graphing, and writing laboratory notebooks. The experiments often had answers or end results that were initially unknown to the students. For example, students had to measure and compare the reaction time for visual stimuli of their group members. They were also asked to measure their reaction time using another measurement method, compare the results of the two methods (one method having an unknown systematic error), and discuss which of the two methods was more precise and which method was more accurate. In this setting, the laboratory guides had supporting questions to guide students’ work.

**Course type B:** In the second and third semester courses, we want students to understand the role of modeling in experimental physics and practice it during their laboratory work [48]. To engage students in the process of modeling, we made use of the Modeling

Framework for Experimental Physics described in the literature [23,48].

In the second semester, in addition to laboratory work, we had a seminar session in which we introduced the Modeling Framework for Experimental Physics [23]. In the third semester course, we used only the laboratory setting.

As an example of a laboratory for this type of course (two three-hour laboratory sessions), students had to design, build, and calibrate an optical spectrometer capable of resolving the sodium doublet. For this laboratory, students were provided with an optical table, several lenses and gratings, a line camera, and atomic lamps like Na, Ne, and Hg.

**Course type C:** In this type of course, we want students to continue engaging with modeling, but also work on experimental design and communication. The additional goals include

1. Learn how to use and program microcontrollers (Arduino) and sensors to perform simple experiments with them using the process of modeling; and
2. Practice how to plan and design an experiment based on students' own ideas and write a proposal about it.

After an introduction session on Arduinos microcontrollers and their programming language, students performed several experiments (at home) that aimed to foster modeling. Students were given little guidance on how to set up experiments in their home and how to take the data, but were given a specific problem to work on and were encouraged to engage in the process of modeling. In the second part of the course, there was one session focused on how to write proposals. Students then planned their own experiment using Arduinos and sensors, wrote a proposal about it, and peer-reviewed other students' proposals. For course 7, students were given time for the realization of their idea in the lab.

## B. Datasets

### 1. GE-CLASS

Using the GE-CLASS system, we obtained 111 valid student survey responses. Responses are considered valid only in case the anonymous code and the course identification code described above matched in the pre- and post-tests. Additionally, a control question was used to validate pre- and post-tests independently, in which students are explicitly asked to respond with a particular answer. Among the 111 responses, 89 responses are from courses for the first year (FY) and 22 are from courses beyond the first year (BFY) (Table I).

### 2. E-CLASS

One of our research questions is to compare the sample of responses at the University of Potsdam to responses from U.S. institutions broadly. Since we do not have enough data

TABLE II. Data used in this study. The numbers for GE-CLASS represent all of the data from University of Potsdam. The data from E-CLASS represent a filtered dataset to match the ratio of FY to BFY responses present in the GE-CLASS data set. All data are from physics majors.

	GE-CLASS	E-CLASS
Institutions	1	90
FY courses	4	306
BFY courses	5	241
Matched student responses (FY)	89	1073
Matched student responses (BFY)	22	264
Total student responses	111	1337

to be able to split our GE-CLASS data into FY and BFY subsets, and all of the students at Potsdam considered in this study are physics majors, we select data from the U.S. dataset to make more direct comparisons. To do this, we first filter the U.S. data to include only physics majors. Then, we randomly choose students in both FY and BFY courses such that the ratio of these two populations is the same for the U.S. dataset and Potsdam dataset. We note this is a limitation of our work in Sec. V F.

The details of the resulting filtered datasets for the E-CLASS and GE-CLASS are shown in Table II.

## C. Data analysis

We used two different schemes for comparing students' answers to the ER. One for creating the reports for instructors (i.e., for instructional purposes) and one for investigating our results for research purposes. We present both types of analysis as the stated goals for this paper include encouraging instructors to use GE-CLASS, so it is important to present results that they will see in their reports, and beginning to investigate the impact of courses on students' views for use by education researchers, so we include the second format of our results. In both cases, we first reduced the five-point Likert scale (see Sec. II) to a three-point Likert scale, by collapsing "strongly (dis)agree" and "(dis)agree" into a single category. For the instructor reports, we used a binary scale analysis scheme. This was realized by assigning a numerical score of +1 for answers in agreement with the ER and a score of 0 otherwise. Such a binary analysis is easy to interpret for instructors as it directly provides information on what percentage of the class agrees with experts and has an easy to interpret visual representation. For research studies, we use a three-point scale analysis. Here, we assign a numerical score of +1 for agreement with the ER, 0 for neutral, and -1 for disagreement with the ER. These schemes are the same ones used for instructors and researchers for E-CLASS.

In this paper, the results in Figs. 1 and 2 are obtained using a two-point analysis, while results in Figs. 3 and 4 are obtained using a three-point analysis.



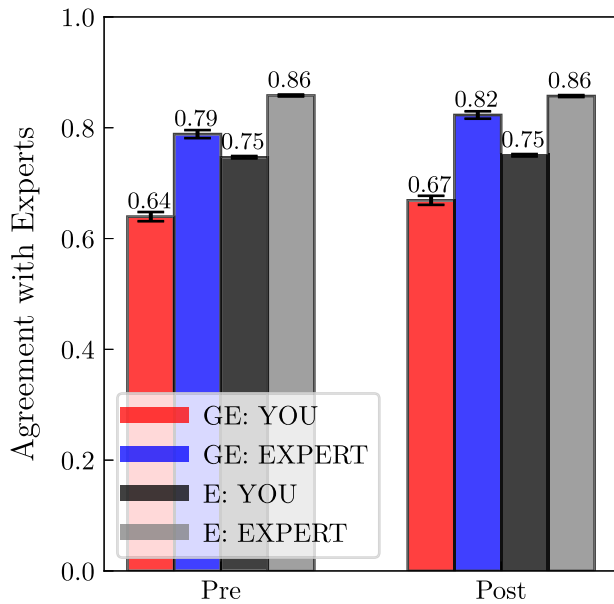


FIG. 1. Overall GE-CLASS agreement with experts for the pre and post-tests. Perfect agreement with experts corresponds to one, zero means no agreement with experts. In red are the GE-CLASS results for YOU questions and in blue for the EXPERT questions. For comparison, E-CLASS results are indicated in the figure as well (in dark and light gray). The overall mean shown here averages over all students and all items on the survey. Error bars are standard deviations of the mean.

In this paper, we used two statistical (nonparametric) tests, the Mann-Whitney U-test [49], and the Anderson-Darling test for  $k$  samples [50]. The Mann-Whitney U procedure was used to test if the observed differences between two sample distributions were statistically significant. We used the null hypothesis that the two samples are coming from the same population. We rejected the null hypothesis for a  $p$  value  $p < 0.05$ . To evaluate practical significance, we calculate an effect size using Cohen's  $d$  [51]. In order to distinguish if the cumulative distributions in Fig. 4 were statistically different, we used the Anderson-Darling  $k$ -samples test, with null hypothesis that the two-samples are drawn from the same population. For this test we used  $k = 2$  and a significance level of 0.05.

#### D. Results

The overall agreement with the ER is shown in Fig. 1. It was calculated using the 2-point scale analysis described above, by averaging over all students and all items. Figure 1 shows both results for students personal views, as well as students perspective of experts (labeled as YOU and EXPERT, respectively) for both the GE-CLASS and E-CLASS (labeled as GE and E, respectively). This type of graph is also provided to instructors by the automated GE-CLASS system [47].

For the YOU questions, the overall GE-CLASS agreement with the ER is  $0.640 \pm 0.008$  before instruction and

$0.667 \pm 0.008$  after instruction (see Fig. 1). This shift is statistically significant, as tested using a Mann-Whitney U-test ( $p = 0.0058 \ll 0.05$ ) [49] with a very small effect size (Cohen's  $d = -0.062$  [51]).

We also observe that students perspectives of what experts think (EXPERT questions) have much larger preinstruction agreement with the ER ( $0.789 \pm 0.007$ ) than the YOU questions. The students' perspective of experts is also positively influenced by instruction, as we measure  $0.823 \pm 0.007$  for the post-test. This positive shift is statistically significant with a Mann-Whitney U-test  $p \ll 0.05$  and a very small effect size (Cohen's  $d = -0.087$ ).

In Fig. 1, we also show the results from the E-CLASS dataset. We found that the overall E-CLASS agreement with ER for the YOU questions both before and after instruction are higher than the scores from the University of Potsdam. Before instruction, the E-CLASS overall score is  $0.747 \pm 0.002$ . The difference between E- and GE-CLASS is statistically significant (Mann-Whitney U-test  $p \ll 0.001$ ) with a small effect size (Cohen's  $d = -0.24$ ). After instruction, the E-CLASS overall score is  $0.750 \pm 0.002$ . This difference between E- and GE-CLASS results is also statistically significant (Mann-Whitney U-test  $p \ll 0.001$ ) with a small effect size (Cohen's  $d = -0.20$ ).

For the EXPERT questions, the overall expert agreement score before instruction is for the E-CLASS  $0.858 \pm 0.002$ . This value is higher than for the GE-CLASS. The difference between these values is statistically significant (Mann-Whitney U test  $p \ll 0.001$ ) with a small effect size (Cohen's  $d = -0.20$ ). However, the gap between E- and GE-CLASS results becomes smaller for the EXPERT questions after instruction ( $0.857 \pm 0.002$  for the E-CLASS). The difference is statistically significant with a smaller size effect than for the pre-instruction (Cohen's  $d = -0.10$ ).

It is important to note that no change upon instruction (for both YOU and EXPERT questions) is observed in the E-CLASS for the dataset used here.

In Fig. 2, we present a detailed item-by-item analysis of the GE-CLASS responses. This type of graph is also provided to instructors by the automated GE-CLASS system [47]. Once again, here, a 2-point scale analysis was used. Results are ordered based on the level of agreement with the ER in the preinstruction YOU questions. In Fig. 2, we also show the results of the E-CLASS for comparison.

The EXPERT questions show larger agreement with the ER than the YOU questions in general and, in particular, for questions Q29, Q6, Q14, Q10, and Q5 in both E-CLASS and GE-CLASS data. Thus, there is a pronounced gap between students personal view and students' knowledge of experts' views for these questions.

Moreover, we observe that in 30% of the survey's items, the values of ER agreement of YOU questions in the GE-CLASS data are significantly lower than those of the E-CLASS results. For the EXPERT questions, this is the case for about 7% of the items. For determining significance here, we considered the 95% confidence

### What do YOU think ... vs What do EXPERTS think ...

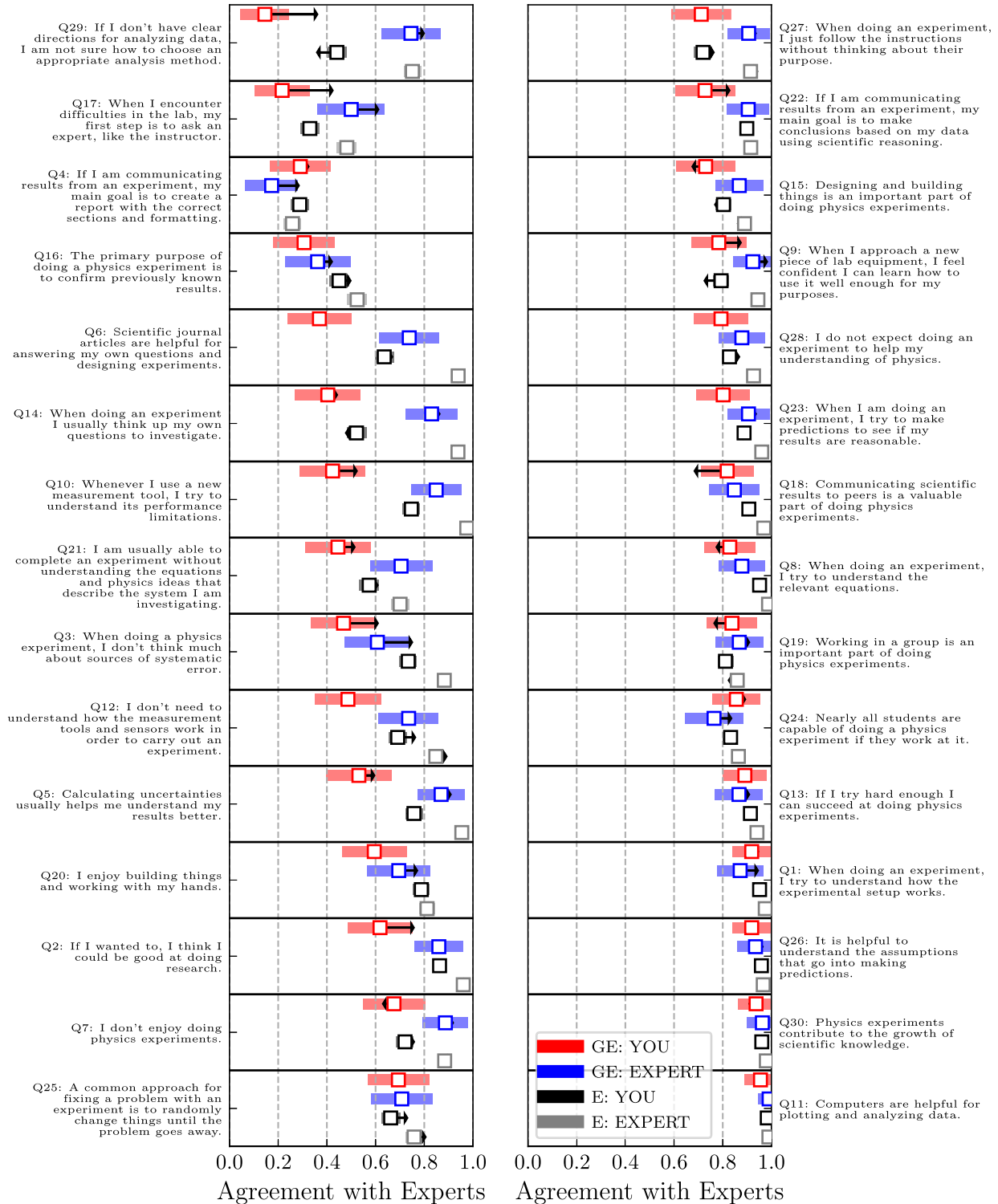


FIG. 2. Fraction of responses that agree with the ER for each item of the survey. Squares indicate the mean for the pre-instruction responses for YOU-question (red) and EXPERT-question (blue) views. Black and gray squares are the E-CLASS preinstruction responses for YOU-question and EXPERT-question view, respectively. Shaded bars indicate the 95% confidence intervals of the preinstruction fraction. These intervals are not visible for E-CLASS results since they are smaller than the data point symbols. The end of the arrows indicate the postinstruction agreement with experts. Note the questions in the figure are ordered according to the preinstruction personal views for the GE-CLASS data.

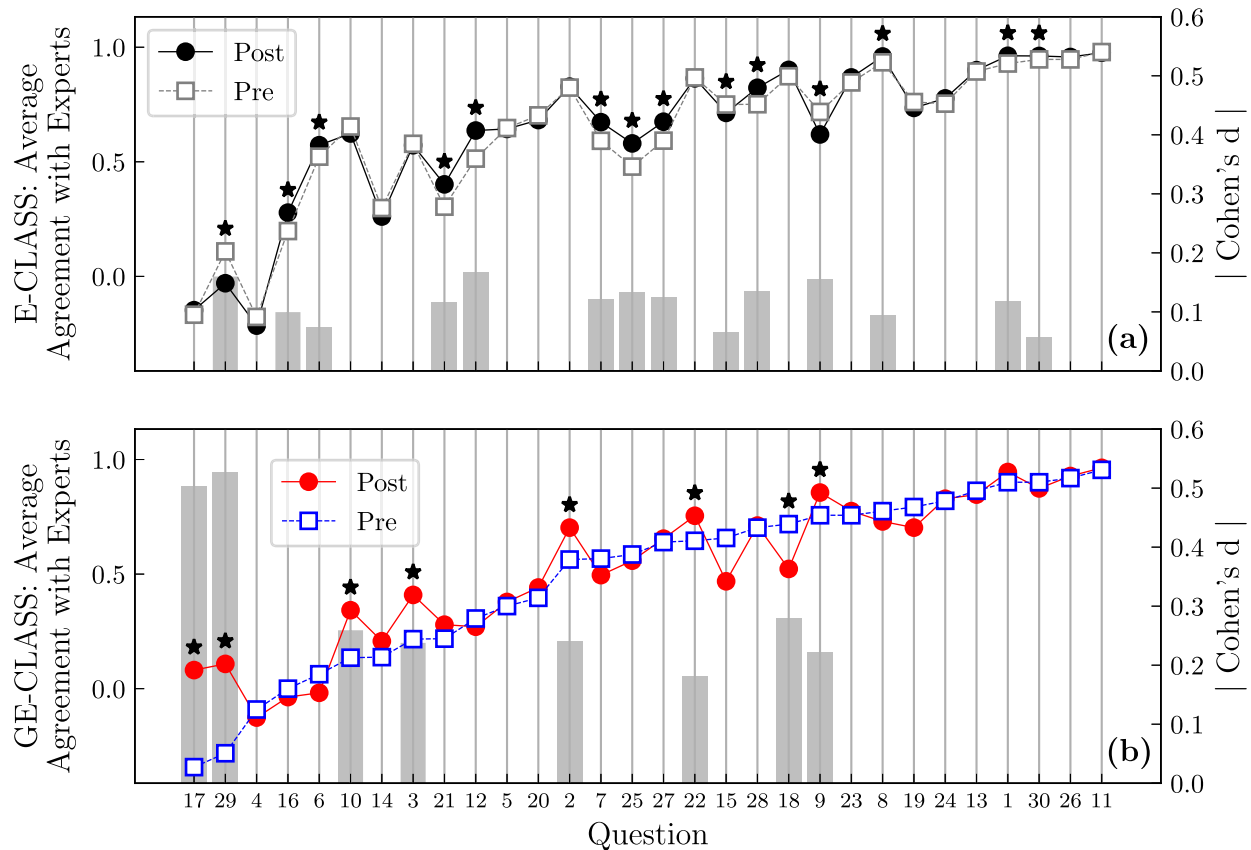


FIG. 3. Overall agreement with the ER for YOU questions, item by item, before and after instruction for E-CLASS (top graph) and for the GE-CLASS (lower graph). The results are obtained using a 3-point scoring method (left axis). Stars indicate pre-post changes that are statistically significant, as obtained by using the Mann-Whitney U-test. The absolute values of the Cohen's  $d$  are also indicated in the figure as bars for the items with a statistically significant change after instruction (right axis).

intervals indicated in Fig. 2. Therefore, we found that students' personal beliefs at the University of Potsdam are, for a considerable amount of items of the survey, lower than for students participating in the E-CLASS.

Looking now at changes upon instruction, we observe the largest positive shifts in the GE-CLASS for the YOU-questions Q29 (*If I don't have clear directions for analyzing data, I am not sure how to choose an appropriate analysis method*) and Q17 (*When I encounter difficulties in the lab, my first step is to ask an expert, like the instructor*). Those items are followed by Q3 (*When doing a physics experiment, I don't think much about sources of systematic error.*) and Q2 (*If I wanted to, I think I could be good at doing research*). The largest negative shift is observed for question Q18 (*Communicating scientific results to peers is a valuable part of doing physics experiments*).

To gain further insights about the item-by-item E- and GE-CLASS comparison, we looked at the differences between E- and GE-CLASS results using a 3-point scale analysis as described above. The results are shown in Fig. 3. The survey questions in the figure have been ordered based on the values of the preinstruction GE-CLASS average agreement with the ER. Statistically significant shifts upon

instruction are indicated by stars in Fig. 3. The effect size of those statistically significant shifts has been analyzed by calculating the Cohen's  $d$  values and are indicated in the figure using gray bars referring to the right axis. For the preinstruction data in Fig. 3, the order of items from low to high agreement with ER shows a similar trend for E- and GE-CLASS results. Interestingly, the four questions with the lowest agreement, as well as the three questions with highest agreement, are the same questions for E- and GE-CLASS. This indicates that both groups score highest and lowest (i.e., have expertlike and nonexpertlike thinking) on items regarding the same aspects of experimental physics. Additionally, in the case of the GE-CLASS we found that for seven items there are statistically significant positive shifts (see Fig. 3), while for one item [Q18 (*Communicating scientific results to peers is a valuable part of doing physics experiments*)] there is a statistically significant negative shift. The fact that we observed negative shifts for items Q15 and Q18 of the GE-CLASS, motivated us to specifically put effort in creating laboratory activities to foster communication and apparatus building.

For the E-CLASS data in Fig. 3, a statistically significant positive shift occurs for 11 items, while for three of the

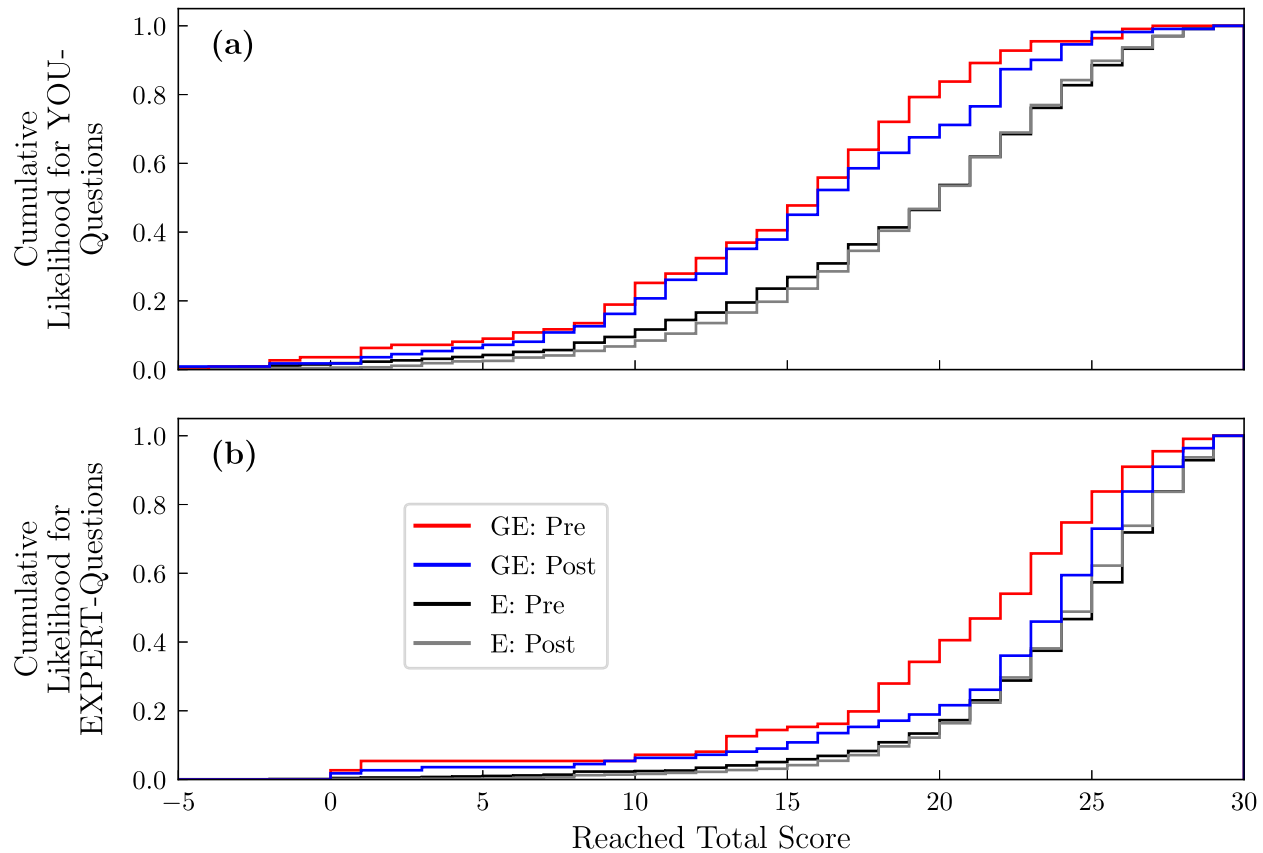


FIG. 4. Cumulative distribution of the agreement with experts as a function of total GE- and E-CLASS score for pre- and post-tests in (a) for YOU questions and (b) for EXPERT questions. The reached total score has been calculated using a 3-point data analysis (see text for more information).

items a statistically significant negative shift is observed (for items 29, 15, and 9). For those items, slightly larger effect sizes are found for the GE-CLASS than for the E-CLASS results, as indicated by the gray bars in Fig. 3, which show the values of the Cohen's  $d$ .

To further compare the E- and GE-CLASS results and the changes upon instruction, we analyzed the changes of the cumulative likelihood distribution of the total score. This allows us to examine the distribution of scores and not just the mean. This distribution is shown in Fig. 4 for both before and after instruction for both the YOU and EXPERT questions. This type of graph shows the integrated fraction of students that received up to a certain total score on the survey, where a student's reached total GE and E-CLASS score is obtained by summing up the scores on each of the 30 items using a 3-point analysis scheme. The reached total score can therefore range from a minimum of  $-30$  (disagreement with experts for all survey items) to a maximum of  $+30$  (agreement with experts for all survey items). The best possible distribution would be zero everywhere and a sharp peak at reached total score =  $+30$ . This would mean that all students agree for all items with experts. Additionally, the smaller the area under the curve the more the distribution of responses that aligns with experts.

Comparing the E- and GE-CLASS results before instruction, we see that the point where the cumulative distribution of students at University of Potsdam begins to increase is shifted to the left with respect to the one for the E-CLASS data, meaning that more students at University of Potsdam start with lower survey scores. Moreover, the curve reaches near 1.0 at a lower reached total score for the GE-CLASS data, meaning that the highest survey scores in the GE-CLASS dataset are lower than for the E-CLASS dataset. This trend is particularly pronounced for the YOU-questions. At University of Potsdam, fewer students reach the highest survey scores before instruction (the maximum total reached score is in fact for the YOU-questions  $+28$ ) as compared to the E-CLASS results where there are students who reach up to  $+30$ .

Using an Anderson-Darling  $k$ -samples statistical test, we investigated if the observed differences between the E-CLASS and GE-CLASS cumulative distributions in both pre- and post-surveys and for YOU and EXPERT questions are statistically significant. We found for all these cases that the two samples of the E- and GE-CLASS do not originate from the same distribution. For both YOU and EXPERT questions, we observe larger shifts upon instruction of the distribution for GE-CLASS as compared to the E-CLASS.



Anderson-Darling  $k$ -samples statistical tests indicate that changes upon instruction are statistically significant only in the case of the EXPERT questions of the GE-CLASS data.

## E. Discussion

We now discuss the results focusing on three main aspects: (i) levels of expertlike thinking, (ii) differences between students' personal perspective and students' perspective of experts, and (iii) changes upon instruction.

### 1. Levels of agreement with experts

The results of the data analysis in Figs. 1 and 2 indicate that there is a clear difference between the level of "expertlike" thinking about experimental physics of students at Potsdam and at other (mostly U.S.) institutions. Such a difference is also observed when we look at students' opinions about expert physicists. We conclude that further efforts to improve the views of students on the nature of experimental physics at the University of Potsdam is a worthwhile endeavor. Moreover, these results lead to additional question, such as are these differences arising from a lack of understanding of the nature of experimental physics of German students in general as compared to other educational environments or is this result specific to the University of Potsdam. To answer this important question, we need to further investigate laboratory courses at other German-speaking institutions using the GE-CLASS. The GE-CLASS web-based system described in this paper allows this as a next step. Regarding the data presented in Fig. 2, we conclude, from an instructional point of view, that we need to put particular effort in facilitating the aspects probed by the items in the left side of this figure. Those are the aspects that show the lowest agreement with the ER.

Looking at the ranking of the survey items in Fig. 3, we conclude that the aspects with the most or least expertlike views of experimental physics do not depend on the educational environments (countries) investigated in this study. These findings are informative for both instructors and the physics education community, as we work to broaden educational improvement efforts beyond our own countries.

### 2. Students' contradictions

As discussed in the previous section and in accordance to the literature [52], we found in this study for physics majors (for both E- and GE-CLASS) that the overall agreement with experts for the YOU questions is lower than the one for the EXPERT questions. Students often know what are experts views and attitudes in the laboratory, but do not personally practice those perspectives when working in laboratory courses. This possible contradiction might arise from a perception of students that laboratory courses are not authentic experimental physics experiences or the type

of activities they engage in are not well aligned with authentic practice.

From Fig. 2, we find that the aspects for which we measured the largest differences between personal and expert views are the same for the two datasets GE and E-CLASS (see, for example, items 29, 6, 14, 10, and 5). The reason why we observed the largest gaps for the same items is unknown, but may be due to a few reasons. For example, the discrepancy for question 5 (*Calculating uncertainties usually helps me understand my results better*) can possibly be explained considering previous studies on students' understanding of measurements uncertainty, which show a variety of student challenges with these concepts and practices in courses in many different countries [9,53–55].

For the items where students hold different views than their perception of experts, we must look at the practices and structures of our courses to posit why these may be different and how we might address this gap.

### 3. Changes upon instruction

We now discuss changes upon instruction for the E- and GE-CLASS data. As outlined in the previous section, we found for the GE-CLASS, for both YOU and EXPERTS questions statistically significant shifts between pre- and postsurveys (Fig. 1). Our transformed curriculum has therefore a small, but positive, influence on students views and attitudes towards experimental physics. By comparing those results with the E-CLASS results, in which no significant shift is observed in Fig. 1, we argue that this small shift may be due to the fact that we identified items covered on the survey as an explicit goal of our laboratory class. For example, the largest positive shifts for questions Q29 and Q17 (Q29 *If I don't have clear directions for analyzing data, I am not sure how to choose an appropriate analysis method* and Q17 *When I encounter difficulties in the lab, my first step is to ask an expert, like the instructor*) were addressed in the course transformation. We started the transformation by adapting the student laboratory manuals to remove the precise instructions on how to perform the experiment and data analysis and instead added self-guiding questions to help students reflect on how to solve problems or how to perform the data analysis. With this approach, we allowed for student decision making in the laboratory. This practice may have led to positive shifts on these two questions. Further qualitative research (classroom observations and student interviews) would need to be conducted to confirm this idea. We do note that the positive shift for Q17 may have been impacted by the COVID-19 pandemic and the resulting course modality. A previous study looked at how E-CLASS scores shifted from pre-pandemic to Fall 2020 and found a strong positive shift for Q17, which was attributed to, at least in part, the remote nature of the laboratories where access to instructor help may not have been as readily available [38].

On the other end, the largest negative shifts for the GE-CLASS are observed for YOU question Q18 (*Communicating scientific results to peers is a valuable part of doing physics experiments.* and Q19 (*Working in a group is an important part of doing physics experiments*), which are related to communication and working in a group. We wonder if this effect is due to changes in the instructions we had to make due to the COVID-19 pandemic ( $N = 42$  pre- and  $N = 83$  post-COVID-19), for which group work was less encouraged compared to before the pandemic and students had to work alone a lot. Regardless, we are now working on specifically fostering communication and team work by activities that emphasize the importance and advantages of team work and peer discussions in an authentic manner.

### F. Limitations

There are a few limitations to the work presented here. First, we note that a large fraction of the data of the GE-CLASS sample was obtained during the COVID-19 pandemic. This might have influenced our results in unpredictable ways, including impacts of the stress and uncertainty of the pandemic on the students. For various reasons, the response rates are low for the GE-CLASS. Thus, the small sample of students from each class may be biased, as students who were experiencing high levels of stress may have chosen not to complete the surveys.

Second, in the case of the E-CLASS, students are often given an incentive for doing the survey, such as a small amount of course credit for completion. Since this incentive was missing in the GE-CLASS, it is possible that students who were more engaged in the course and its contents answered the survey and thus represent a biased sample. We believe that this form of possible bias does not invalidate the results, but must be taken into account when considering the stronger shifts upon instruction of the GE-CLASS.

Finally, this low response rate led to a relatively small dataset to analyze for the GE-CLASS. In an ideal world, we would wait to collect enough data from the University of Potsdam to be able to split the data into FY and BFY datasets. However, due to the pandemic and low response rates caused by adhering to GDPR rules, this would take many years. Critically, this would delay the broad scale use of GE-CLASS across Germany and, thus, delay improvements in German lab classes along this one important dimension.

## VI. CONCLUSIONS

We presented the process of how the E-CLASS was adapted to the German-speaking context. It is now available for instructors to assess laboratory classes in German-speaking institutions through an online automated system. This system will allow instructors and researchers to gain insights into the influence of laboratory classes on students'

epistemology and views about experimental physics in German-speaking institutions, a field that has been unexplored until now.

First results using the GE-CLASS have been obtained at the University of Potsdam, as an example use of the GE-CLASS. The results of the GE-CLASS dataset help us to guide our laboratory course transformation. This study using the GE-CLASS allowed us to make a first comparison with the results obtained from the E-CLASS. From this comparison, we found that there are similarities, but also differences between instructional contexts. For example, in both E- and GE-CLASS results, students have a good understanding about what experts think about experimental physics, but they do not always share those views during their own laboratory class. The aspects of experimental physics for which students have the most and least expert-like personal views are the same for the E-CLASS and GE-CLASS, indicating that the attitudes towards experimental physics show similar characteristics even considering these different educational environments. On the other hand, we found that the level of "expert thinking" is systematically lower for students who took the GE-CLASS at the University of Potsdam as compared to students who took the E-CLASS in the U.S., and that the effects upon instruction are larger for the GE-CLASS. The analysis of the cumulative expert likelihood distribution shows that students at Potsdam reached lower scores both pre- and postinstruction than students did on E-CLASS for the YOU questions. These results indicate the need to further address students' epistemology and views about experimental physics at the University of Potsdam and leads to questions about if these findings would be similar if studies were carried out at other German institutions.

In the future, we hope to use the GE-CLASS to answer additional research questions using an expanded dataset collected from both the University of Potsdam and many additional institutions across Germany. Possible questions include (i) to what extent do students develop habits of mind, experimental strategies, enthusiasm, and confidence in doing experimental physics in German-speaking laboratory courses broadly? (ii) Are different teaching approaches in German-speaking laboratory courses impacting students attitudes and beliefs about experimental physics? (iii) Are there any differences between the U.S. and German-speaking results when we compare the results using the data from a broad selection of laboratory courses across German-speaking countries? The answers to these research questions will be of great value for the further education of physics students in German-speaking, as well as other European, laboratory courses.

## ACKNOWLEDGMENTS

We would like to thank J. Beck, M. Gühr, and M. Robinson for helping with the translation of the E-CLASS. We also thank M. Path and R. Zumpe for

helping students taking the GE-CLASS survey during the laboratory courses. This project was funded by the Stifterverband and the Baden-Württemberg Stiftung (Project No. H1205228500830925) and by National Science Foundation (Grant No. PHY-1734006). Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project No. 491466077.

#### APPENDIX: GERMAN TRANSLATION OF THE E-CLASS

Here, we report a list of all E-CLASS items and their translation into German. In the parenthesis at the end of the E-CLASS items, we indicate the answer of the ER. “A” stands for AGREE and “D” stands for DISAGREE.

Q1: When doing an experiment, I try to understand how the experimental setup works. (ER: A)

Bei der Durchführung eines Experimentes versuche ich zu verstehen wie der Versuchsaufbau funktioniert.

Q2: If I wanted to, I think I could be good at doing research. (ER: A)

Ich denke ich könnte gut in der Durchführung von Forschung sein, falls ich wollte.

Q3: When doing a physics experiment, I don't think much about sources of systematic error. (ER: D)

Während der Durchführung eines physikalischen Experiments denke ich nicht viel über Quellen systematischer Fehler nach.

Q4: If I am communicating results from an experiment, my main goal is to create a report with the correct sections and formatting. (ER: D)

Wenn ich über die Ergebnisse eines Experiments berichte, ist mein Hauptziel einen Bericht mit korrekten Abschnitten und Formatierung zu erstellen.

Q5: Calculating uncertainties usually helps me understand my results better. (ER: A)

Das Berechnen von Messunsicherheiten hilft mir üblicherweise meine Ergebnisse besser zu verstehen.

Q6: Scientific journal articles are helpful for answering my own questions and designing experiments. (ER: A)

Artikel aus wissenschaftlichen Fachzeitschriften helfen mir meine eigenen Fragen zu beantworten und beim Entwurf von Experimenten.

Q7: I don't enjoy doing physics experiments. (ER: D)

Ich führe physikalische Experimente ungern durch.

Q8: When doing an experiment, I try to understand the relevant equations. (ER: A)

Während ich ein Experiment durchführe, versuche ich die relevanten Gleichungen zu verstehen.

Q9: When I approach a new piece of lab equipment, I feel confident I can learn how to use it well enough for my purposes. (ER: A)

Wenn ich auf ein neues Laborgerät stoße, bin ich mir sicher, dass ich lernen kann, für meine Zwecke gut genug damit umzugehen.

Q10: Whenever I use a new measurement tool, I try to understand its performance limitations. (ER: A)

Immer wenn ich ein neues Messgerät nutze, versuche ich die Grenzen seiner Leistungsfähigkeit zu verstehen.

Q11: Computers are helpful for plotting and analyzing data. (ER: A)

Computer sind hilfreich zur graphischen Darstellung und Analyse von Daten.

Q12: I don't need to understand how the measurement tools and sensors work in order to carry out an experiment. (ER: D)

Ich muss nicht verstehen, wie Sensoren und Messinstrumente funktionieren, um ein Experiment durchzuführen.

Q13: If I try hard enough I can succeed at doing physics experiments. (ER: A)

Wenn ich mich genug anstrengt, kann ich erfolgreich physikalische Experimente durchführen.

Q14: When doing an experiment I usually think up my own questions to investigate. (ER: A)

Wenn ich ein Experiment durchführe überlege ich mir normalerweise meine eigenen Fragen zu erforschen.

Q15: Designing and building things is an important part of doing physics experiments. (ER: A)

Das Entwerfen und Aufbauen von Gegenständen ist ein wichtiger Teil von physikalischen Experimenten.

Q16: The primary purpose of doing a physics experiment is to confirm previously known results. (ER: D)

Der Hauptzweck eines physikalischen Experimentes besteht in der Bestätigung bereits bekannter Resultate.

Q17: When I encounter difficulties in the lab, my first step is to ask an expert, like the instructor. (ER: D)

Wenn ich auf Schwierigkeiten im Labor stoße ist mein erster Schritt, einen Experten, wie z.B. den Betreuer, zu fragen.

Q18: Communicating scientific results to peers is a valuable part of doing physics experiments. (ER: A)

Die Vermittlung von wissenschaftlichen Resultaten an Kommilitonen ist ein wertvoller Teil des physikalischen Experimentierens.

Q19: Working in a group is an important part of doing physics experiments. (ER: A)

Gruppenarbeit ist ein wichtiger Teil der Durchführung physikalischer Experimente.

Q20: I enjoy building things and working with my hands. (ER: A)

Ich baue gerne Sachen und arbeite gerne mit meinen Händen.



Q21: I am usually able to complete an experiment without understanding the equations and physics ideas that describe the system I am investigating. (ER: D)

Meistens kann ich ein Experiment abschließen, ohne dass ich die Gleichungen und die physikalischen Konzepte verstehe, die das untersuchte System beschreiben.

Q22: If I am communicating results from an experiment, my main goal is to make conclusions based on my data using scientific reasoning. (ER: A)

Für mich ist bei der Präsentation meiner experimentellen Ergebnisse am wichtigsten, dass die Schlussfolgerungen aus meinen Daten unter Nutzung wissenschaftlicher Überlegungen hervorkommen.

Q23: When I am doing an experiment, I try to make predictions to see if my results are reasonable. (ER: A)

Wenn ich ein Experiment durchführe, versuche ich Vorhersagen zu treffen, um zu sehen ob meine Resultate sinnvoll sind.

Q24: Nearly all students are capable of doing a physics experiment if they work at it. (ER: A)

Fast alle Studierenden sind fähig ein physikalisches Experiment durchzuführen wenn sie sich Mühe geben.

Q25: A common approach for fixing a problem with an experiment is to randomly change things until the problem goes away. (ER: D)

Um ein Problem am Experiment zu beheben ändert man üblicherweise zufällig Sachen bis das Problem verschwindet.

Q26: It is helpful to understand the assumptions that go into making predictions. (ER: A)

Es ist hilfreich die vorausgesetzten Annahmen zu verstehen, die in Vorhersagen einfließen.

Q27: When doing an experiment, I just follow the instructions without thinking about their purpose. (ER: D)

Während ich ein Experiment durchführe folge ich einfach der Anweisung ohne über ihren Zweck nachzudenken.

Q28: I do not expect doing an experiment to help my understanding of physics. (ER: D)

Ich erwarte nicht, dass die Durchführung eines Experiments mir beim Verstehen von Physik hilft.

Q29: If I don't have clear directions for analyzing data, I am not sure how to choose an appropriate analysis method. (ER: D)

Ohne klare Anweisungen zur Datenanalyse bin ich mir unsicher über die angemessene Analyse-methode.

Q30: Physics experiments contribute to the growth of scientific knowledge. (ER: A)

Physikalische Experimente tragen zum Wachstum des wissenschaftlichen Wissensschatzes bei.

The following questions are asked only in the post-test: How important for earning a good grade in this class was...

Wie wichtig für eine gute Note im Praktikum war:

... understanding how the experimental setup works?  
... es zu verstehen, wie der experimentelle Aufbau funktioniert?

... thinking about sources of systematic error?  
... das Nachdenken über systematische Fehlerquellen?

... communicating results with the correct sections and formatting?  
... Ergebnisse mit korrekter Gliederung und Formatierung zu vermitteln?

... calculating uncertainties to better understand my results?  
... die Berechnung von Messunsicherheiten zum besseren Verständnis meiner Ergebnisse?

... reading scientific journal articles?  
... das Lesen von Artikeln in wissenschaftlichen Fachzeitschriften?

... understanding the relevant equations?  
... das Verständnis der relevanten Gleichungen?

... learning to use a new piece of laboratory equipment?  
... das Erlernen der Nutzung von neuen Laborinstrumenten?

... understanding the performance limitations of the measurement tools?  
... das Verstehen der Leistungsgrenzen der Messinstrumente?

... using a computer for plotting and analyzing data?  
... die Nutzung eines Computers zur graphischen Darstellung und Analyse der Daten?

... understanding how the measurement tools and sensors work?  
... es zu verstehen wie die Messinstrumente und Sensoren funktionieren?

... thinking up my own questions to investigate?  
... meine eigenen Fragen zur Erforschung zu überlegen?

... designing and building things?  
... der Entwurf und Aufbau von Gegenständen?

... confirming previously known results?  
... die Bestätigung bereits bekannter Resultate?

... overcoming difficulties without the instructors help?  
... die Überwindung von Schwierigkeiten ohne die Hilfe des Betreuers?

... communicating scientific results to peers?  
... die Vermittlung wissenschaftlicher Ergebnisse an Kommilitonen?

... working in a group?  
... die Arbeit in der Gruppe?

... understanding the equations and physics ideas that describe the system I am investigating?



- ... das Verständnis der Gleichungen und der physikalischen Konzepte die das untersuchte System beschreiben?
- ... making conclusions based on data using scientific reasoning?
  - ... basierend auf Daten und wissenschaftlicher Denkweise, Rückschlüsse zu ziehen?
- ... making predictions to see if my results are reasonable?
  - ... es, Vorhersage zu treffen um zu sehen ob meine Ergebnisse sinnvoll sind?
- ... randomly changing things to fix a problem with the experiment?
  - ... Zufällig Sachen am Experiment zu ändern bis ein Problem am Experiment verschwindet?

- ... understanding the approximations and simplifications that are included in theoretical predictions?
  - ... das Verständnis der Näherungen und Vereinfachungen die in die theoretischen Vorhersagen eingehen?
- ... thinking about the purpose of the instructions in the lab guide?
  - ... das Nachdenken über den Zweck der Anleitungen in der Versuchsanleitung?
- ... choosing an appropriate method for analyzing data (without explicit direction)?
  - ... die Auswahl einer angemessenen Methode zur Datenanalyse (ohne explizite Anleitung)?

- 
- [1] R. Trumper, The physics laboratory—A historical overview and future perspectives, *Sci. Educ.* **12**, 645 (2003).
  - [2] American Association Physics Teachers Committee on Laboratories, *AAPT Recommendations for the Undergraduate Physics Laboratory Curriculum* (AAPT, College Park, MD, 2015).
  - [3] AAPT, Goals of the introductory physics laboratory, *Am. J. Phys.* **66**, 483 (1998).
  - [4] C. Nagel, R. Scholz, and K. Weber, Umfrage zu den Lehr/Lernzielen in physikalischen Praktika, in *PhyDid B, Didaktik der Physik, Beiträge zur DPG-Frühjahrstagung Dresden, 2018*, edited by V. Nordmeier and H. Grötzebauch (2018), p. 97, <http://phydid.physik.fu-berlin.de/index.php/phydid-b/issue/view/285>.
  - [5] N. G. Holmes and H. J. Lewandowski, Investigating the landscape of physics laboratory instruction across North America, *Phys. Rev. Phys. Educ. Res.* **16**, 020162 (2020).
  - [6] C. Wieman and N. G. Holmes, Measuring the impact of an instructional laboratory on the learning of introductory physics, *Am. J. Phys.* **83**, 972 (2015).
  - [7] N. G. Holmes and C. Wieman, Introductory physics labs: We can do better, *Phys. Today* **71**, No. 1, 38 (2018).
  - [8] E. Etkina, A. Karelina, M. Ruibal-Villasenor, D. Rosengrant, R. Jordan, and C. Hmelo-Silver, Design and reflection help students develop scientific abilities: Learning in introductory physics laboratories, *J. Learn. Sci.* **19**, 54 (2010).
  - [9] S. Pillay, A. Buffler, F. Lubben, and S. Allie, Effectiveness of a GUM-compliant course for teaching measurement in the introductory physics laboratory, *Eur. J. Phys.* **29**, 647 (2008).
  - [10] T. S. Volkwyn, S. Allie, A. Buffler, and F. Lubben, Impact of a conventional introductory laboratory course on the understanding of measurement, *Phys. Rev. ST Phys. Educ. Res.* **4**, 010108 (2008).
  - [11] B. M. Zwickl, N. Finkelstein, and H. J. Lewandowski, The process of transforming an advanced lab course: Goals curriculum, and assessments, *Am. J. Phys.* **81**, 63 (2013).
  - [12] B. M. Zwickl, T. Hirokawa, N. Finkelstein, and H. J. Lewandowski, Epistemology and expectations survey about experimental physics: Development and initial results, *Phys. Rev. ST Phys. Educ. Res.* **10**, 010120 (2014).
  - [13] B. R. Wilcox and H. J. Lewandowski, A summary of research-based assessment of students' beliefs about the nature of experimental physics, *Am. J. Phys.* **86**, 212 (2018).
  - [14] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010101 (2006).
  - [15] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics, *Am. J. Phys.* **66**, 212 (1998).
  - [16] K. Perkins, W. Adams, S. Pollock, N. Finkelstein, and C. Wieman, Correlating student beliefs with student learning using the Colorado Learning Attitudes about Science Survey, *AIP Conf. Proc.* **790**, 61 (2005).
  - [17] L. Lising and A. Elby, The impact of epistemology on learning: A case study from introductory physics, *Am. J. Phys.* **73**, 372 (2005).
  - [18] J. Handelsman, D. Ebert-May, R. Beichner, P. Bruns, A. Chang, R. DeHaan, J. Gentile, S. Lauffer, J. Stewart, S. M. Tilghman, and W. B. Wood, Scientific Teaching, *Science* **304**, 521 (2004).
  - [19] E. Etkina and A. V. Heuvelen, Investigative science learning environment: Using the processes of science and cognitive strategies to learn physics, in *Proceedings of the 2021 Physics Education Research Conference, Rochester, NY*, edited by K. Cummings, S. Franklin, and J. Marx (AIP, New York, 2001).
  - [20] E. Etkina, A. V. 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).

- [21] A. Karelina and E. Etkina, Acting like a physicist: Student approach study to experimental design, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020106 (2007).
- [22] N. G. Holmes, C. E. Wieman, and D. A. Bonn, Teaching critical thinking, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11199 (2015).
- [23] B. M. Zwickl, N. F. D. Hu, and H. J. Lewandowski, Model-based reasoning in the physics laboratory: Framework and initial results, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020113 (2015).
- [24] D. Dounas-Frazer, K. S. Johnson, S. E. Park, J. T. Stanley, and H. J. Lewandowski, Student perceptions of laboratory classroom activities and experimental physics practice, in *Proceedings of the 2020 Physics Education Research Conference*, Virtual Conference (AIP, New York, 2020), pp. 131–136.
- [25] <https://www.physport.org/assessments>.
- [26] J. Day and D. Bonn, Development of the concise data processing assessment, *Phys. Rev. ST Phys. Educ. Res.* **7**, 010114 (2011).
- [27] C. Walsh, K. N. Quinn, C. Wieman, and N. G. Holmes, Quantifying critical thinking: Development and validation of the physics lab inventory of critical thinking, *Phys. Rev. Phys. Educ. Res.* **15**, 010135 (2019).
- [28] H. Theyßen, M. Dickmann, K. Neumann, H. Schecker, and B. Eickhorst, Measuring experimental skills in large scale assessments: A simulation-based test instrument, in *Proceedings of the Bi-Annual Conference of the European Science Education Research Conference (ESERA)*, edited by J. Dolin and P. Kind (University of Cyprus, Nicosia, 2013), p. 1598.
- [29] H. Theyßen, H. Schecker, K. Neumann, B. Eickhorst, and M. Dickmann, Messung experimenteller Kompetenz ein computergestützter Experimentierertest, *PhyDid A* **15**, 26 (2016), <http://phydid.physik.fu-berlin.de/index.php/phydid/issue/view/Phydid-A%20Nr.%2015%20Band%201>.
- [30] D. Rehfeldt and V. Nordmeier, Lehrqualität naturwissenschaftlicher Hochschulpraktika. Befunde zu Chemie- und Physikpraktika, sowie Block- und Semesterpraktika, *PhyDid A* **17**, 34 (2018), <http://phydid.physik.fu-berlin.de/index.php/phydid/article/view/821/964>.
- [31] B. R. Wilcox and H. J. Lewandowski, Students epistemologies about experimental physics: Validating the Colorado Learning Attitudes about Science Survey for experimental physics, *Phys. Rev. Phys. Educ. Res.* **12**, 010123 (2016).
- [32] B. Priemer, Ein diagnostischer Test zu Schüleransichten über Physik und Lernen von Physik - eine deutsche Version des Tests "Views About Science Survey", *Zeitschrift für Didaktik der Naturwissenschaften* **9**, 160 (2003).
- [33] H. Michel and I. Neumann, Nature of science and science content learning: The relation between students nature of science understanding and their learning about the concept of energy, *Sci. Educ.* **25**, 951 (2016).
- [34] D. Woitkowski, L. Rochell, and A. B. Bauer, German university students' views of nature of science in the introductory phase, *Phys. Rev. Phys. Educ. Res.* **17**, 010118 (2021).
- [35] B. R. Wilcox, B. M. Zwickl, R. D. Hobbs, J. M. Aiken, N. M. Welch, and H. J. Lewandowski, Alternative model for administration and analysis of research-based assessments, *Phys. Rev. Phys. Educ. Res.* **12**, 010139 (2016).
- [36] B. R. Wilcox and H. J. Lewandowski, Open-ended versus guided laboratory activities: Impact on students beliefs about experimental physics, *Phys. Rev. Phys. Educ. Res.* **12**, 020132 (2016).
- [37] B. R. Wilcox and H. J. Lewandowski, Developing skills versus reinforcing concepts in physics labs: Insight from a survey of students beliefs about experimental physics, *Phys. Rev. Phys. Educ. Res.* **13**, 010108 (2017).
- [38] M. F. J. Fox, J. R. Hoehn, A. Werth, and H. J. Lewandowski, Lab instruction during the COVID-19 pandemic: Effects on student views about experimental physics in comparison with previous years, *Phys. Rev. Phys. Educ. Res.* **17**, 010148 (2021).
- [39] J. M. Aiken and H. J. Lewandowski, Data sharing model for physics education research using the 70000 response Colorado Learning Attitudes about Science Survey for experimental physics dataset, *Phys. Rev. Phys. Educ. Res.* **17**, 020144 (2021).
- [40] B. R. Wilcox and H. J. Lewandowski, Impact of perceived grading practices on students beliefs about experimental physics, in *Proceedings of the 2017 Physics Education Research Conference, Cincinnati, OH* (AIP, New York, 2017).
- [41] B. Van Dusen, M. Shultz, J. M. Nissen, B. R. Wilcox, N. G. Holmes, M. Jariwala, E. W. Close, H. J. Lewandowski, and S. Pollock, Online administration of research-based assessments, *Am. J. Phys.* **89**, 7 (2021).
- [42] <http://geclass.physik.uni-potsdam.de/auth/login>.
- [43] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).
- [44] P. Virtanen *et al.*, and (SciPy 1.0 Contributors), SciPy 1.0: Fundamental algorithms for scientific computing in python, *Nat. Methods* **17**, 261 (2020).
- [45] C. R. Harris *et al.*, Array programming with NumPy, *Nature (London)* **585**, 357 (2020).
- [46] J. D. Hunter, Matplotlib: A 2d graphics environment, *Comput. Sci. Eng.* **9**, 90 (2007).
- [47] <http://geclass.physik.uni-potsdam.de/report/example.pdf>.
- [48] D. R. Dounas-Frazer and H. J. Lewandowski, The modeling framework for experimental physics: Description, development, and applications, *Eur. J. Phys.* **39**, 064005 (2018).
- [49] H. B. Mann and D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, *Ann. Math. Stat.* **18**, 50 (1947).
- [50] F. W. Scholz and M. A. Stephens, K-sample Anderson-Darling tests, *J. Am. Stat. Assoc.* **82**, 918 (1987).
- [51] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Lawrence Erlbaum Associates, Hillsdale, NJ, 1988).
- [52] K. E. Gray, W. K. Adams, C. E. Wieman, and K. K. Perkins, Students know what physicists believe, but they don't agree: A study using the class survey, *Phys. Rev. ST Phys. Educ. Res.* **4**, 020106 (2008).

- [53] B. C. S. Allie, A. Buffler, and F. Lubben, First-year physics students perceptions of the quality of experimental measurements, *Int. J. Sci. Educ.* **20**, 447 (1998).
- [54] B. Pollard, R. Hobbs, J. T. Stanley, D. Dounas-Frazer, and H. J. Lewandowski, Impact of an introductory lab course on students understanding of measurement uncertainty, in *Proceedings of the 2017 Physics Education Research Conference, Cincinnati, OH* (AIP, New York, 2017), pp. 312–315.
- [55] B. Pollard, A. Werth, R. Hobbs, and H. J. Lewandowski, Impact of a course transformation on students reasoning about measurement uncertainty, *Phys. Rev. Phys. Educ. Res.* **16**, 020160 (2020).