

**Examining students' views about validity of experiments: From introductory to Ph.D. students**

Dehui Hu\* and Benjamin M. Zwickl

*School of Physics and Astronomy, Rochester Institute of Technology, Rochester, New York 14623, USA*

(Received 22 August 2017; published 11 April 2018)

We investigated physics students' epistemological views on measurements and validity of experimental results. The roles of experiments in physics have been underemphasized in previous research on students' personal epistemology, and there is a need for a broader view of personal epistemology that incorporates experiments. An epistemological framework incorporating the structure, methodology, and validity of scientific knowledge guided the development of an open-ended survey. The survey was administered to students in algebra-based and calculus-based introductory physics courses, upper-division physics labs, and physics Ph.D. students. Within our sample, we identified several differences in students' ideas about validity and uncertainty in measurement. The majority of introductory students justified the validity of results through agreement with theory or with results from others. Alternatively, Ph.D. students frequently justified the validity of results based on the quality of the experimental process and repeatability of results. When asked about the role of uncertainty analysis, introductory students tended to focus on the representational roles (e.g., describing imperfections, data variability, and human mistakes). However, advanced students focused on the inferential roles of uncertainty analysis (e.g., quantifying reliability, making comparisons, and guiding refinements). The findings suggest that lab courses could emphasize a variety of approaches to establish validity, such as by valuing documentation of the experimental process when evaluating the quality of student work. In order to emphasize the role of uncertainty in an authentic way, labs could provide opportunities to iterate, make repeated comparisons, and make decisions based on those comparisons.

DOI: [10.1103/PhysRevPhysEducRes.14.010121](https://doi.org/10.1103/PhysRevPhysEducRes.14.010121)**I. INTRODUCTION**

Epistemology is an area of philosophy concerned with the nature and justification of human knowledge and generally refers to the theory of knowledge, knowing, and learning [1,2]. A growing area of interest for psychologists and education researchers is that of personal epistemological development and epistemological beliefs: how individuals come to know, the theories and beliefs they hold about knowing, and the manner in which such epistemological premises are a part of and an influence on the cognitive processes of thinking and reasoning [1]. In the past few decades, much research in educational psychology, physics education, and science education has focused on students' attitudes and views about knowledge and learning, which may shape and are shaped by their learning experiences [3–6].

A number of validated survey instruments have been developed to measure students' beliefs and attitudes

towards science and learning science [7–9]. One focus is students' views about learning in physics lecture courses, especially problem solving and conceptual understanding in physics, which includes the Views About Science Survey (VASS) [8], the Maryland Physics Expectations Survey (MPEX) [9], and the Colorado Learning Attitudes about Science Survey (CLASS) [5]. The other research focus is students' views about the nature of science, which includes the suite of the Views of Nature of Science (VNOS) surveys [6]. The third focus is students' views about learning physics in laboratory courses, which includes the Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS) [10,11]. Our goal is to study students' personal epistemology on the nature of science with a specific focus on physics experiments.

Theory and experiment are twin pillars of physics research, and they provide two sources of knowledge: one in the form of model-based reasoning and the other being observations and measurements from experiment. Similarly, physics education has a tradition of emphasizing theory and concepts in lecture and also providing experimental opportunities for students in lab courses. Our study investigated students' views about physics with a specific focus on experiments across a broad student population.

Halloun and Hestenes provide a helpful division of personal epistemology in science into two major dimensions: the scientific dimension and cognitive dimension [12].

\*Corresponding author.  
dxhsps@rit.edu

*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.*

The scientific dimension consists of structure, methodology, and validity of knowledge in the discipline. The cognitive dimension consists of self-efficacy, reflective thinking, and personal relevance. Within physics education, personal epistemology includes students' views about the nature of knowledge and the nature of learning physics [4]. We adapted the three subcategories from Halloun's scientific dimension of students' personal epistemology [12] into our study design: structure, methodology, and validity. We developed, administered, and analyzed results from an open-ended survey that were given to physics students ranging from college freshmen to Ph.D. students. Although the survey addresses all three subcategories (i.e., structure, methodology, validity) to varying degrees, this article only focuses on the validity dimension and the other two dimensions were discussed in an earlier publication [13].

In Sec. II, we review previous epistemological surveys as well as previous studies on the validity of measurements. Then, in Sec. III we discuss the development and administration of our open-ended survey. Finally, in Sec. IV, we present our main results and discuss the implications of this study.

## II. BACKGROUND

How individuals come to believe their knowledge is valid or trustworthy, also known as *justification for knowing*, has been one of the four aspects (i.e., certainty of knowledge, simplicity of knowledge, sources of knowledge, and justification for knowing) in several existing models of personal epistemology [14]. Justification for knowing has been defined as "how individuals justify what they know and how they evaluate their own knowledge and that of others." [1] How to establish the validity of knowledge is a key question in scientific investigations. In physics and astronomy, scientists need to establish validity of any theoretical predictions, computational simulations [15], astronomical observations, and experimental measurements [16]. In this section, we will review two established areas of education research that relate to validity of knowledge. The first area of research is on students' views regarding the validity of their knowledge of science in general and the second area is about students' understanding of measurement and uncertainty concepts in statistics and science.

### A. Prior surveys on the validity of knowledge in science

Previous epistemological surveys in physics and science education research have explored various aspects of students' personal epistemology, including how students justify knowledge that is learned in class and elsewhere. Individuals may justify their beliefs in a number of ways, such as on the basis of their past experiences, through their own observations and assessments, or through knowledge from an authority. Through a review of all the survey items

in several previous epistemological surveys [6–9,17], we have found several common themes related to students' views about the validity of knowledge in physics and science. One theme addresses sources of authority that establish validity of knowledge; such sources could reside within the student or the student could defer to another source such as a textbook, instructor, or government. A second main theme relates to the appropriate use of mathematical equations and derivations to establish knowledge as valid. A third theme deals with the utilization of prior knowledge, experiences, or expectations (either qualitative or quantitative) to justify knowledge as valid.

The following examples from existing surveys address either one or two of the themes. One item from MPEX-II [17] said "Tamara just read something in her physics textbook that seems to disagree with her own experiences. But to learn physics well, Tamara shouldn't think about her own experiences; she should just focus on what the book says." This item investigated students' views about knowledge construction when two sources of knowledge provided different answers, one from authority and the other from personal experiences. One Likert-scale question from CLASS [5] was "In doing a physics problem, if my calculation gives a result very different from what I'd expect, I'd trust the calculation rather than going back through the problem." This item probed if students were in favor of their own expectations or mathematical calculations when justifying the correctness of a result for a problem. Another item from CLASS [7] was "When I am solving a physics problem, I try to decide what would be a reasonable value for the answer." It involved the justification of knowledge on the basis of what feels reasonable although it is not clear how the student decides what value is reasonable.

Those surveys adapted a relatively narrow view of validity in scientific knowledge that mostly focused on juxtaposing authority (e.g., textbooks or teachers) with personal experience or mathematics as the justification for validity. There are also a variety of methodologies used in previous work to probe students' epistemological views. However, these previous works primarily focused on problem solving or knowledge construction in a lecture course instead of an experimental context that includes various methodological approaches [16]. Hammer's [18] adapted case study interviews to investigate students' beliefs about the structure of knowledge, content of knowledge, and the learning of knowledge, primarily in lecture courses, problem solving, and exams. Other work [19] includes making inferences about college physics students' epistemological beliefs about knowledge and learning from their self-reflection through weekly report on how they learned specific physics content. When reflected on how they learned physics knowledge, students described a variety of categories, including authority, logical reasoning, practice, common sense, as well as experimental evidence.

## B. Students' understanding of measurements and uncertainty

Physics is an empirical science, the core of which is experimentation and measurement. Experimentation, observation, and measurement provide the evidence that grounds scientific knowledge. However, one question that arises for any experiment is, "How do we know experimental results are valid?" The validity of experimental results is influenced by almost every aspect of the experiment procedure, including experimental design, use of measurement tools, data collection techniques, documentation of work, data analysis. When communicating findings from an experiment, uncertainty analysis is critical for justifying the validity of an experiment. Physicists typically report uncertainty with measured values to justify the quality of measurements and to facilitate meaningful comparisons with other measurements or theoretical predictions. Every measured value has some degree of uncertainty, and the idea of uncertainty is involved in every phase of a physics experiment, including experimental design, data collection, data processing, and data comparison. Each phase requires that students know what it means to take a measurement and be able to apply this knowledge along with an understanding of the associated uncertainty. Making measurements and quantifying their uncertainty are widely considered to be some of the most fundamental and important components of a student's science education [20,21]. As John Taylor said in his textbook on error analysis, "Because the whole structure and application of science depends on measurements, the ability to evaluate these uncertainties and keep them to a minimum is crucially important" [21].

An understanding of measurement and uncertainty is also critical for making informed decisions in many different situations, such as making legal decisions depending on the accuracy of scientific data [22]. Research studies in statistics and mathematics education have given attention in the development of statistical literacy and numerical skills of all students and citizens [23]. According to Gal, statistical literacy refers to people's ability to interpret, critically evaluate, and express their opinions regarding statistical information, data-related arguments, or stochastic phenomena [24]. In statistical education, reasoning about uncertainty is defined as "understanding and using ideas of randomness, chance, likelihood to make judgments about uncertain events; knowing that not all outcomes are equally likely; knowing how to determine the likelihood of different events using an appropriate method [23]." Studies have looked at students' ability to use probability to make justifications about specific events or situations, and they found that even people who can correctly compute probabilities tend to apply faulty reasoning when asked to make an inference or judgment about an uncertain event, relying on incorrect intuitions [25]. Although physics educators rarely use the term "statistical literacy," physics laboratory

courses are well suited to help students develop statistical reasoning skills through measurements and data processing, and it is often listed as a key learning goal for physics labs [26–28].

Several prior studies in physics education research have examined students' perceptions of measurements and uncertainty in physics laboratory courses [22,29–31]. Those studies focused on students' conceptual understanding of uncertainty in various levels of physics laboratory courses as well as instructional strategies to improve their understanding. They examined several of the broader issues related to measurements: the reasons for repeated measurements, concepts about accuracy and precision, random versus systematic errors, assessing the quality of measured data by the mean and spread, and the determination of uncertainty. In a study conducted by Volkwyn, Allie, and Buffler [32], after receiving the instructions from a conventional laboratory course, most students were able to link repeating measurements to "uncertainty" or "standard deviation" (i.e., focusing on the spread of data) for data collection and data processing; however, many of those students still focused on individual readings or means of several readings when comparing data. In general, strategies used in conventional introductory laboratory courses were unsuccessful in improving students' understanding of uncertainty beyond the appropriation of the numerical routines [32,33]. Buffler, Allie, and Lubben proposed the use of a probabilistic approach to teach measurements and uncertainty and the implementation of a carefully designed curriculum based on this approach in freshman level showed the effectiveness in improving students' understanding [34,35].

In addition to the epistemological role of uncertainty in establishing the validity of measurements, there has also been education research related to uncertainty analysis as a scientific ability. Day and Bonn probed students' abilities related to measurement, uncertainty, and handling data through their Concise Data Processing Assessment (CDPA), a ten question, multiple-choice diagnostic instrument [36]. Those abilities included data fitting, error propagation, and accounting for uncertainties arising from a digital measurement display. Multiple questions asked whether or not a model fits a data set given particular measurement uncertainties. The instrument was administered to first-year physics students, second-year, fourth-year, graduate students, and faculty at the University of British Columbia. The results showed that this instrument was able to distinguish between populations in the novice-to-expert spectrum in regards to their data-handling abilities.

More recently, Holmes, Wieman, and Bonn developed instructional strategies and activities that targeted students' quantitative critical thinking skills (i.e., making quantitative comparisons between data sets and between data and models) [37]. Students in the experimental condition received explicit instructions about how to determine uncertainties in measurements or fit parameters, how to

use quantitative analysis tools (e.g., chi-square test), and how to make decisions about the comparisons, including devising and carrying out a plan to improve the quality of measurements. Students in the experimental condition showed significant improvement in making refinement about experimental methods, as well as identifying and explaining the limitations of a model using their data. Those students also showed the ability to transfer those skills to a subsequent course the following year.

Across those previous studies on measurement and uncertainty, we gained a better understanding of how students interpreted and utilized uncertainties, as well as effective strategies to improve students' abilities to deal with measurements and uncertainty. More specifically, those studies addressed the quantification and utilization of uncertainty when making decisions about data collection (e.g., the number of measurements to take), data processing (e.g., data fitting), and data comparison. However, the concept of uncertainty was not examined in relationship to personal epistemology. In this study, we investigated students' attitudes and views about the justification of validity. Thus, the key research questions in our study are as follows: (i) How do students justify whether or not an experimental result is acceptable or trustworthy? What gives students confidence that the data is trustworthy? (ii) How do students perceive the purpose of uncertainty analysis? By asking about what makes data trustworthy, it offers students a chance to discuss many possible approaches to establishing validity of results, which may include uncertainty evaluation. Because uncertainty analysis has been a pervasive emphasis in lab courses, though evidence suggested students recognize it more as another calculation than as a tool for establishing validity (as experts do). By directly asking students about the role of uncertainty analysis, we offer a modest amount of cuing to see if students describe uncertainty analysis as primarily about validity or if they provide other purposes for conducting uncertainty analysis in a lab setting.

### III. METHODOLOGY

#### A. Survey development and data collection

The scientific dimension with three subcategories of students' personal epistemology about the nature of science (i.e., structure, methodology, and validity) guided the design of our open-ended survey [8]. The main purpose of this study is to probe students' personal epistemology when engaging in scientific practices and not merely learning about science ideas through lecture courses. Halloun's framework perfectly captures students' personal epistemology about the nature of science through scientific practices in the lab environment. The structure dimension is explored through students' views about the relationship between theoretical knowledge and experimental knowledge. The methodology dimension involves the use of

observations and experiments as tools to collect data or evidence and form theoretical ideas. The validity dimension requires students to use existing resources to make judgments and decisions about the data, the experimental setup, and the conclusion.

The development of the open-ended survey consisted of several sequential steps. We first generated several themes following the three subcategories of the scientific dimension, including role of experiments, definition of theory, relation between theory and experiment, and validity of experiment. Based on those themes, we developed a semistructured interview protocol. Next, we conducted 1 hr individual interviews with eight students, two from an introductory algebra-based physics course, two from an introductory calculus-based physics course, two physics majors enrolled in an upper division physics lab, and two physics Ph.D. students. Results from the interviews were used to create an open-ended survey which was then converted to an online version hosted on the Qualtrics survey platform. The online survey was tested again with a small sample of around 10 students. Based on their feedback as well as time considerations, the survey was refined again and shortened to eight questions as shown in Table I. Lastly, we disseminated the survey through Qualtrics, which formed the set of data collected and presented in this paper.

The survey was administered to four different populations with various levels of physics laboratory experiences described in Table II. All undergraduate students were from a large private university. All introductory algebra-based students were enrolled in the same course section, and this course was taken primarily by life science, pre-med, and engineering technology students. Students took two 110-min workshop sessions aligned with a 110-min lecture each week. Most of the workshop sessions involved problem-solving or paper-pencil activities, and there was a lab activity every other week. During the lab activities, students followed an activity manual with step-by-step instructions and built-in guiding questions. All calculus-based physics students were from a single integrated lecture and lab class, and the majority of them were majoring in engineering or computing fields. The class met for six hours each week and about 15%–20% of the in-class time was spent on labs. Students followed a lab manual to do measurements, calculations, plotting, and occasional uncertainty analysis.

The modern physics lab course was for 3rd-year physics majors. The course met each week for a 1-hr prelab to develop a preliminary understanding of the experiment and apparatus and then for a 3-hr period for the main experiment. The course focused on developing proficiencies in statistical uncertainty analysis, computing for data visualization and fitting, and using standard lab equipment such as oscilloscopes and multimeters. Physics Ph.D. students were from the same large public research university. All of them finished their first-year coursework and were involved in different stages of graduate research work.

TABLE I. Survey.

No.	Survey questions
1	In your opinion, why are experiments a common part of physics classes? Provide examples or any evidence to support your answer.
2	In your opinion, why do scientists do experiments for their research? Provide examples or any evidence to support your answer.
3	In your opinion, what defines a scientific theory?
4	How do theory and experiment relate? Provide examples or any evidence to support your answer.
5	What kind of difficulty or difficulties do you typically encounter in a lab?
6	How do you try to resolve those difficulties?
7	How do you know if an experimental result is acceptable or trustworthy? What gives you confidence that your data is trustworthy?
8	Why is uncertainty analysis a common part of physics labs and experiments? How is uncertainty used in physics experiments?

TABLE II. Information of survey respondents.

Population	Intro algebra	Intro calculus	Upper-level	Ph.D.
Course	Algebra-based physics	Calculus-Based Physics	Modern physics lab	Physics research
Institution	Large private university	Large private university	Large private university	Large public research university
Number of responses	75	28	20	31
Administered during	End of semester	End of semester	End of semester	Beginning of summer

A survey link was sent to students via Email, and they were asked to take the survey outside class within two weeks of receiving the link. In introductory courses, the instructors agreed to offer extra credit to students who completed the survey. Upper-division undergraduate students and Ph.D. students were provided with gift cards to compensate for their time.

### B. Coding process

All responses were imported into QSR International's NVivo 11 qualitative data analysis software. The coding process took four steps: (i) Open coding—The primary coder (i.e., the first author D. H.) highlighted key words and grouped similar words and phrases, and then made initial codes. (ii) Focused coding—Two coders (D. H. and the second author B. M. Z.) identified emergent themes in the initial codes and created a hierarchical structure of the coding schema. (iii) Codebook refinement—A draft of the code dictionary was used by the second coder (B. M. Z.), and applied to all the data. Results were discussed with the primary coder (D. H.); then, codes and code definitions were renegotiated resulting in a revised coding dictionary. (iv) Interrater reliability—the coding scheme was given to a third coder who was not involved in the initial creation of the coding dictionary, and coding results were compared with the primary coder (D. H.). If there was a major disagreement on a specific code, the two coders went through the references and coding criteria together and made modifications to their coding or to the coding dictionary as necessary. The final percentage agreement between the two coders for all codes was above 90%.

## IV. RESULTS

Our analysis focuses on survey questions 7 and 8, both of which address the validity dimension of personal epistemology. We present major codes and code distributions by student subgroup for each question, and then we discuss the overall patterns of students' epistemology through a network analysis of codes from the two questions.

### A. Students' views about the validity of experimental results

Question 7 asked, "How do you know if an experimental result is acceptable or trustworthy?" The major codes identified from student responses to question 7 are shown below.

#### 1. Code definitions for Q7: Validity

*Comparison with theory*—Experimental results are trustworthy when they match with theoretical or predicted values. One introductory algebra-based student explained a result was trustworthy "If it is close to what was calculated in the prelab."

*Comparison with others*—Experimental results are trustworthy if they match with those from another source, for example, a published article, or a peer. A graduate student described, "if someone can measure the value using a completely different method and arrive at the same result, taking into account error, this can add confidence."

*Repeatability*—Experimental results are trustworthy when similar results are achieved under multiple trials. Students sometimes used phrases such as, "repeating the

same experiments multiple times” (often meaning the same person doing the repetition) or “reproducibility” (by the same or different experimenters). A graduate student explained, “If you can repeatedly measure the same thing and get reasonably close data, then you may conclude that your data are trustworthy.”

*Uncertainty evaluation*—Measurement uncertainty is calculated and the results are considered trustworthy if the uncertainty is within a certain range. An introductory student said, “If [the result] falls within a given error or set value that is proven to be the constant, then I believe it is acceptable and trustworthy.”

*Quality work*—The experimenter is confident about the methodology, procedure of the lab, and clear documentation of the experiment. This code was mostly used by Ph.D. students. For example, one graduate student explained that “My own perception of how well I set up the experiment and whether I understood everything I saw right up until the taking of data plays a large role...”

*Authority figures*—Experimental results are trustworthy when students receive confirmation from authority, such as the instructor or teaching assistant (TA). An introductory physics student explained, “We usually pass it by the professor or TA, which gives us confidence in our answers/data.”

**2. Results and discussion for Q7: Validity**

Although students at all levels of physics lab experience described several criteria to justify the validity of experimental results, three major categories emerged in their reasoning: *theory-based justification*, *experiment-based justification*, and *authority-based justification* as shown in Fig. 1. Theory-based justification establishes the validity of experimental results by demonstrating agreement with theoretical predictions. Experiment-based justification includes two subcategories—*results-based justification* (i.e., the comparison of experimental results among multiple trials by the same experimenter or different experimenters) and *process-based justification* (i.e., the justification of the experimental process). The final category, authority-based justification, relies upon knowledge from authority figures to establish the validity of experimental results.

Figure 2 shows the frequency of occurrence of all major codes that are defined in Sec. IV A 1. Looking down a single column allows for identifying the most common codes within a subpopulation. For example, the first column “Intro algebra” shows the fraction of introductory algebra-based students’ responses with each code. The two most frequent codes are “comparison with theory” and “comparison with others.” Looking across a row allows for identifying which subpopulation most frequently used a particular type of explanation. For example, the first row comparison with theory shows the fraction of responses in each subpopulation where that code was applied. The two subpopulations where

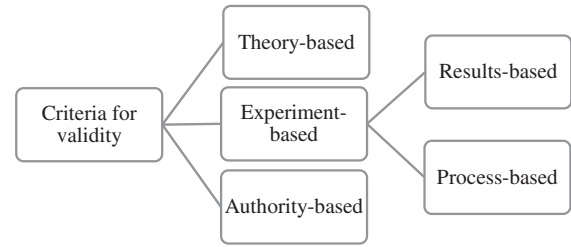


FIG. 1. Criteria for validity.

this code most frequently occurred are introductory calculus-based students and Ph.D. students.

In Fig. 2, comparison with others and repeatability are categorized as result-based justification because they both rely upon a comparison of experimental results across different trials, experiments, or researchers. Uncertainty evaluation and quality work are categorized as process-based justification because they base validity upon an understanding of the experimental setup and the procedure for making measurements. For example, uncertainty evaluation includes the evaluation of the limitations of measurement tools and uncertainty involved in the measurement techniques.

Theory-based justification is one of the most frequently occurring line of reasoning within introductory students’ responses, though it is common across all subpopulations. One example from an introductory algebra-based student was, “If [the results] correspond to the theory answers. This is the best way to check if the results are acceptable.”

**Q7: How do you know if an experimental result is acceptable or trustworthy?**

Criteria for validity	Codes	Codes				
		Intro algebra	Intro calculus	Upper	Ph.D.	
Theory-based	Comparison with theory	0.39	0.68	0.35	0.50	
Experiment-based	Results-based	Comparison with others	0.37	0.29	0.80	0.35
		Repeatability	0.15	0.39	0.25	0.71
	Process-based	Uncertainty evaluation	0.17	0.25	0.55	0.32
		Quality work	0.08	0.11	0.05	0.39
Authority-based	Authority figures	0.12	0.07	0.00	0.00	

FIG. 2. Fraction of student responses with a particular code in the subpopulation (e.g., intro algebra-based physics). Students’ responses may be coded under several codes when appropriate so fractions do not add up to 1.00. Five levels of gray scale shading were applied based on the fraction,  $f$  (from lightest to darkest):  $f < 10\%$  (white);  $10\% \leq f < 25\%$ ;  $25\% \leq f < 50\%$ ;  $50\% \leq f < 75\%$ ;  $f \geq 75\%$  (black).

Another student said, “Compare it to the theoretically calculated answer.” In these examples, experiments done in class do not provide evidence to support theories. Rather, theoretical predictions provide evidence that the experiment worked correctly and yielded the correct result. Although this view seems contrary to an expert view of the nature of science that prioritizes observation and data to establish the validity of theories, the role of theory-based justification does have a useful place in constraining possible outcomes. As one Ph.D. student said, “If the experiment matches some theory prediction. It’s important to have some idea of what the data should look like and that the data is reproducible and not a result of some measurement errors.”

The second major category is results-based justification. Introductory algebra-based students most frequently relied on the comparison of their own results with those of others, especially their peers, to establish validity ( $p < 0.01$  when using a chi-square test to compare the codes comparison with theory or comparison with others with all other codes). One typical phrase from introductory algebra-based students was “If data matches up with other data received from other groups...” Another example of comparing with others refers to the comparison with known or existing results from other sources, such as online resources or published results. One typical example was, “Normally, after doing an experiment, you can check online for published results and compare and see the differences and similarities. If my results match up, I know I did something right!” Upper-division physics majors also mentioned a comparison with literature as well as the results from other research groups as a way to justify the trustworthiness of their experimental results. Ph.D. students differed in their emphasis with over 70% graduate students mentioned repeatability as the major criteria for justifying the trustworthiness of experimental results. An example from Ph.D. students was “Repeating the experiment over multiple days, with different equipment, etc., can all give confidence in the results...” It is possible that time constraints of lab courses limit the use of repeatability in undergraduate lab courses, which leads to a reliance on comparison with others.

The third category of justification, the process-based justification, which includes uncertainty evaluation and quality work, was much more common among Ph.D. students and upper-division undergraduate students ( $p < 0.05$  when using a chi-square test to compare the codes uncertainty evaluation and quality work across different levels of students). Even though introductory students are often asked to calculate uncertainty, prior research shows that introductory students still have difficulties with concepts of uncertainty and using uncertainty for data comparison [27], meaning it is less likely that uncertainty analysis will be used to establish validity of results. However, the upper-division physics majors were

taking a junior-level advanced physics lab course that involved the measurement of fundamental physical constants, such as the electron charge to mass ratio  $e/m$ . Almost all of these experiments required a calculation of uncertainty and making conclusions based on uncertainty, which explains the high percentage (55%) of upper-level students who mentioned uncertainty evaluation. In addition, Ph.D. students also gained confidence by doing quality work on procedural aspects of experiments, including methodology, design, maintaining well-functioning equipment, and keeping good documentation of the experimental process. One typical phrase from graduate students was “having an understanding of the effects of each knob I can turn in my system makes me more confident in my understanding of what I am seeing.” Ph.D. students are often involved in complex experiments that require a good understanding of the physical system, experimental design, procedure, and data. Very often there are no existing results that can be directly compared with, but the results can be trustworthy through a systematic, detail-oriented, and careful approach to their research.

In authority-based justification, there was a small fraction of students (12% for introductory algebra-based students) who mentioned relying upon confirmation from authority figures, such as their professors or teaching assistants, in order to trust their data. An intro algebra-based physics student said, “I have confidence that my data is trustworthy if my professor says it looks right.” None of the upper-level and graduate students mentioned the use of authority figures to justify the validity of their experimental results.

In summary, there are distinct features among introductory students and advanced students in their criteria for the validity of experimental results. Introductory algebra-based students relied primarily upon comparison with theory and comparison with others as the main criteria. Advanced students tended to use multiple criteria, mainly based on the experiment itself, to justify their results. We found that more than 70% of Ph.D. students mentioned about repeatability as a useful criterion, and they almost distinctively used quality work for justifying results. Also, many upper-level and Ph.D. students used the evaluation of uncertainty to justify results.

These general trends are interesting though somewhat unsurprising based on our teaching experience. Introductory students are typically asked to finish a lab within three hours or less. It is challenging for some students to finish one round of data collection within this time limit, and it is very unlikely that an entire experiment is repeated multiple times. The introductory labs done by students in this sample were fairly procedural, and students had little autonomy in designing and performing an experiment. In upper-division and research labs, there are fewer constraints and less rigid procedures, meaning students may have had more flexibility and more responsibility for

designing and conducting experiments. This may have led to the use of a wider range of strategies for establishing validity, such as repeatability, uncertainty evaluation, and quality work.

**B. Views on the role of uncertainty analysis**

Question 8 asked “Why is uncertainty analysis a common part of physics labs or experiments?” This question explored how students perceived the role and purpose of uncertainty analysis in physics labs. The major codes that emerged from students’ responses towards question 8 are shown below.

**1. Code definitions for Q8: Uncertainty**

*Imperfection of experiment*—Uncertainty characterizes the imperfectness or nonideal aspects of an experiment. One introductory algebra-based student explained that uncertainty analysis was a part of labs “because uncertainty is always present in experiments. It is because it is real world testing in comparison to theory which is perfect.”

*Data variability*—Uncertainty describes or quantifies the variations or fluctuations in data. An introductory algebra-based student explained, “Uncertainty is common in physics labs and experiments because it accounts for slight variations between the experiments being performed, allowing a range of ‘acceptable’ results. The uncertainty is used in the experiment to account for variations in your experimental setup.”

*Human mistakes*—Uncertainty characterizes the possible mistakes in an experiment. One introductory student said, “it is common because sometimes in a difficult experiment, scientists could make lots of mistakes when measuring.”

*Quantifying reliability*—Uncertainty helps determine the range of reliable data or to determine the “acceptable range of results.” One introductory student said, “Uncertainty analysis help show where the answer must fall in to be acceptable.”

*Making comparison*—Uncertainty is used to make comparisons between experimental results and theory or to compare results from different sources. One introductory algebra-based student explained, “Uncertainty is necessary when comparing the theoretical value to the experimental value because in the experimental value there are typically errors.”

*Refinement*—Uncertainty analysis is used to identify mistakes or to guide improvements of an experiment. One Ph.D. student said, “The uncertainty is a measure of the confidence in our data and results. If we analyze the uncertainty and find it to be quite high (relative to the quantity we are measuring), then we might have a problem with the apparatus or analysis (for example, might need to average over many trials or remove a source of measurement noise). In this way, the uncertainty analysis can be used to improve the experimental design and improve the resulting data.”

*Inherent aspect*—Uncertainty is inherent in any measurement, and the uncertainty analysis emphasizes an important aspect of teaching the empirical nature of science to students. One Ph.D. student said, “Scientific labs are mainly about measuring things in one form or another. Taking a ruler up to nature does not always yield numbers with the utmost certainty; therefore, when we study nature through labs, we need uncertainties to have any meaningful numbers.”

**2. Results and discussion for Q8: Uncertainty**

When responding to question 8 regarding the role of uncertainty analysis in physics labs, students’ reasoning can be placed into three major categories as shown in the left column of Fig. 3: *Representational*, *Inferential*, and *Teaching nature of science*. The use of the terms “representational” and “inferential” to describe the role of uncertainty analysis is adapted from statistics education [38]. The representational role refers to uncertainty and statistical concepts that are used to describe the features of data. The representational role of uncertainty is similar to the category of descriptive statistics which provides descriptions of a population or series of measurements (e.g., mean, standard deviation). The inferential role refers to concepts that are used to make inferences and conclusions about data (e.g., evaluate if a hypothesis is true or not). The inferential role of uncertainty is similar to the category of inferential statistics.

**Q8: Why is uncertainty analysis a common part of physics labs/experiments?**

Role of uncertainty	Codes	Subpopulation			
		Intro algebra	Intro calculus	Upper	Ph.D.
Representational	Imperfection of experiment	0.36	0.64	0.35	0.26
	Data variability	0.41	0.21	0.45	0.42
	Human mistakes	0.25	0.04	0.00	0.03
Inferential	Quantifying reliability	0.25	0.32	0.65	0.58
	Making comparison	0.16	0.21	0.50	0.39
	Refinement	0.01	0.07	0.05	0.13
Teaching nature of science	Inherent aspect	0.01	0.21	0.15	0.52

FIG. 3. Fraction of student responses with a particular code in the subpopulation (e.g., intro algebra-based physics). Students’ responses may be coded under several codes when appropriate so fractions do not add up to 1.00. Five levels of gray scale shading were applied based on the fraction,  $f$  (from lightest to darkest):  $f < 10\%$  (white);  $10\% \leq f < 25\%$ ;  $25\% \leq f < 50\%$ ;  $50\% \leq f < 75\%$ ;  $f \geq 75\%$  (black).



In the specific context of students' ideas about uncertainty analysis, the representational role of uncertainty describes features of the data or experiment (e.g., imperfections in the experiment or variability of data), while the inferential role focuses on how to use uncertainty to make inferences about data or experiment in a decision-making process (e.g., justifying the reliability of data or guiding the refinement of an experiment).

The representational role of uncertainty includes describing the imperfection or nonideal features of an experiment (linked to code imperfection of experiment), variation in the measured data (code data variability), as well as accounting for any possible human mistakes in an experiment. The inferential role of uncertainty includes quantifying reliability to determine the quality of the data, making comparison with theory or other measurements, and using uncertainty analysis for the refinement of the experimental design or process. The third category, teaching nature of science, is considered as a separate category because it emphasized that measurement uncertainty is an essential and fundamental aspect of how science progresses. Thus, these students felt it was important for them to appreciate the empirical nature of science beyond any particular representational or inferential roles.

The representational role of uncertainty was frequently addressed by all students, although it was the dominant role discussed among introductory students. In imperfection of experiment, many introductory students simply mentioned that the measurements or experiments are not perfect or the real world is not ideal, and other introductory students listed specific examples to illustrate the causes of the imperfection of experiment, such as "errors from machines" or "instrument error." Those students realized the existence of uncertainty in measurements, and many of them pointed out that uncertainty was inherent in experiments due to the difficulties in controlling real-world situations. One introductory student said, "there is always uncertainty because measurements won't be perfect." In general, imperfection of experiment accounted for a range of issues including the experimental design, insufficient control of factors, limitations of apparatus, and error in operations of apparatus.

Imperfection of experiment, which focused on the non-ideal nature of the experiment, was distinguished from human mistakes, which also appeared in several introductory algebra-based students' responses. Human mistakes included any possible user errors or mistakes made during an experiment. Mistakes were typically avoidable when experiments were performed in a careful and professional manner. One introductory student said, "Humans make errors, and any instruments has errors." Another student said, "We are not machines we can make a lot of error, that can affect the results." Accounting for human mistakes was a rationale primarily provided by introductory students and occurred only once among upper-level and Ph.D. students.

Finally, in the representational role category, data variability focused on uncertainty analysis as a way to describe fluctuations or variations in results. One example from an introductory algebra-based student was "There is error in everything we do so the uncertainty allows the data to fluctuate within a certain range and it still be accurate."

The inferential role of uncertainty was more frequently addressed by upper-level and Ph.D. students ( $p < 0.06$  when using a chi-square test to compare the codes quantifying reliability, making comparison, and refinement, across different levels of students). In quantifying reliability, students made an explicit connection between uncertainty and the quality of results. Many upper-level students associated uncertainty with the precision or accuracy of results although sometimes responses did not distinguish the two. One upper-division student said that uncertainty analysis "tells us something meaningful about the accuracy of our measurements. We use uncertainty to tell us how accurate our lab equipment is, and hence how accurate our final value is." Other students related uncertainty to the level of confidence they had in their data. One Ph.D. student explained, "The uncertainty is a measure of the confidence in our data/results. If we analyze the uncertainty and find it to be quite high (relative to the quantity we are measuring), then we might have a problem with the apparatus or analysis..."

Making comparison was another main theme in the inferential role of uncertainty analysis. Uncertainty is used to justify if an experimental result agrees with results from other sources and theoretical predictions or how well it fits a model. One example from introductory students was "uncertainty is necessary when comparing the theoretical value to the experimental value because in the experimental value there is typically errors." An example from an upper-level undergraduate student was "In our experiment uncertainty gives us a better chance of being able to fit our data to that of the accepted model, but is undeniably larger than that of any respectable institution (e.g., NIST)."

A small portion of Ph.D. students mentioned that uncertainty analysis can guide improvements or refinements to the experimental design or measurement techniques. One Ph.D. student said that uncertainty analysis "...gives meaning to number.  $5 \pm 6$  is meaningless in most cases;  $5 \pm 1$  is something. You need to know how well your measurement is defining the number you seek. *Large uncertainty can tell you something isn't working correctly or that the measurement method is not appropriate* [emphasis added]." Although refinement was not a frequently occurring code, it does demonstrate an additional inferential role of uncertainty analysis that is distinct from quantifying reliability or making comparisons.

Lastly, teaching the nature of science as a rationale for uncertainty analysis was almost exclusively addressed by Ph.D. students ( $p < 0.01$  when using a chi-square test to compare code teaching the nature of science across different

levels of students). In addition to discussing the representational and inferential roles of uncertainty analysis, Ph.D. students additionally explained that uncertainty was a fundamental aspect of any measurement, and students needed to appreciate this fact. Teaching uncertainty analysis conveys fundamental lessons about measurement and how science works. Many Ph.D. students emphasized uncertainty as “extremely important” or “as important as results.” For example, one Ph.D. student said “An experiment is essentially meaningless without an analysis of uncertainty...”

In students’ reasoning about the role of uncertainty analysis, introductory students tended to focus more on the representational roles of uncertainty (i.e., to describe the imperfection of experiment, variations in data, or mistakes made during experimental work). Students with more lab experience were more likely to discuss the inferential roles of uncertainty analysis in addition to the representational roles, which includes evaluating the reliability or giving confidence about the quality of data, making comparisons and, guiding refinements to the experiment to improve the quality of results.

Although uncertainty analysis is an important component in many introductory laboratory courses and detailed procedures are often provided for calculations of uncertainty (e.g., standard deviation of repeated measurements or uncertainty propagation), the ultimate purpose of those calculations may be hidden from many students. Our data indicate that students with less experience tended to emphasize representational roles of uncertainty such as describing imperfections and human error within an experiment and are less aware of the inferential and decision-making roles of uncertainty, such as justifying the quality of data and making comparisons with other results or theoretical calculations.

### C. Network analysis across questions

The analysis of students’ responses towards questions 7 and 8 gave us several insights into students’ views about the criteria for establishing trustworthy results as well as the rationale behind conducting uncertainty analysis in physics labs. However, it was common for students to invoke several different ideas within the same response, so we were curious to explore students’ overall reasoning patterns within and across questions by investigating the co-occurrence of the different reasoning categories regarding their views toward justification of experimental results (Q7) and their views about the role of uncertainty analysis (Q8). The specific questions we addressed were as follows: (i) How are the codes related within a particular subpopulation? (ii) How do the relationships between codes differ across subpopulations with varying levels of lab experiences? In order to answer those questions, we conducted a network analysis using the R programming language. We first exported a coding matrix from QSR International’s NVivo 11 qualitative analysis software, which shows

whether a code appears in a student’s response or not for all codes in Q7 and Q8. Then this coding matrix was imported and visualized in R as shown in Fig. 4.

A network plot enables the visualization of the structure of reasoning for each subpopulation as well as comparisons across subpopulations. The ideas and techniques are similar to those used in social network analysis. Social network analysis is primarily used to characterize social network structures in terms of nodes (e.g., individual people) and edges (i.e., links) that connect the nodes (e.g., relationships or interactions between people). The network analysis conducted here is similar to social network analysis because both are used to explore the relationships among objects. However, the primary difference is that the objects in social network analysis are people, while the objects in our network analysis are particular reasoning codes present within students’ responses.

Figure 4 shows students’ reasoning patterns across questions 7 (orange circles) and 8 (red circles). We plotted all codes as well as edges (i.e., connections between codes) for each subpopulation. The size of a node is proportional to how often the code appeared. A link between nodes was made whenever individual students displayed both codes within their responses. The thickness and darkness of a link is proportional to the fraction of students that mentioned both codes in their responses. However, only links that represented 10% or more of the subpopulation are shown in the plots. The cutoff was chosen to minimize the visual impact of a large number of very infrequent links between codes. Also, only codes that occurred in 10% or more of the subpopulation are labeled.

Network plot provides an alternative way to visualize the main results from Figs. 2 and 3 by evaluating the nodes within each subpopulation. Additionally, the links between codes form clusters, which may serve as a practical tool for exploring the structure of students’ epistemological beliefs. In Fig. 4, we can easily observe the complexity of students’ reasoning in terms of the distribution of codes and their connections. Students with more advanced lab experiences demonstrated epistemological reasoning that was more complex as evidenced by the larger number of nodes and increased number of connections. Introductory algebra-based students often provided only one single idea, while Ph.D. students tended to recall multiple ideas together. The average connectivity index (i.e., the average number of links) for each subpopulation is 3.2 for introductory algebra-based students, 5.4 for introductory calculus-based students, 7.2 for upper-level physics majors, and 9.2 for physics Ph.D. students.

When looking more closely at the connections between codes from questions 7 and 8, there are a few additional interesting results. First, introductory students’ responses rarely made a connection between the purpose of uncertainty and the validity of experimental results. The majority of introductory students did not discuss using uncertainty

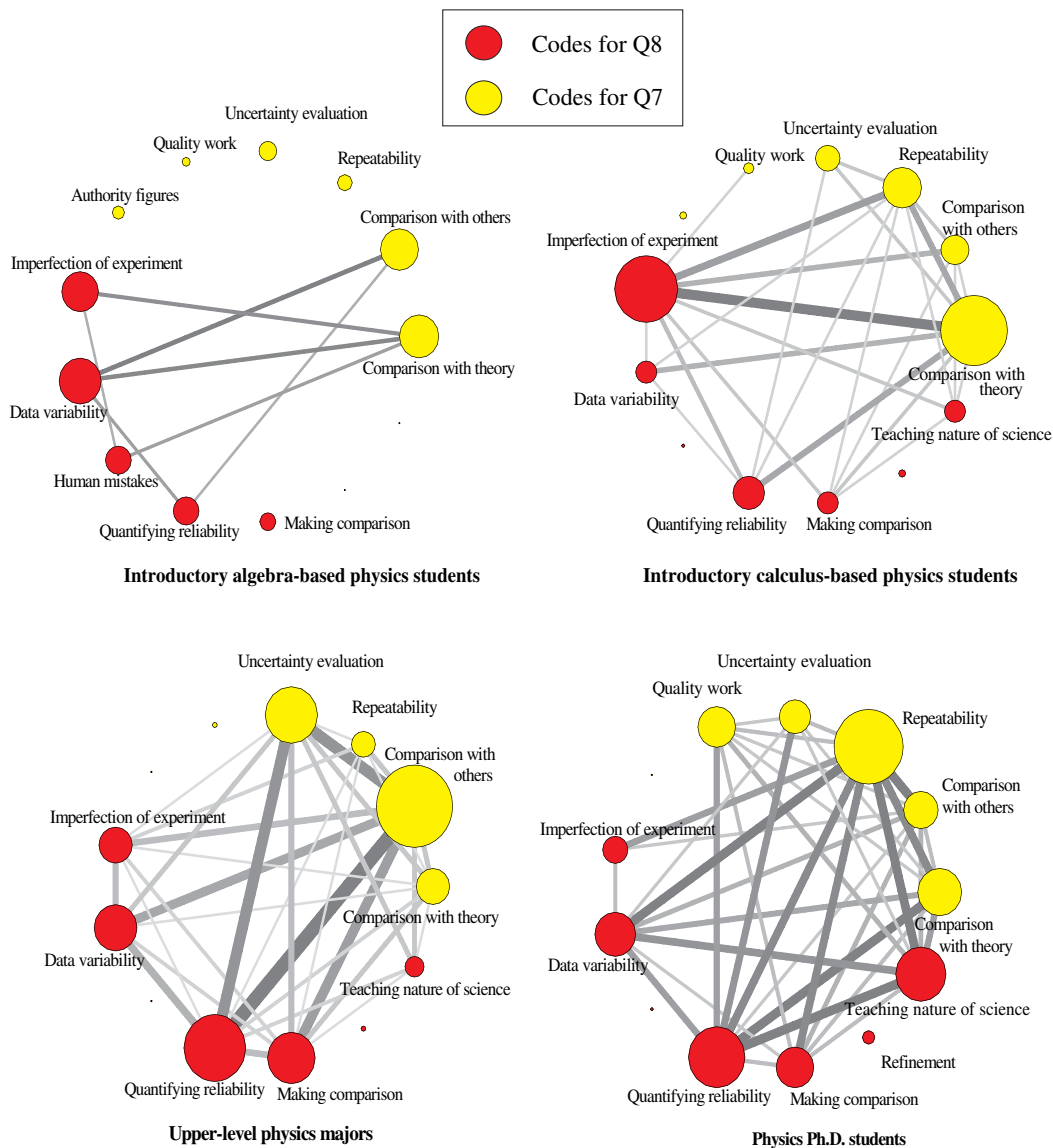


FIG. 4. Network analysis [Red circle represents codes for Q8, and yellow circle represents codes for Q7. The size of each circle is proportional to the percentage of responses (i.e., the frequency of code divided by the size of the subpopulation). The thickness of the edges is also normalized and proportional to the percentage of weights of edges (i.e., frequency of connections divided by the total number in that population). In each of the network plots, the gray scales of edges vary linearly depending on the weights, and darker lines correspond to greater weights].

to establish the validity of experimental results when responding to Q7, and they also did not recognize that one of the important goals of performing uncertainty analysis was to quantify the reliability of data.

Second, introductory algebra-based and calculus-based students rarely discussed the use of uncertainty analysis and statistical tools to support their conclusions despite often discussing the comparison of data to theory or the comparison of data to data from another source in order to establish the trustworthiness of their experimental results. The network plot shows there was rarely a connection between this “comparison” type of reasoning for establishing validity (from Q7) and the use of uncertainty analysis or

statistical tests to estimate reliability of the data and then make comparisons (from Q8). Comparisons were then being made without the use of formal quantitative statistical tools, such as uncertainty analysis. Rather the data suggest these introductory students tended to check how similar their own experimental results were to others’ results or theoretical predictions and then used uncertainty as a justification of any disagreement because it characterizes the nonideal feature of the physical world.

In contrast, both upper-level physics majors and Ph.D. students recognized quantifying reliability as the most important goal for uncertainty analysis. Additionally, upper-level physics majors linked the quantitative role of

uncertainty (code quantifying reliability) to data comparison in order to make conclusions about how similar their results are. Ph.D. students' responses also mentioned the quantitative role of uncertainty together with the code repeatability which often refers to checking the consistency of data through multiple trials.

## V. DISCUSSIONS AND IMPLICATIONS

In this study we developed an open-ended survey to probe students' personal epistemological views about the validity of experiments. The survey was given to students with various levels of physics laboratory experiences ranging from introductory physics students to physics Ph.D. students. We identified differences between introductory, upper-level, and Ph.D. students in the ways they justified the validity of experimental results.

When asked about their criteria for establishing trustworthy experimental results, introductory students almost exclusively discussed comparing their results with theoretical predictions or results of others. Ph.D. students in research labs utilized a range of approaches to establish the validity of experimental results (e.g., repeatability, comparison with theory, quality work, comparison with others, and uncertainty evaluation). A view of validity that over-emphasizes agreement with theory may distort the nature of science by implying that experimental results are the problem if they disagree with a highly idealized theory, rather than questioning the assumptions in the theoretical models as the problem. Further, a limited view of validity also provides students with a limited toolbox to evaluate their experimental results, especially if they encounter data about complex phenomena where there is little prior understanding. Implications for instruction could include designing labs where a variety of approaches are used to establish validity, including requiring careful documentation (e.g., having students hand off their work to another student who builds on it), understanding of the measurement tools (e.g., developing models of sensors and knowing their limitations), comparison with theory in limiting test cases (but not as the only means), repeatability using multiple trials or multiple approaches. Allowing students some freedom to develop experimental designs may also facilitate a discussion what makes the various approaches more or less trustworthy.

Another significant finding was that students at the introductory level rarely recognized uncertainty analysis as a tool to establish validity of results. Rather, they emphasized how it describes imperfections in the experiment or variations in data. Network analysis also showed few links between uncertainty analysis and validity. Some curricular designs may obscure the role of uncertainty as a means to quantify validity. Particularly, if the sole use of uncertainty is to compare with theoretical predictions (which are assumed to be true), it focuses students' attention on uncertainty as a way to quantify the impacts

of imperfections in the real world. An implicit message is that theory is perfect, while experiments are not. One way to shift students' attention from comparing data with existing theory is to develop experimental activities that go beyond known theoretical models (or at least known to students). For example, an exercise focused on experimentally exploring phenomena could ask students to find patterns or trends from data. In that kind of exploratory activities, uncertainty could be used as an effective tool to establish whether apparent trends are a result of random variation or more likely due to a real physical change in the behavior of the system. Uncertainty also be linked to validity in labs that require comparing sets of data in the absence of theory (e.g., which battery brand has the highest capacity) or to make decisions (e.g., which battery should be chosen to minimize the likelihood of a battery running out in a particular usage scenario).

Regarding Q8, we found that uncertainty plays several roles in physics labs: representational, inferential, and for teaching the nature of science. However, introductory students tended to ignore inferential roles, especially the use of uncertainty as a tool to quantify confidence in results, independent of theory. When teaching physics laboratory courses, uncertainty analysis could be used for more than answering a simple question about whether or not theory and experiment agree, although making such comparisons is an important role. Labs could instead provide opportunities to iterate, make repeated comparisons, and engage in decision making based on those comparisons. Uncertainty analysis is a means to establish and quantify a scientist's confidence in their experimental results, and it is part of a larger scientific process. Uncertainty analysis could be used as a tool for guiding the experimental process, which could include identifying when more data is needed, identifying sources of largest uncertainty that should be improved, and distinguishing between systematic discrepancies and random errors. One example by Holmes and others [37] uses an iterative cycle to help students engage in activities of making and acting on comparisons of their data (i.e., make a comparison, reflect on the comparison, and act on the comparison).

Regarding future research, we envision expanding our data set beyond cross-sectional snapshots of students involved in various stages of physics laboratory experiences to include longitudinal tracking and a larger span of lab course formats with larger sample size. This will likely require a modified survey format that is easier to analyze across a much larger population.

## ACKNOWLEDGMENTS

We thank Michael Rinkus and Kayla Adgate for assisting in the interrater reliability. This work was funded by National Science Foundation Grant No. DUE-1432301.

- [1] B.K. Hofer and P.R. Pintrich, The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning, *Rev. Educ. Res.* **67**, 88 (1997).
- [2] H. Siegel, What's in a name?: Epistemology, epistemology, and science education, *Sci. Educ.* **98**, 372 (2014).
- [3] M. Schommer, Effects of beliefs about the nature of knowledge on comprehension., *J. Educ. Psychol.* **82**, 498 (1990).
- [4] A. Elby, Getting started with research on epistemologies and expectations, *Getting Started in PER* (2011), Vol. 2, <http://www.per-central.org/items/detail.cfm?ID=10578>.
- [5] 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).
- [6] N. G. Lederman, F. Abd-El-Khalick, R. L. Bell, and R. S. Schwartz, Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science, *J. Res. Sci. Teach.* **39**, 497 (2002).
- [7] W. K. Adams, K. K. Perkins, M. Dubson, N. D. Finkelstein, and C. E. Wieman, The design and validation of the Colorado learning attitudes about science survey, *AIP Conf. Proc.* **790**, 45 (2005).
- [8] I. Halloun, Views about science and physics achievement: The VASS story, *AIP Conf. Proc.* **399**, 605 (1997).
- [9] E. F. Redish, Student expectations in introductory physics, *Am. J. Phys.* **66**, 212 (1998).
- [10] 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).
- [11] 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).
- [12] I. Halloun and D. Hestenes, Interpreting VASS Dimensions and Profiles for Physics Students, *Sci. Educ.* **7**, 553 (1998).
- [13] D. Hu and B. M. Zwickl, Examining students' personal epistemology: The role of physics experiments and relation with theory, in *Proceedings of the Physics Education Research Conference, 2017, Cincinnati, OH*, edited by L. Ding, A. Traxler, and Y. Cao (AIP, New York, 2017), pp. 11–14.
- [14] B. K. Hofer, Exploring the dimensions of personal epistemology in differing classroom contexts: Student interpretations during the first year of college, *Contemp. Educ. Psychol.* **29**, 129 (2004).
- [15] W. S. Parker, Franklin, Holmes, and the Epistemology of Computer Simulation, *Int. Studies Philos. Sci.* **22**, 165 (2008).
- [16] A. Franklin, *The Epistemology of Experiment*, edited by D. Gooding *et al.* (Cambridge University Press, Cambridge, England, 1989), pp. 437–460.
- [17] A. Elby, J. Frederiksen, C. Schwarz, and B. White, Epistemological beliefs assessment for physical science (EBAPS) (2003); Recuperado de <http://www2.physics.umd.edu/~elby/EBAPS/home.htm>.
- [18] D. Hammer, Epistemological beliefs in introductory physics, *Cognit. Instr.* **12**, 151 (1994).
- [19] D. B. May and E. Etkina, College physics students epistemological self-reflection and its relationship to conceptual learning, *Am. J. Phys.* **70**, 1249 (2002).
- [20] S. Duggan and R. Gott, What sort of science education do we really need?, *Int. J. Sci. Educ.* **24**, 661 (2002).
- [21] J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements* (University Science Books, Herndon, VA, 1997), p. 109.
- [22] D. L. Deardorff, Ph. D. thesis, North Carolina State University, Raleigh, NC, 2001.
- [23] I. Gal, Expanding conceptions of statistical literacy: An analysis of products from statistics agencies, *Stat. Educ. Res. J.* **2**, 3 (2016).
- [24] I. Gal, Adults' statistical literacy: Meanings, components, responsibilities, *Int. Stat. Rev.* **70**, 1 (2016).
- [25] J. Garfield and A. Ahlgren, Difficulties in learning basic concepts in probability and statistics—implications for research **19**, 44 (1988).
- [26] R. L. Kung, Teaching the concepts of measurement: An example of a concept-based laboratory course, *Am. J. Phys.* **73**, 771 (2005).
- [27] S. Allie, A. Buffler, B. Campbell, F. Lubben, D. Evangelinos, D. Psillos, and O. Valassiades, Cape Town, and South Africa, teaching measurement in the introductory physics laboratory, *Phys. Teach.* **41**, 394 (2003).
- [28] Subcommittee Membership, American Association of Physics Teachers, Technical Report, 2014.
- [29] S. Allie, A. Buffler, B. Campbell, F. Lubben, and L. Kaunda, First-year physics students' perceptions of the quality of experimental measurements, *Int. J. Sci. Educ.* **20**, 447 (1998).
- [30] J. Ryder and A. Clarke, Teaching and learning about "sources of error", on university physics courses, *A Research Study supported by the Learning and Teaching Support Network-Physical Sciences* (University of Leeds, Leeds, England, 2001).
- [31] R. Trumper, What do we expect from students' physics laboratory experiments?, *J. Sci. Educ. Technol.* **11**, 221 (2002).
- [32] T. 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).
- [33] R. L. Kung and C. Linder, University students' ideas about data processing and data comparison in a physics laboratory course, *Nordic studies in science education*, **2**, 40 (2006).
- [34] A. Buffler, S. Allie, and F. Lubben, Teaching measurement and uncertainty the GUM way, *Phys. Teach.* **46**, 539 (2008).
- [35] A. Buffler, S. Allie, F. Lubben, and B. Campbell, Introduction to measurement in the physics laboratory, *A Probabilistic Approach*, 3.4th ed. (2007), <http://www.phy.uct.ac.za/people/buffer/labmanual.html>.
- [36] J. Day and D. Bonn, Development of the concise data processing assessment, *Phys. Rev. ST Phys. Educ. Res.* **7**, 010114 (2011).
- [37] N. G. Holmes, C. E. Wieman, and D. A. Bonn, Teaching critical thinking, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11199 (2015).
- [38] A. Bakker and J. Derry, Lessons from inferentialism for statistics education, *Math. Thinking Learning* **13**, 5 (2011).