Assessing modeling in the lab: Uncertainty and measurement

N.G. Holmes¹ and Carl E. Wieman^{1,2}

¹*Department of Physics, Stanford University, Stanford, CA*

²*Graduate School of Education, Stanford University, Stanford, CA*

(Dated: August 20, 2015)

Abstract. Many introductory physics labs include goals related to learning about measurement uncertainty or error propagation. In this article, we present evidence from a new survey about experimentation and modeling in physics labs indicating that student difficulties with these concepts may stem in part from the language used. We demonstrate that students conflate measurement uncertainty, systematic effects, and measurement mistakes under a single umbrella. After a course that explicitly distinguished these three terms, students' paid more attention to precision and sources of random variability, rather than systematics or experimenter mistakes. We use this preliminary analysis to evaluate the survey as an instrument to assess learning in physics labs.

PACS: 01.40.-d, 01.40.G-, 01.50.Qb

I. INTRODUCTION

With the American Association of Physics Teachers' list of recommended goals for the physics lab curriculum [1] comes a need for assessment. Instructional lab programs require validated instruments to assess whether they are meeting these goals, few of which exist at present. The "PhysPort" website [2], an online resource to support physics instructors with research-based teaching resources, has amassed 47 validated assessment instruments for physics courses. Only one of them, the Concise Data Processing Assessment [3], evaluates lab skills.

In this paper, we present preliminary work and results on the development of a lab skills survey. The assessment targets students' understanding of modeling experimental systems. We define modeling in this context as used by Zwickl, Finkelstein, and Lewandowski [4], where an experiment involves models of a measurement and a physical system. The measurement system model includes understanding the experimental design and sources of uncertainty or systematics in that design. The physical system model includes understanding systematic effects due to invalid approximation or assumptions. Engaging with the models can include designing the experiment to reduce the effects of uncertainty or address systematics, to analyze data, and to use the data to further inform the measurement or physical system models. The survey currently includes 10 questions that probe students' understanding of the full experimental system model.

In what follows, we focus on student responses to the first two questions, which probe experimental design and sources of uncertainty. We study how students combine mistakes, systematic effects, and random variability when asked to list sources of uncertainty. They also use a variety of terms related to reasons the measurement is erroneous, not correct, or inaccurate, even when instructions include no reference to the word 'error'.

Research has consistently demonstrated student difficulties with and misconceptions about measurement uncertainty [5–10]. Interpretations of uncertainty have been

classified into a point and a set paradigm [11]. Under the set paradigm, any single measurement is believed to be an estimate of a physical quantity and deviation between measurements is random. Measurements are, therefore, characterized through probability distributions. Under the point paradigm, importance is placed on individual data points, such that any measurement could be the true value. This leads students to interpret measurement error (when meant as uncertainty) to characterize why the measurement differs from an expected value [6].

In labs, the point paradigm is reinforced by the percent error equation or through 'closeness' comparisons, neither of which take into account the uncertainty in the measurements [8, 9]. This interpretation of 'error' as a measurement mistake leads students to believe that measurement error or uncertainty could be reduced to zero [7, 11, 12]. These misconceptions have broader implications for students' understanding of the nature of science.

The work done so far on categorizing students' understanding of measurement uncertainty has improved how we teach these as statistical concepts, rather than mathematical procedures [5, 6, 8]. These studies used teaching methods that focused on the probabilistic nature of measurement variability and were shown to increase the number of students who attend to uncertainty when describing measurements with variability or comparing measurements with uncertainty. This focus on uncertainty as a statistical concept, however, does not necessarily give physical purpose to the analysis or relate it to the scientific concepts and models being explored [10]. Many lab courses still encourage students to list all the possible sources of error that may have impacted their measurements at the end of a lab report, regardless of the nature of the error (someone misread a meter versus the probabilistic nature of radioactive decay, for example), thereby implying equivalence. These lists are often treated by students as reasons for why measurements are different from a true, theoretical, or textbook value, rather than demonstrating the limits of confidence specific to the physical measurement.

In this paper, we explore this physical interpretation

of measurement uncertainty, rather than the statistical nature. We do not aim to suggest replacing the statistical treatment of uncertainty; rather, we argue for the need for a physical treatment in addition.

II. METHODS

Data were collected using an instrument under development to explore students' ideas about experimental system models. The data presented in this paper were collected as part of an early validation process.

The survey uses a cover story of two groups of students making measurements of the period of oscillation of a spring to evaluate the (given) model:

$$T = 2\pi\sqrt{\frac{m}{k}}. \quad (1)$$

The survey lists several assumptions that this model makes, as well as the available equipment, and explains that the students used a stopwatch to time several oscillations of the spring. The first two questions are:

1. *What are possible sources of uncertainty for measuring the time for a certain number of bounces? List as many as you can think of.*
2. *What are some ways to reduce the impact of this uncertainty on their measurements?*

and are followed by 8 more questions that probe students' interpretation of sample methods and collected data.

The survey was given to 148 students during the first and last week of a laboratory course associated with an introductory, calculus-based electricity and magnetism course at a large, research-intensive institution. The majority of students were science and engineering majors and 50% of them were in their first-year of study. Results from a pre-test using the Conceptual Survey of Electricity and Magnetism [13] demonstrated that students in the lab course had a stronger background in electricity and magnetism than typical students in equivalent undergraduate physics courses at large selective public institutions at pre-test: this group, $M = 50\%$, $SD = 15\%$, $n = 125$; comparison group 1, $M = 21\%$, $SD = 12\%$, $n = 389$ [13]; comparison group 2, $M = 32\%$, $SD = 10\%$, $n = 168$ [14]. This suggests that the students were above average in their physics preparation, though it is unclear how this would translate to their lab performance.

The lab course included explicit learning goals related to developing an understanding of uncertainty and measurement. In each experiment, students were asked to list and quantify all sources of uncertainty in their measurements. There were also learning activities related to separately defining sources of uncertainty (random variability), systematic effects (invalid assumptions of models, for example), and measurement mistakes (including incorrectly calibrating equipment).

Student responses to the first two questions were open coded and codes were refined as common categories emerged. The total number of sources of uncertainty

(question one) and methods for reducing uncertainty (question two) were determined for each student. We also evaluated whether the methods to reduce uncertainty mapped onto their listed sources of uncertainty. Finally, student responses were separately coded for five key error-like terms: 'human error', 'accuracy', 'exact' (or 'inexact'), 'incorrect' (or 'not right'), and 'error' (other than when used in human error).

III. RESULTS

The common response categories at pre- and post-test are shown in Figures 1a (sources of uncertainty listed for question 1) and 1b (methods to reduce uncertainty for question 2), along with the fraction of students making comments in each category.

A. Interpretations of uncertainty versus error

At both pre- and post-test, most students listed uncertainty sources due to timing, though their descriptions of this uncertainty varied widely. At pre-test, many of these sources related to reasons the measurement was incorrect (e.g. inaccurate pressing of the stopwatch, human error, delays from reaction time), rather than varying randomly. At post-test, many more students listed the precision of the instruments (i.e. number of decimal places) as a major source of uncertainty.

These issues are further illuminated by examining students' use of error-like language (Figures 2a and 2b). 77% of students used one or more of these key terms when listing their sources of uncertainty at pre-test. This decreased to 56% of students doing so at post-test. Students used far fewer error-like terms when listing their methods for reducing the sources of uncertainty (21% at pre-test and 17% at post-test).

B. Connecting uncertainty sources with physical procedures

The average number of sources of uncertainty and methods to reduce the sources of uncertainty are given at pre- and post-test in Table I. Students at post-test listed significantly more sources of uncertainty than at pre-test. We also coded for whether students' sources of uncertainty mapped onto the methods to reduce them. At pre-test, 34% of students listed methods in Q2 to reduce each of their uncertainty sources, while 45% of students listed uncertainty sources in Q1 that did not have corresponding methods to reduce them in Q2. In contrast, only 8% of students had additional methods in Q2 that did not correspond to uncertainty sources listed in Q1. At post-test, 27% of students wrote methods in Q2 to reduce each of their uncertainty sources in Q2, 73% listed additional sources in Q1 that did not have corresponding

(a) Fraction of students writing specific sources of uncertainty for Q1 at pre- and post-test.

(b) Fraction of students writing specific methods to reduce uncertainty for Q2 at pre- and post-test.

FIG. 1: Most common emergent codes for students' sources of uncertainty and methods.

(a) Frequency of error-like key terms on Q1

(b) Frequency of error-like key terms on Q2

FIG. 2: Student's use of error-like terms

methods to reduce them in Q2, and 27% had additional methods in Q2 that did not correspond to uncertainty sources listed in Q1.

TABLE I: Average number of sources of uncertainty (Q1) and methods to reduce uncertainty (Q2) listed by students at pre- and post-test.

	Sources of uncertainty	Methods
Pre-test	$M=2.72, SD=1.07$	$M=1.95, SD=1.00$
Post-test	$M=3.12, SD=1.14$	$M=2.02, SD=0.96$
	$t(276.87)=3.05, p=.003^{**}$	$t(282.28)=0.59, p=.554$

M = mean; SD = standard deviation; $^{**} p < .01$

Students may have grasped the procedural or statistical ways to improve measurements, since many students recognized the benefit of conducting multiple trials or measuring over several oscillations. At pre-test, 46% of students listed trials, and this increased to 61% at post-test. Most of the students who suggested trials, however, did not include any description as to why trials or repeated measurements would reduce uncertainty. Since the instrument did not ask students to explain or justify their proposed methods, it is unclear whether students had learned a rote procedure or whether they understood how it would improve their various sources of uncertainty. The second most common method to reduce uncertainty, however, was to do a completely different experiment (with automated timing and release system or video analysis at high frame rates). At pre-test, 32% of students suggested this method. After the lab course with explicit

goals related to understanding measurement and uncertainty, only 15% of students proposed this as a method to reduce uncertainty, paying more attention to repeated trials and improving the precision of the instruments.

IV. DISCUSSION

In this paper, we have described preliminary analysis from the first two questions on an instrument for assessing modeling in physics labs. We see that students define sources of uncertainty both as sources of random variability and reasons measurements are not exactly equal to a theoretical value. Students often list more sources of error and uncertainty than methods to deal with them. These results are further evidence that students list sources of uncertainty as catch-all lists of excuses for why measurements are not perfect [10].

From these data, we propose that the language used to describe measurement uncertainty and variability (the word ‘error’) may be a cause of students’ novice, point-like interpretations of measurement seen elsewhere [11]. While the word ‘uncertainty’ suggests there is a degree of limitation to how well we know a measurement, ‘error’ suggests a mistake. When ‘error’ is used synonymously with ‘uncertainty’, students may interpret their measurement uncertainty as reasons that their experiment is not perfect, reinforcing point-like thinking.

Using the word ‘uncertainty’ instead of ‘error’ has been previously recommended for instructional physics labs in the past [5, 7] and for reporting scientific data in general [15]. Distinguishing uncertainty from error (as mistakes), however, leaves ambiguity with regards to systematic errors. A modeling framework for advanced physics lab activities describe systematic errors as invalid assumptions or approximations about the physical or measurement system models [4]. This definition is somewhat dis-

tinct from the notion of measurement mistakes such as incorrectly calibrating an instrument or reading a measurement device incorrectly. Instead, this definition focuses more on understanding the assumptions being made about a physical or measurement system.

As a whole, this motivates three distinct terms to reflect three distinct concepts, as used in the study here: *Uncertainty* due to inherent limitations and random variability of a measurement; *Systematic effects* due to assumptions or approximations about a physical or measurement model being invalid for a given experiment; and *Measurement mistakes* in the measurement process.

This preliminary analysis also informs the development of the lab skills survey. The first two questions will be expanded into a larger set of questions about experimental design, probing systematic effects and sources of uncertainty separately. It will also ask for explanations linking the methods with the sources of uncertainty. Further work will evaluate the subsequent questions in detail similar to the analysis done here. From the open-responses, distributed to more students at multiple institutions, a set of multiple choice options will be generated for each question. Interviews will be carried out with students and experts with the multiple choice format. We plan to use a test structure similar to the Coupled Multiple Response format used in ref. [16], where items probing answers to the what and why are paired, and consistency between responses is evaluated.

ACKNOWLEDGEMENTS

We would like to thank the following people for insightful contributions to the discussion about these ideas and concepts: S. Allie, P. Barmby, J. Heafner, H.J. Lewandowski, B.J. Ramshaw, A. Rundquist, C. Rutherford, and B. Vancil.

-
- [1] American Association of Physics Teachers, Tech. Rep., American Association of Physics Teachers (2014).
 - [2] URL www.physport.org/assessments/.
 - [3] J. Day and D. Bonn, Phys. Rev. Spec. Top.-PH **7**, 010114 (2011).
 - [4] B. M. Zwickl, N. Finkelstein, and H. J. Lewandowski, Am. J. Phys. **82**, 876 (2014).
 - [5] S. Allie, A. Buffler, B. Campbell, F. Lubben, D. Evangelinos, D. Psillos, and O. Valassiades, Phys. Teach. **41**, 394 (2003).
 - [6] A. Buffler, S. Allie, and F. Lubben, Phys. Teach. **46**, 539 (2008).
 - [7] D. Evangelinos, D. Psillos, and O. Valassiades, in *Teaching and Learning in the Science Laboratory*, edited by D. Psillos and H. Niedderer (Springer Netherlands, 2002), vol. 16, p. 179.
 - [8] R. L. Kung, Am. J. Phys. **73**, 771 (2005).
 - [9] M.-G. Séré, M. Fernandez-Gonzalez, J. A. Gallegos, F. Gonzalez-Garcia, E. D. Manuel, F. J. Perales, and J. Leach, Res. Sci. Educ. **31**, 499 (2001).
 - [10] M. Séré, R. Journeaux, and C. Larcher, Int. J. Sci. Educ. **15**, 427 (1993).
 - [11] A. Buffler, S. Allie, and F. Lubben, Int. J. Sci. Educ. **23**, 1137 (2001).
 - [12] J. Leach, R. Millar, J. Ryder, M.-G. Séré, D. Hammelev, H. Niedderer, and V. Tselfes, Working Paper 4, European Commission - Targeted Socio-Economic Research Programme (1998).
 - [13] D. P. Maloney, T. L. O’Kuma, C. J. Hieggelke, and A. Van Heuvelen, Am. J. Phys. **69**, S12 (2001).
 - [14] S. Pollock, in *Phys. Educ. Res. Conf.* (AIP Conf. Proc., 2008), p. 171.
 - [15] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML, Tech. Rep., Organization for Standardization (2008).
 - [16] B. R. Wilcox and S. J. Pollock, Phys. Rev. Spec. Top.-PH **10**, 020124 (2014).