

# Developing quantitative critical thinking in the introductory physics lab

N.G. Holmes

*Stanford University, Department of Physics,  
382 Via Pueblo Mall, Stanford, CA, 94305*

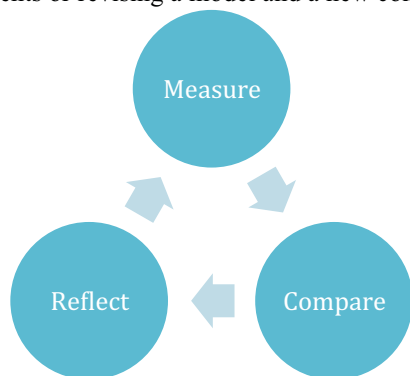
**Abstract.** While the goals for instructional labs have been highly debated, there is consensus that labs have unique potential to develop students' scientific and experimentation abilities. I present a simple scaffold for introductory labs that uses iterative cycles of comparisons (either between data sets or between data and models) to develop students' epistemologies, experimentation behaviors, and critical thinking abilities. By focusing the iterations on improving measurements, students explore the limits of physical models in the real world and engage in the evaluation and refinement of these models. In a controlled research study, students adopted these behaviors and continued to use them even after instruction to do so had been removed.

**PACS:** 01.40.G-, 01.40.-d, 01.55.+b, 01.50.Qb

## I. INTRODUCTION

Thinking critically about data is an important skill for scientists and non-scientists alike, whether to make sense of published scientific studies, conclusions drawn by the media, or everyday events. Instructional labs, where students work with data to make sense of physical systems and models, are uniquely suited to provide experiences developing these skills. The AAPT's recently endorsed set of lab goals [1], however, demonstrate the complexity of the learning activities involved in physics labs: modeling, constructing knowledge, developing technical and practical skills, designing experiments, analyzing and visualizing data, and communicating physics. Critical thinking skills are necessary to make progress with and connect across all of these elements. In this paper, I will present a simple pedagogical framework that focuses on critical thinking with data and leads to significant improvements in each of these areas, as well as in their thinking and reasoning.

We define critical thinking by three key behaviors (Fig. 1): First to take measurements and then compare them, either internally between the measurements or to a model; Then to reflect on and interpret this comparison; and, finally to decide on a follow-up action, leading one to new measurements or revising a model and a new comparison.



**FIG 1.** Three step comparison cycles used to structure critical thinking behaviors.

## II. EXAMPLE I: LR CIRCUITS EXPERIMENT

In an introductory physics lab course, students complete a circuits experiment with an inductor ( $L$ ) and a resistor ( $R$ ) in parallel with a square-wave alternating current power source. They measure the voltage across the resistor as a function of time to test the model that the time constant ( $\tau$ ) for the voltage decay across the resistor goes as:

$$\tau = L/R. \quad (1)$$

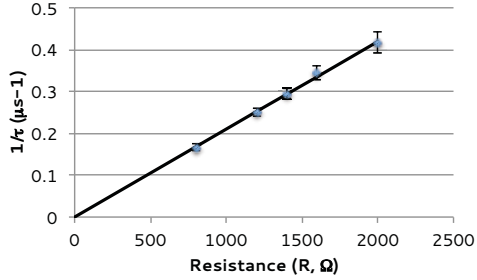
From this, a graph of  $1/\tau$  versus the resistance,  $R$ , should produce a straight line with an intercept through the origin.

Figure 2a shows measurements made by a student during such an experiment with the given model, with a one-parameter linear fit with the intercept fixed through the origin drawn as the solid line. Comparing the model and the data, the line appears to be a good fit to the data. One may argue, however, that the data provides little support for the choice of a fixed intercept through the origin, since there is no data near the origin. A good way to act on this comparison, then, is to take additional data at smaller values of  $R$ , as in Fig. 2b.

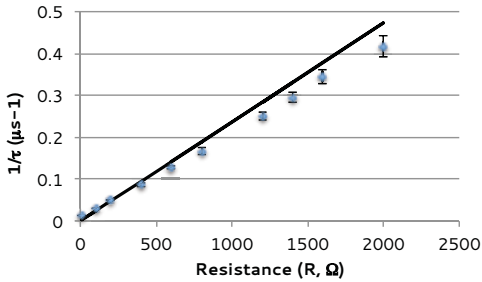
It is clear from this additional data that  $\tau$  is approaching a non-zero intercept, which suggests that the one-parameter model (solid line) may be invalid. This leads to a new action, to fit the data to a two-parameter model with the intercept as a free parameter (Fig. 2c). This new model is, visually, a much better fit to the data, which leads to a new reflection. What does this non-zero intercept mean? Did the student make a mistake when measuring? Was the equipment mislabeled? Is there a problem with the model? A number of follow-up actions can lead the student to discover that the issue is related to the physical system model: that there is additional resistance in other components of the circuit.

From this sequence of comparisons, reflections, and actions, the student explores the limits of the physical and measurement models [2], thinking critically about the experimentation process. However, there was a critical first step. The first data set and fit looked quite good and

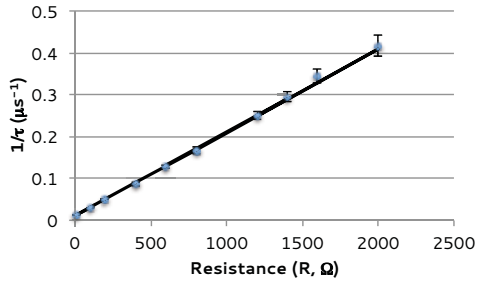
verified the given model. Given that students rarely choose to repeat or improve measurements [3,4], how do we motivate them to challenge their results and dig deeper? One option is to tell them what data to take and to direct their attention to the issue. Such cookbook tasks, however, transfer poorly to different contexts. A better option is to structure (or scaffold) their inquiry behaviors to lead them to discover it (seemingly) on their own.



(a) Initial data set fit to a one-parameter model



(b) Second data set fit to a one-parameter model



(c) Second data set fit to a two-parameter model

**FIG 2.** Sample data sets and fits for an LR circuits experiment show plausible progressions when cycling through comparisons and iterations.

### III. QUANTITATIVE SCAFFOLDS

We provide students with a scaffold that is based in the quantitative, statistical tools that would be used to make comparisons. The first such tool is an index comparing two measurements with uncertainties:

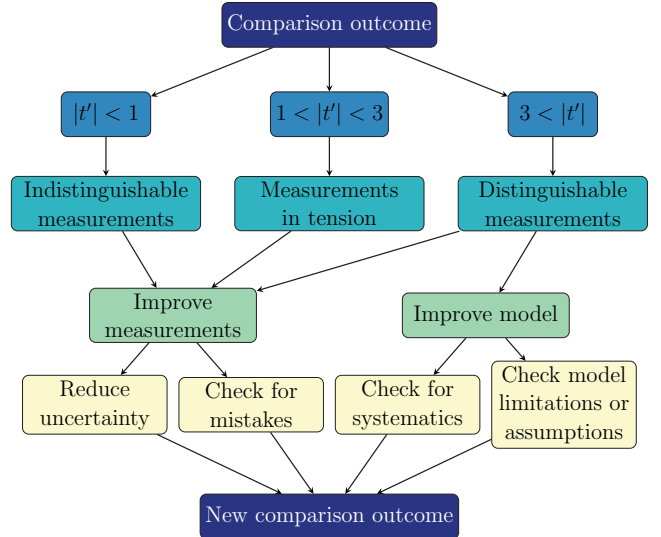
$$t' = \frac{A-B}{\sqrt{\delta_A^2 + \delta_B^2}} \quad (2)$$

where  $A$  and  $B$  are measured values and  $\delta_A$  and  $\delta_B$  are their uncertainties, respectively. This index is the difference between the two measurements in units of uncertainty. It

replaces the more common comparison to check whether the uncertainty ranges of two measured values overlap, with which students often have difficulties [5]. Instead of a binary comparison (either the ranges overlap, meaning agreement, or do not overlap, meaning disagreement), the  $t'$ -score puts the comparison on a continuous scale (to what degree are the measurements distinguishable). This is also a more set-like treatment for comparisons [6], where set-like thinking is the notion that any single measurement is an estimate of the value being measured, and that there is an associated probability distribution that characterizes the measurement. This is in contrast to the undesirable point paradigm, where special importance is placed on single data points, often without attributing an uncertainty range or probability distribution to any measured value.

Beyond quantifying the comparison, reflecting on the  $t'$ -score supports the critical thinking cycle (Fig. 1). Values less than one may suggest agreement, but can also suggest over-estimated uncertainties (a large denominator). Values much greater than three suggest a disagreement. A decision tree can then be built from these interpretations (Fig. 3).

If students obtain a small  $t'$ -score, one follow-up action is to design a way to improve the quality of their data (especially to reduce their uncertainty). If students obtain a large  $t'$ -score, they check for mistakes in their data or methods, or search for systematic effects, or evaluate their model to identify limitations or assumptions that may not be valid. This set of interpretations and actions can be mapped onto our comparison cycles (Fig. 1), providing support for the decision making process. It also provides opportunities for students to design their own experiments, which has been shown to improve students' scientific abilities [7]. Early in the course, this is constrained to designing improvements to measurements, rather than designing wholly new experiments.



**FIG 3.** Interpretations of and follow-up behaviors from a  $t'$ -score comparison between two measurements.

## IV. EXAMPLE II: PENDULUM EXPERIMENT

Students in our introductory physics lab course are given a 1-m long, high-efficiency pendulum, a protractor, and a stopwatch. They are asked to measure the period of the pendulum at an angle of amplitude of  $10^\circ$  and  $20^\circ$  and compare the two measurements using a  $t'$ -score. They have seen the period ( $T$ ) defined as:

$$T = 2\pi \sqrt{\frac{L}{g}} \quad (3)$$

where  $L$  is the length of the pendulum, and  $g$  is the acceleration due to gravity. Equation 3 suggests that the period is independent of the angle of amplitude.

While this seems fairly ‘cookbook’ at first glance, we give the students no instructions on *how* to measure the period of the pendulum. In effect, we tell them to make a quiche, not *how* to make a quiche. They have to decide how many trials to take, how many oscillations to measure in each trial, how to control the angle of amplitude, and how to assign uncertainties to their measurements. After an initial comparison, students are told to iterate according to the decision tree in Fig. 3, but, again, are not told how.

It may be clear to the reader, though it is certainly not clear to introductory physics students, that in deriving Eqn. 3, one must make a small angle approximation. The size of the effect from this approximation can be estimated as follows. The first-order correction to the small angle approximation would give:

$$T = 2\pi \sqrt{\frac{L}{g}} \left(1 + \frac{1}{16} \theta_0^2\right) \quad (4)$$

where  $\theta_0$  is the angle of amplitude. A 1-m long pendulum would have a period of oscillation around 2s, and the  $10^\circ$  and  $20^\circ$  amplitudes would differ by 0.01s. This means that the uncertainty in measurements of the period need to be less than 0.0027s (relative uncertainty less than 0.14%) in order to discern this difference at the  $t' > 3$  level.

When using this experiment in our course (see [8] for more details), most students' initial methods involve taking 10 trials of single-swing measurements. This gives an uncertainty around 0.03s (with stopwatch timing uncertainty around 0.1s), a 1.6% relative uncertainty for a 1-m long pendulum, and a  $t'$ -score of 0.26. Through the explicit instructions to iterate using the  $t'$ -score decision tree, 70% of the students designed a measurement that reduced their uncertainty and 64% of students used measurements of multiple swings to do so. Their final measurements involved, on average, 11 trials of 9-swing measurements, which gives an uncertainty of 0.003s and a  $t'$ -score around 2.5, in the ‘tension’ area.

Only 11% of the students identified a disagreement between the two measurements. Several students even explicitly described their iteration designs as attempts to decrease their  $t'$ -scores to obtain better agreement. The following week, however, we told students to make measurements at several angles between  $5^\circ$  and  $25^\circ$ , which allowed them to visualize the quadratic second-order

behavior of the amplitude dependence. In this case, 70% of students identified the disagreements between their measurements and 12% related it to the small angle approximation.

As in the LR experiment (Sec. II), an initial, low-quality measurement can validate the simplified model seen in class. A high-quality data set is needed to reveal its limitations. The decision tree, based uniquely on a student’s measurements and comparisons, can push them towards better measurements to a point where they construct new knowledge. The pendulum experiment results, and other studies, show us that they need extensive and deliberate practice [9] within the framework, however, to be comfortable reflecting on and evaluating authoritative models [10].

## V. EXTENDING THE FRAMEWORK

Another common comparison in physics is to fit data to models, such as in the LR experiment described earlier. The appropriate tool for this is the  $\chi^2$  value in least-squares fitting. It is structurally similar to the  $t'$ -score in that it, too, is an index in units of uncertainty:

$$\chi^2 = \frac{1}{N} \sum_{i=1}^N \left( \frac{y_i - f(x_i)}{\delta y_i} \right)^2 \quad (5)$$

where  $N$  is the number of data points,  $x_i$  and  $y_i$  are the measured values,  $f(x_i)$  is the model value at  $x_i$ , and  $\delta y_i$  is the uncertainty in  $y_i$ . The decision tree in Fig. 3, therefore, directly maps onto least-squares comparisons, though the cut-off for disagreement would be at a  $\chi^2$  value of 9 (rather than 3). This supports the use of the cycles framework in a multitude of physics lab experiments and courses, especially since  $f(x_i)$  can be any model form.

After sufficient practice with these quantitative tools and explicit iterations through the cycle, the framework behaviors can be less explicitly scaffolded. Students should eventually iterate and revise their methods and models continuously, with both quantitative tools (such as  $t'$ -scores and  $\chi^2$  values) and qualitative tools (such as the visual analysis described in Sec. II or evaluating residual plots).

## VI. EVALUATING STUDENT BEHAVIOURS

The effectiveness of the cycles framework and quantitative decision tree scaffold was recently evaluated in an 8-month long controlled study (see [11,12] for further details). Here we summarize the results. In-lab behaviors of an experimental group of students ( $n=133$ ) who were taught using the framework were compared with those of a control group who were not ( $n=131$ ). The two groups were composed of students registered in two different cohorts of the same introductory physics lab course. Both groups worked with the same set of physics experiments and learned a similar set of analysis tools (the control group compared measurements using overlapping uncertainty ranges, while the experiment group used  $t'$ -scores; both

groups worked with  $\chi^2$  values). While both groups regularly made comparisons between data sets or between data and models, only the experiment group worked with the critical thinking cycles (Fig. 1) and the decision tree (Fig. 3). In early experiments, they were given explicit instructions to move through the stages of the cycles. This structure was slowly faded until their instructions were identical to those of the control group. Towards the end of the year, students in both conditions were given an unstructured version of the LR circuits experiment, described in Sec. II. Their in-lab behaviors were evaluated through their written lab notes.

In the control group, only 2% of students proposed or made changes to their measurements during the experiment. Consistent with other literature [3,4], we see that students rarely repeat or improve measurements on their own. In contrast, 69% of students in the experimental group made or proposed making such changes. The instruction to iterate in earlier labs became habit to the experimental group such that they continued to do so on their own.

Obtaining a high quality data set in this experiment should lead students to notice the need for an intercept in the fit. In the control group, 18% of students identified the need for an intercept and interpreted it physically as extra resistance in the circuit. In the experimental group, 77% of students identified the need for the intercept, with 52% also interpreting it physically. That is, more than 4 times as many students reflected critically on their data to identify a disagreement with a model, and nearly 3 times as many interpreted the analytic parameter (intercept) physically.

This demonstrates a significant difference in students' behaviors and task orientation. Students who had engaged in the iterative comparison cycles were thinking critically about their data and the models and engaging in more expert-like scientific behaviors. These students had also begun using their data to make claims about the validity of authoritative models. Given that few students in the control group made this claim at the end of the course, this suggests a shift in students' epistemological frames. That is, they had shifted their ideas about how they, as students, can generate scientific knowledge in the lab. This also has implications for their identities as scientists, which deserves a much more detailed examination than there is room for in this

paper (see [12]). Students in the experimental group used creativity, ownership, and scientific habits of mind by designing and carrying out ways to improve their data without instruction to do so.

## VII. DISCUSSION

In this paper, I have presented a relatively simple pedagogical framework for structuring students' critical thinking in introductory physics labs. In the introduction, I claimed that focusing instruction on developing critical thinking skills could result in gains in a variety of other lab-specific learning goals recommended by the AAPT [1]. Scaffolding the critical thinking process when interpreting quantitative comparisons provides clear opportunities to develop students' skills at *analyzing and visualizing data*. The focus on improving measurement uncertainty helps to develop students' understanding of the measurement tools they are using, including the limitations of devices, as related to *technical and practical skills*. This focus also results in opportunities to break down the elements of *experimental design*, thinking critically about how small changes in measurement procedures affect outcomes. Through the iterations, students explicitly explore the limitations and assumptions in *models*, even engaging in opportunities to *construct knowledge* that is new to them. When they make claims that a model is insufficient, they must clearly *communicate* and articulate how their data supports that claim.

Through a controlled study, this framework was shown to be highly effective at engaging students with iterating and improving their measurements and evaluating limitations and assumptions of models. Beyond the in-lab behaviors, this framework also supports a number of epistemological and affective shifts as students engage in a variety of scientific activities.

## ACKNOWLEDGEMENTS

All work presented in this paper was done in collaboration with Doug Bonn. I would also like to acknowledge support from Carl Wieman and Heather Lewandowski.

- [1] AAPT. Technical report, 2014.
- [2] B.M. Zwickl, N. Finkelstein, and H.J. Lewandowski. *Am. J. Phys.* **82**(9), 876 (2014).
- [3] M.-G. Séré, R. Journeaux, and C. Larcher, *Int. J. Sci. Educ.* **15**(4), 427 (1993).
- [4] Z. Kanari and R. Millar, *J. Res. Sci. Teach.* **41**(7), 748 (2004).
- [5] R.F. Lippmann, Ph.D. Thesis, University of Maryland, 2003.
- [6] A. Buffler, S. Allie, and F. Lubben, *Int. J. Sci. Educ.* **23**(11), 1137 (2001).
- [7] E. Etkina, A. Karelina, and M. Ruibal-Villasenor, *Phys. Rev. ST-PER* **4**(2), 020108 (2008).
- [8] N.G. Holmes and D.A. Bonn, *Phys. Teach.* **53**(6), 352 (2015).
- [9] K.A. Ericsson, R.T. Krampe, and C. Tesch-Romer, *Psych. Rev.* **100**(3), 363 (1993).
- [10] B. Ibrahim, A. Buffler, and F. Lubben, (2009). *J. Res. Sci. Teach.* **46**(3), 248 (2009).
- [11] N.G. Holmes, C. Wieman, and D.A. Bonn, *P. Natl. A. Sci.* **112**(36), 11199 (2015).
- [12] N.G. Holmes, Ph.D. Thesis, University of British Columbia, 2015.