Partially supported by Grant No. 1611482

Assessment of Critical Thinking in Physics Labs

Cole Walsh, Katherine N. Quinn & N. G. Holmes
Laboratory of Atomic & Solid State Physics, Cornell University, Ithaca, NY, 14853

Physics Lab Inventory of Critical thinking (PLIC)

- Two groups conduct an experiment to test the model: $T = 2\pi\sqrt{\frac{m}{k}}$.
- Likert-style questions ask respondents to evaluate how well data agrees with a model or how well a particular group tested the model.
- 'Select all that apply' questions ask respondents to elaborate on their reasoning to the Likert-style questions and to suggest what the group should do next.
  - There are between 5-10 options to choose from, and students are limited to selecting no more than three options.

Group 1

- Measure 10 repeated trials
- Use 2 masses
- Calculate k in each case and compare

Students described "evaluating a model" as finding k

Group 2

- Measure two repeated trials
- Use 10 different masses
- Linearized plot, residuals, find k

Trend motivates need for intercept

Data Sources

Data presented here was collected over the 2017-2018 academic year and includes:

- data from 25 courses across 12 institutions.
- a total of 2681 valid responses.
- matched pre- and post-instruction data from 726 students.

Scoring the PLIC

Item Classification	Criteria
Expert (E)	Picked by >50% of experts
Partial-Expert (P)	Picked by 30-50% of experts
Novice (N)	Picked by <10% of experts
Other	Other

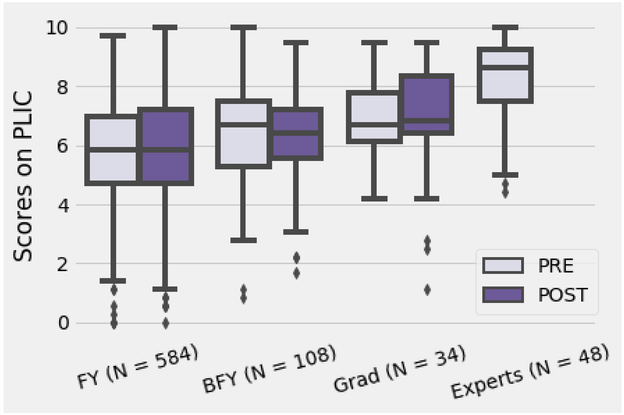
Possible Scores	Criteria
1	At least one E response, no N responses
0.75	At least one E response, at least one N response
0.5	At least one P response, no E responses, no N responses
0.25	At least one P response, at least one N response, no E responses
0	Other

Interested in using the PLIC in your class?

You can fill out the course information survey at http://cperl.lassp.cornell.edu/PLIC and you will be automatically sent a link to the PLIC for your course.

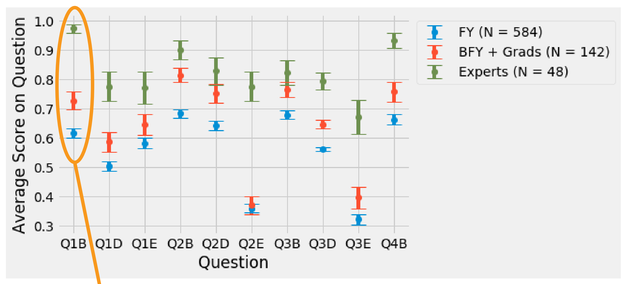
Validating the PLIC

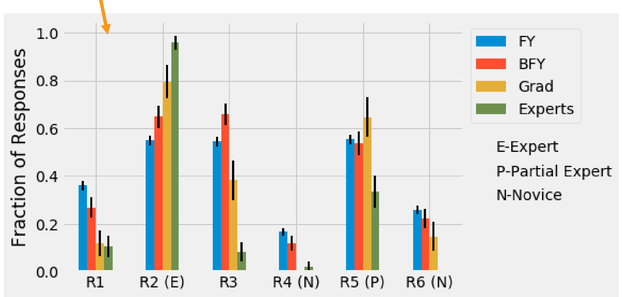
Performance is higher with greater physics maturity



- Data split by type of lab respondent participated in while taking the PLIC:
  - FY—students in first-year labs
  - BFY — students in beyond-first-year labs
  - Grad — students in graduate level labs
- Pre- and post-instruction means are not statistically different between any two groups of students.
  - Effect sizes are small or very small (Cohen's d ≤ 0.1).
- Pre-instruction means are statistically different between all groups (unpaired t-test, $p \ll 0.01$) except students in BFY and graduate level labs ($p = 0.12$).
  - Effect sizes ranged from small (d = 0.31) between students in BFY and graduate level labs to very large (d = 1.6) between students in FY labs and experts.

Experts score higher than students in BFY labs who, in turn, score higher than students in FY labs…ON EVERY QUESTION!



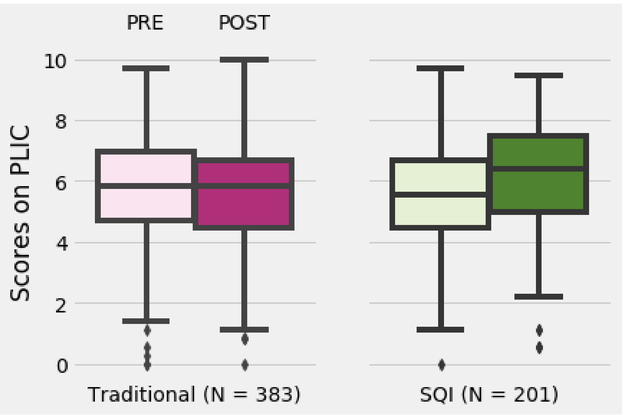


What features were most important in comparing the two k-values?

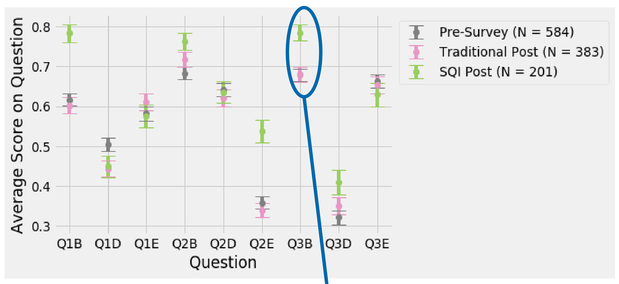
- ❑ R1 — the difference between the two k-values
- ❑ R2 — the difference between the k-values compared to the uncertainty
- ❑ R3 — the percent difference between the k-values
- ❑ R4 — the difference between the two periods
- ❑ R5 — the size of the uncertainty
- ❑ R6 — how they accounted for human error

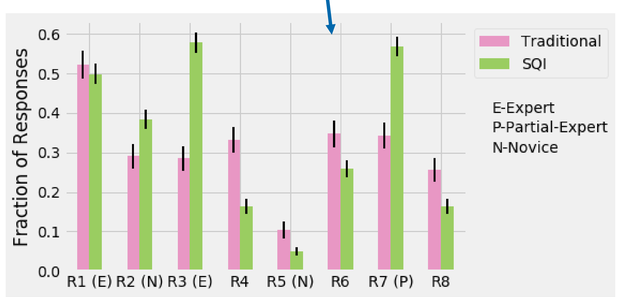
Students taught in Structured Quantitative Inquiry Labs perform better

- Structured Quantitative Inquiry (SQILabs) were designed to teach the skills that the PLIC is designed to measure.
- We expect, then, that performance on the PLIC will be greater for students taught in these re-designed labs.
- ANCOVA results indicate that, controlling for pre-instruction scores, students who particpated in SQILabs perform statistically better than students who participated in traditional labs on the post-instruction PLIC ($p \ll 0.01$).
  - Lab treatment has a medium effect (Cohen's f = 0.18) on post-scores.



Students taught in SQILabs see, on average, increases in performance on all but one question, whereas the average scores for students in traditional labs increase on only 4 out of 10 questions!





What features were most important in comparing the fit to the data?

- ❑ R1 — how close points are to the line compared to the uncertainties
- ❑ R2 — number of points above compared to number below the line
- ❑ R3 — the way points are scattered above and below the line
- ❑ R4 — how close the points are to the line
- ❑ R5 — the number of outliers
- ❑ R6 — the size of the uncertainty
- ❑ R7 — the number of points with uncertainties crossing the line
- ❑ R8—how well the data fall in a straight line