Interview validation of the Physics Lab Inventory of Critical thinking (PLIC)

K.N.Quinn[1], Carl Wieman[2], N.G. Holmes[1]

[1]Cornell University [2]Stanford University

DUE-1611482- 01

Current Assessment Format

Goal: To assess critical thinking skills of students in college-level, introductory lab courses.

Design: The PLIC is a closed-response survey. A traditional format of "choose one multiple choice" would be inadequate for our purpose. Here different follow-up questions appear based on student response. The design and format were modeled after Wilcox and Pollock's (2016 Phys. Rev. PER.) multiple-choice assessment.

Sample Question

Q1c. How well do you think Group 1's method tested the model for the period of the spring bounces? Use a scale where 1 means the method was very bad and 5 means the method was very good.

Method very bad — Method very good

1 2 3 4 5

How well Group 1's method investigated the model

Q1d. Which items below reflect your reasoning in Q1c? Please select at most 3 answers.

- ☐ The amount of data collected
- ☐ The way they collected data (or the type of data collected)
- ☐ The way they presented their methods and data
- ☐ The quality of the data they got
- ☐ Other

Q1d1. Which items below reflect your reasoning in Q1c, regarding The amount of data collected?

Too Few	Sufficient Number	Too Many
☐ They did not use enough bounces of the spring per trial	☐ They used a good number of bounces of the spring per trial	☐ They used too many bounces of the spring per trial
☐ They took too few repeated trials for each mass	☐ They took a good number of repeated trials for each mass	☐ They took too many repeated trials for each mass
☐ They did not test enough different masses (too few)	☐ They tested a good number of different masses	☐ They tested too many different masses
☐ Other (Please describe)		

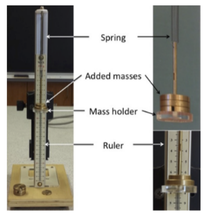
Q1d2. Which items below reflect your reasoning in Q1c, regarding The way they collected data (or the type of data collected) ?

☐ They evaluated possible other variables	☐ They didn't evaluate possible other variables
☐ They did account for human error or reaction time	☐ They didn't account for human error or reaction time
☐ Using stopwatches is a good technique	☐ Using stopwatches is a bad technique
☐ Other (Please describe)	

Context for PLIC

Given a set of standard lab equipment, test the model of harmonic motion from Hooke's law for a mass on a spring:

$$T = 2\pi\sqrt{\frac{m}{k}}$$



Group 1

Total mass: 30 g

Time for 5 oscillations (s)	Period (s)
2.18	0.554
2.68	0.536
2.76	0.552
2.70	0.540
2.67	0.534
2.82	0.564
2.66	0.532
2.72	0.544
2.56	0.512
2.57	0.514

T(average) = 0.5384 ± 0.00531 s

⇒ k = $\frac{4\pi^2 m}{T^2}$ = 4.086 ± 0.0807 N/m

Total mass: 50 g

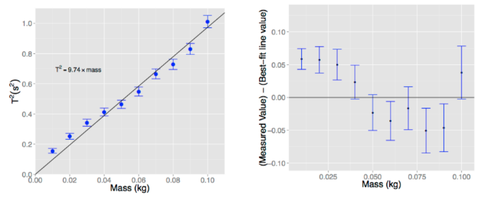
Time for 5 oscillations (s)	Period (s)
3.39	0.678
3.35	0.670
3.40	0.680
3.42	0.684
3.40	0.680
3.45	0.690
3.31	0.662
3.55	0.710
3.60	0.720
3.46	0.692

T(average) = 0.6866 ± 0.00553 s

⇒ k = $\frac{4\pi^2 m}{T^2}$ = 4.181 ± 0.0675 N/m

Calculate spring constant of 2 masses.

Group 2



Graph square of period for 10 different masses.

Two fictional groups have different methods to test the simple model of a mass on a spring. Students are asked to assess the methods, the fictional data, suggest next steps and compare the two groups.

Thinking Behaviors

Through think-aloud validation interviews, we identified three main thinking behaviors exhibited by students as they take the assessment, results which are useful not only to the PLIC but for all similar assessments.

Cuing to Key Words

- N = 5 students
- Select options with keywords learned in class, such as:
  - Percent error
  - "Human error"
- Common when answering questions related to Group 1 methodology.

To switch students from cuing to discerning, familiar material needs to be presented in an unfamiliar way (such as data in Group 2 rather than Group 1)

Discerning

- N = 10 students
- Considering all options presented, and selecting only a few after careful evaluation.
- Different students prioritize different answers, such as those related to:
  - "Human error"
  - Data collection
  - Model breakdown
- The answers selected reflect the sophistication (novice-like or expert-like) of their critical thinking.
- Some students, when exhibiting this behavior, change their initial judgment or assessment after seeing all other options presented.

Only when students are discerning in their choices can we assess their critical thinking.

Selecting all Options

- N = 3 students
- Selecting all options that appeal, regardless of priority or importance.
- All closed response option for the PLIC are "correct" in some way, and so this behavior is not associated with critical thinking and may be a symptom of the assessment design.

To switch students from selecting all options to discerning, students must be told to select a limited number of answers (no more than three)

Major			Academic Level				Gender		Race/Ethnicity			
Physics	Non-Declared STEM	Other Declared STEM	Freshman	Sophomore	Junior	Grad	F	M	African-American	Asian/ Asian-American	White	Other
N = 8	N = 2	N = 2	N = 6	N = 3	N = 1	N = 2	N = 6	N = 6	N = 2	N = 3	N = 5	N = 2

Demographic breakdown of all 12 students interviewed for this study. Students switch behavior during interview, making the total of all behavior groups greater than 12.

Developing the Assessment

The development framework for this assessment followed guidelines from Adams and Wieman (2011 Int. Jour. of Sc. Ed.) with validation process from Wilcox and Lewandowski (2016 Phys. Rev. Phys. Educ. Res.), and protocol from Madsen, McKagan, and Sayre (2017 Am. Jour. of Phys.). Arrows indicate the iterative nature of this process, where we may return to previous steps depending on findings.

Phase 1: Conception

- Delineation of the purpose of test and chose topic.
- Collect and create data.
- Initial data created by a physicist conducting the experiment.
- Initial questions based on their self-questions related to their choice of methods, data and collection.

Phase 2: Open-Response

- Development of open-response version.
- Student interviews conducted.
- Student written responses gathered from introductory class.
- Experts consulted.
- Results previously presented.[5]

Phase 3: Closed-Response

- Development of the closed-response version based on open responses.
- Currently using the online survey software *Qualtrics*.
- Field testing though think-aloud validation interviews.
- Develop scoring.

We are currently in Phase 3.

Phase 4: Operational Use

- Assemble and evaluate the test for operational use.
- Large-scale validation, reliability and validity testing.
- Easy access for instructors, with automation for setup, reminders and report-back scoring similar to Wilcox *et. al.* (2016 Phys. Rev. PER)