# Physics Evidence-Centered Assessment Project (PE-CAP) Vision

Rachel Henderson, Andrew Heckler, Dena Izadi, and John Stewart

June 2024

## 1 Project Goals

The Physics Evidence-Centered Assessment Project (PE-CAP) will develop a set of 6 to 12 item subscales which measure different domains of qualitative kinematic and dynamics. One scale will be a broad measure of kinematic and dynamics which can be used as a replacement for the Force Concept Inventory (FCI) or the Force and Motion Conceptual Evaluation (FMCE). Other scales will measure more narrow domains of kinematics and dynamics including 1D kinematics, 2D kinematics with projectile motion, Newton's 1st and 2nd law, and Newton's 3rd law with additional scales determined by community input.

All items in these scales will undergo extensive systematic validation including expert review, qualitative think aloud sessions, open response testing, and large scale quantitative testing. The validation process and the results of each step will be made as publicly available as possible. Data taken during the validation process will be made available to other researchers to replicate results and to apply different analyses.

## 2 Motivation

Commonly used legacy instruments, such as the Force Concept Inventory (FCI) and the Force and Motion Conceptual Evaluation (FMCE), have been shown to have serious flaws including substantial psychometric problems which threaten their reliability and validity; these include issues of demographic fairness. For additional discussion, see the motivation argument available at the website.

## 3 Guiding Principles

**Transparency** The development and validation process should be thoroughly documented and publicly available.

**Community Property** The new instruments should be informed by broad community input rather than being the creation of a few researchers, representing their personal preferences.

**Continued Evolution** The new instruments should continue to evolve after initial publication as new research results, delivery methods, and communities needs arise.

**Model Driven** The new instruments should be developed from a thorough model indicating what is being measured (and equally importantly) what is not being measured.

**Evidence Based** The new items and instruments should be supported by convincing evidence that they measure the published measurement model. This evidence should be publicly available.

**Flexibility** Instructors and researchers need valid instruments that can be adapted to their needs. An instrument that returns a single score and requires a full class period to apply is not an optimal solution.

**Fairness** The new instruments should be validated for different populations and demonstrated to be fair.

# 4 Arguments

This project is organized around the principles of Evidence-Centered Design which ultimately involves the construction of an "Argument" that the items, scales, and instruments constructed by this project do what they are intended to do. This argument shows the process of using the results of the assessment data collected to argue that a student has a certain level of mastery of the domain measured. This argument is modeled around a legal argument and should be convincing to both instructors and researchers. This project will produce two arguments: an assessment argument and a validity argument.

## 4.1 Assessment Argument

Modeled after a legal argument, the assessment argument presents the instruments used, artifacts collected, the individual constructs measured, and how this evidence is analyzed to convince an audience of stakeholders, both researcher and physics educators, that a student has achieved some level of mastery with the material in the domain of interest.

## 4.2 Validity Argument

For the assessment argument to be convincing, the stakeholder must be convinced that the measurement methods used accurate measure the constructs of interest, the stakeholders must be convinced the methods are valid. We call

the subset of the assessment argument that discusses the validity of the measurement methods, the Validation Argument. While the Assessment Argument is specific to the domain measured, we expect the Validation Argument to be applicable across domains.

# 5 Products

This project will produce a number of products beyond those covered in the general project goals.

**Item Pool** The project will produce a collection of upward of 200 validated conceptual kinematic and dynamic items each with strong validity evidence and each characterize over a set of models.

**Conceptual Inventories** The project will produce a set of FCI length conceptual inventories drawn from the item pool measuring different sub-domains of kinematics and dynamics. Beyond strong item level properties, each will have strong test level reliability and validity. Each will also provide not only an overall measure, but also more granular measure of sub-domains.

**Conceptual Subscales** Smaller collections (6 to 12 items) of items drawn from the conceptual inventories and statistically similar will be constructed. As much as possible, these will be constructed to factor properly when used in a combined instrument.

**FCI/FMCE Replacement** An approximately 12 item instrument will be selected from the item pool and independent measurements of the FCI and FMCE (not sure this is one or two instruments) so that the scale is cross-normed with the FCI/FMCE allowing the continued use of legacy data.

# 6 Questions

As we work on the project, we continue to general questions. Some can be answered by additional research; some can only be answered by a consensus of the physics community.

- Can algebra and calculus-based classes be measured accurately with the same instrument? (Answer looks like no).
- Can one instrument measure classes with all levels of prior preparation (Answer almost certainly no, but what are the parameters)? If the answer is no, how many instruments do we need.
- Can one instrument replace the FMCE and the FCI and maintain strong instrument linking? (Answer, almost certainly no).
- How important is it to present items in physically realistic contexts? How does this affect item fairness?

- What are optimal item writing practices for conceptual physics?
- How do we maintain accessibility for all students? Can this be done with a single instrument?
- What is the optimal mix of representations for an instrument?
- How do we allow instructors of flexibly and accurately use the item pool?