Measurement and Characteristics of Evidence-Based Instruction in Physics

David E. Meltzer

Arizona State University

NSF DUE #1256333, WIDER: EAGER: Recognizing, Assessing, and Enhancing Evidence-Based Instructional Practices in STEM at Arizona State University, Polytechnic; PI: D. E. Meltzer; Co-PIs: Y. Kang, M. Zandieh

Common Characteristics of Research-Based Active-Learning Instruction in Physics

References: D. E. Meltzer and R. K. Thornton, American Journal of Physics **80**, 478 (2012); J. P. Mestre, Physics Education **36**, 44 (2001).

A. Instruction is informed and explicitly guided by research regarding students' pre-instruction knowledge state and learning trajectory, including:
   1. Specific learning difficulties related to particular physics concepts
   2. Specific ideas and knowledge elements that are potentially productive and useful
   3. Students' beliefs about what they need to do in order to learn
   4. Specific learning behaviors
   5. General reasoning processes

B. Specific student ideas are elicited and addressed

C. Students are encouraged to "figure things out for themselves"

D. Students engage in a variety of problem-solving activities during class time

E. Students express their reasoning explicitly

F. Students often work together in small groups

G. Students receive rapid feedback in the course of their investigative or problem-solving activity

H. Qualitative reasoning and conceptual thinking are emphasized

I. Problems are posed in a wide variety of contexts and representations

J. Instruction frequently incorporates the use of actual physical systems in problem solving

K. Instruction recognizes the need to reflect on one's own problem-solving practice

L. Instruction emphasizes linking of concepts into well-organized hierarchical structures

M. Instruction integrates both appropriate content (based on knowledge of students' thinking) and appropriate behaviors (requiring active student engagement)

Related Notes:

- There exists no clear quantitative measure of how, and in what proportion, the various characteristics of effective instruction need be present in order to make instruction actually effective.
  - ➢ For example, would a score of "4 out of 4" on characteristics *E*, *F*, *G*, and *H* on the above list outweigh a score of (e.g.) "3 out of 4" on characteristics *A*, *B*, *C*, and *D?*
- Instructional developers gather and analyze evidence on *specific instructional implementations of specific curricula*, but not (in general) on the specific characteristics enumerated above
- Firm evidence of effective instructional practice always occurs in the context of a large set of tightly interlinked characteristics, each characteristic (apparently) closely dependent on the others for overall instructional success.

How are Research-Based Physics Instructional Methods Assessed?

Reference: Meltzer and Thornton (2012): Compendium of ≈30 research-validated instructional methods/materials in physics

Each method/material was examined to determine which instruments and techniques were used to provide evidence of instructional effectiveness

Types of Diagnostic Instruments

"Standardized" surveys: ≈20-40 items, usually multiple-choice, qualitative (non-algebraic, non-numerical); Example: Force Concept Inventory (FCI)

Researcher-constructed free-response questions: qualitative emphasis; fewer than 8 items; may or may not require student explanations; Examples: University of Washington assessment items; University of Minnesota "context-rich" problems

Instructor-constructed course assessments: quizzes, exams, homework, grades; emphasis on quantitative and algebraic problem-solving

Comparison Groups

Local: compare to similar courses at home institution that use standard instruction

External: compare to similar courses/student populations at other institutions

National baseline: compare to previously published data reflecting performance in representative equivalent courses at multiple institutions

Diagnostic Survey Instruments

Frequently used:
- Force Concept Inventory (FCI)
- Force and Motion Conceptual Evaluation (FMCE)

Occasionally used:
- Conceptual Survey in Electricity and Magnetism (CSEM)
- Brief Electricity and Magnetism Assessment (BEMA)

Rarely used:
- Electric Circuits Concept Evaluation (ECCE)
- Mechanics Baseline Test (MBT)
- Colorado Upper-Division Electrostatics Diagnostic Quiz [non-MC] (CUE)
- Quantum Mechanics Visualization Instrument (QMVI)
- Quantum Mechanics Assessment Tool [non-MC] (QMAT)
- Quantum Mechanics Conceptual Survey (QMCS)

Other Outcomes Assessed

Attitudes and beliefs (e.g., student ideas about how best to learn physics); *Examples:* Redish , Saul, and Steinberg (1998); Adams et al. (2006)

Facility with physics practices (e.g., ability to design and execute experiments); *Example:* Etkina and Van Heuvelen (2007)

Use of Pre-Instruction Tests to Predict Student Course Performance

Halloun and Hestenes (1985) administered the "Mechanics Diagnostic Test" (early version of FCI) and a mathematics diagnostic test in general physics courses at Arizona State University

- Course performance determined by scores on class exams
- Examined both algebra- and calculus-based courses that used traditional instruction

Findings:

1. Scores on physics concept pretest and on math skills pretest were highly correlated with course performance, but nearly uncorrelated with each other (i.e., they were *independent* factors)

2. Students with *combined* physics + math pretest scores < 43% had only 5% probability of earning course grade over C

Mean physics pretest scores for students earning various letter grades (calculus-based physics, N = 192, 16% earn A's):

A: 63%

B: 55%

C: 47%

D/F: 42%

Comparison #1: Large Western State University

Mean FCI pretest scores for students earning various letter grades (calculus-based physics, N = 412, 24% earn A's):

A: 70%

B: 54%

C: 45%

D/F: 37%

Comparison #2: Arizona State University (2012-2013) [this work; evidence-based instruction]

Mean FCI pretest scores for students earning various letter grades (calculus-based physics, N = 107, 32% earn A's):

A: 57%

B: 41%

C: 41%

D/F: 32%

Correlations Between Course Grade and Diagnostic Pretest Score

Math skills: +0.51 (ASU, 1985; calculus-based, N = 478)

Lawson scientific reasoning test: +0.50 (Small university, algebra-based, N = 238)

Physics concepts:

+0.55 (ASU, 1985, calculus-based, MDT, N = 478)

+0.34 (U. Minnesota, 1997-99, calculus-based, FCI, N = 1645)

+0.48 (Small university, 2006-13, algebra-based, FCI, N = 238)

What Grade is Predicted by FCI Pre-Test Score?

Study #1: Henderson (2002), U. of Minnesota

(N_{total} = 2020; 21% earn A's; pre ≈ 45%, post ≈ 68%; <g> ≈ 0.42)

Pretest Score:

Low (0-30%) [N = 663]	**A:** 10%	**B:** 30%	**C:** 46%
High (63-100%) [N = 349]	**A:** 47%	**B:** 40%	**C:** 9%

Study #2: Arizona State University (2012-13) [this work]

(N_{total} = 107; 32% earn A's; pre ≈ 45%, post ≈ 71%; <g> ≈ 0.47)

Pretest Score:

Low (0-30%) [N = 34]	**A:** 12%	**B:** 44%	**C:** 26%
High (63-100%) [N = 23]	**A:** 65%	**B:** 22%	**C:** 13%

Summary

•There is a strong correlation between final course grade and scores on various pre-instruction diagnostic tests

•A strong correlation clearly persists in the presence of evidence-based instruction

•Pre-instruction tests may be able to give "early warnings" in cases where special intervention might be helpful

•Key question remains: What are primary factors underlying "most successful" student cases of low pre-score/high grade?

Issues of Concern

Most assessments done via multiple-choice survey instruments
- (relatively) easy to implement
- limited insight into student thinking: imprecise, and lack explanations
- limited coverage of instructional intervention (narrow scope of topics)

Most non-survey assessments are unpublished

Most components of each collection of materials go unassessed
- *Exception:* University of Washington; majority of UW Tutorials undergo extensive cycle of iterative assessment and validation