

Evaluation of Teachers—or of Teaching?—for Improving Learning Outcomes

David E. Meltzer

Mary Lou Fulton Teachers College
Arizona State University

Supported in part by PhysTEC through NSF PHYS #0108787

Overview

- For over 40 years, physicists have investigated ways to improve the quality of physics teaching
- Main Findings:
 - a) Instructional methods make a big difference
 - b) Curricular materials make a big difference
 - c) Teacher experience with methods and materials makes a significant difference
 - d) Individual differences among teachers—apart from (a), (b), and (c)—make a relatively small difference
- *Conclusion:* To best improve physics teaching, assess and strengthen teachers' use of and experience with research-validated methods and materials

“The Question”

➤ ...How can teaching “quality” of different teachers in different schools or classrooms be compared?

...or:

➤ What makes a high-quality teacher, and how can you recognize one?

High-Quality Teachers

Q: What is a high-quality teacher?

A: One who is effective at achieving learning goals.

But...

...physics teachers have many goals:

- Non-subject-specific learning goals
- Science process skills
- Science attitudes (“nature of science”)
- Physics content knowledge

Goals for All Teachers

Non-subject-specific learning goals

- *Functioning*: sustained, persistent on-task effort
- *Reasoning*: observation, recognition of trends and patterns
- *Attitudes*: interest, enthusiasm, enjoyment, and perseverance
- *Process*: productive questioning, thoughtful action, reflective practice
- Since these are goals for all teachers, a student's manifestation of these qualities can't be traced to any individual teacher

Science Learning Goals

- Science process skills
 - Develop hypotheses, design and execute controlled experiments, reason from evidence, test conclusions; develop proportional, correlational, and probabilistic reasoning
- Science attitudes
 - Understand the nature of science; have interest, enjoyment, and perseverance in science learning

➤ *Since these are goals for all science teachers, their presence can't be traced to the physics teacher*

(except perhaps with systematic pre-/post-testing in every sequentially taught science class...)

Physics Learning Goals

- Physics process skills
 - Develop hypotheses, design and execute controlled experiments, reason from evidence, test conclusions, *in physical systems*
- Physics attitudes
 - Understanding the nature of physics; have interest, enjoyment, and perseverance in physics learning
- Physics content knowledge; problem-solving skill
 - Understanding of physics concepts and unifying principles, ability to apply ideas to solve problems real contexts
- Use and Application of Physics
 - Appreciate and use physics knowledge in everyday life

What is Teacher Quality?

And how can it be measured?

- The quality (or “effectiveness”) of a teacher is the degree to which learning goals for their students are achieved.
- The quality of a physics teacher incorporates general learning goals, science learning goals, and physics learning goals.
- It is virtually impossible to directly assess a physics teacher’s effectiveness in achieving general learning goals—and perhaps science learning goals—so long as students have multiple teachers.
 - *But these goals are nonetheless key characteristics of quality physics teaching, as any other teaching.*

What is Teacher Quality?

And how can it be measured?

- The ultimate measure of teacher/teaching quality is the impact on the teacher's students
- Assessment of teaching quality by directly assessing students' learning poses monumental logistical challenges
- Indirect or "proxy" measures of teaching quality consist of assessments of the teachers themselves

Assessment Measures

- Indirect measures (“Proxies”)
 - observations of teacher’s classroom functioning and pedagogical style
 - assessments of teachers’ knowledge, skills, and attitudes
- Direct Measures
 - assessment of learning gains made by teacher’s students

Indirect Measures

Observations of teacher's classroom functioning

- Classroom management
 - are students on-task and engaged?
- Planning and implementation
 - Are instructional goals and plan specified and evident?
 - Is there evidence for effectiveness of materials used?
- Instructor engagement with students
 - Do students use inductive and deductive reasoning strategies? Are students' ideas elicited and addressed?
 - Does instructor use inquiry-based questioning strategies?

Assess with rubrics (e.g. RTOP, UTOP) and qualitative observations by experienced instructors.

CAUTION: Rubrics' link to objective learning measures is not well established

Indirect Measures

Teachers' Knowledge and Skills

- Knowledge of physics concepts
- Knowledge of science process skills
- Knowledge of “Nature of Physics” (practices and philosophies of physics community),
- Pedagogical content knowledge (knowledge of and interest in issues related to teaching and learning of specific concepts)
 - Includes ability to implement effective methods and guide student inquiry

Indirect Measures

Teachers' Knowledge and Skills

- Knowledge of physics concepts
- Knowledge of science process skills
- Knowledge of “Nature of Physics” (practices and philosophies of physics community),
- Pedagogical content knowledge (knowledge of and interest in issues related to teaching and learning of specific concepts)
 - Includes ability to implement effective methods and guide student inquiry

Assessment of Pedagogical Content Knowledge

“*Pedagogical Content Knowledge*” (PCK):

Awareness of, interest in, and detailed knowledge of learning difficulties and instructional strategies related to teaching *specific* science concepts, including appropriate assessment tools and curricular materials. (Shulman, 1986)

Assessment of Pedagogical Content Knowledge

- No currently accepted, standard physics-PCK instruments exist
- Those under development incorporate:
 - analysis of preservice teachers' interpretations of problem responses or of discussions offered by hypothetical students (e.g., Thompson, Christensen, and Wittmann, 2011);
 - analysis of preservice teachers' evolving ability to create and apply lesson and unit plans, assessments, questioning strategies, etc. (e.g., Etkina, 2010).

Direct Measures

Assessment of Learning Gains Made by Teacher's Students

- Students' physics attitudes
 - Attitude surveys, e.g., MPEX, CLASS, VASS, EBAPS
- Students' science process abilities in physics
 - Qualitative observations and rubrics to assess skills with physics experiment design, execution, and analysis, e.g., Etkina et al. (2006)
- Students' physics conceptual knowledge
 - Multiple-choice and free-response written diagnostics, interviews; e.g., FCI, FMCE, CSEM, BEMA, UW questions.
 - No standardized instruments for problem-solving ability

Assessment of Physics Concept Knowledge

- *Zeroth approximation*: scores on state tests such as New York State Physics Regents exam.
 - Questionable validity
- *First approximation*: pre- and post-tests using standardized diagnostics such as FCI, FMCE, CSEM, and BEMA.
 - Confounding factors
- *Second approximation*: University of Washington tutorial-style questions; “Think-aloud” problem-solving interviews with students.
 - Logistically forbidding

State Tests Have Questionable Reliability and Validity

- They are not explicitly based on peer-reviewed research by physics education experts
- Questions are often ambiguous, misleading, superficial, or incorrect
 - *Cf.* Kanim and Loverude, “Content exams for pre-service physics teachers: A failing grade,” *Phys. Teach.* **44**, 101(2006).
- Emphasis is often on factual knowledge recall, simple algorithmic calculations, or vocabulary
 - *Cf.* R. Steinberg, *An Inquiry into Science Education* (Sense Publishers, 2011).

The University of the State of New York
REGENTS HIGH SCHOOL EXAMINATION

PHYSICAL SETTING
PHYSICS

Wednesday, June 15, 2011 — 1:15 to 4:15 p.m., only

The answers to *all* questions in this examination are to be written in your separate answer booklet. Be sure to fill in the heading on the front of your answer booklet.

You are to answer all questions in all parts of this examination according to the directions provided in the examination booklet. All work should be written in pen, except for graphs and drawings, which should be done in pencil. You may use scrap paper to work out the answers to the questions, but be sure to record all your answers in the answer booklet.

When you have completed the examination, you must sign the statement printed on the first page of your answer booklet, indicating that you had no unlawful knowledge of the questions or answers prior to the examination and that you have neither given nor received assistance in answering any of the questions during the examination. Your answer booklet cannot be accepted if you fail to sign this declaration.

Notice. . .

A scientific or graphing calculator, a centimeter ruler, a protractor, and a copy of the *2006 Edition Reference Tables for Physical Setting/Physics*, which you may need to answer some questions in this examination, must be available for your use while taking this examination.

The use of any communications device is strictly prohibited when taking this examination. If you use any communications device, no matter how briefly, your examination will be invalidated and no score will be calculated for you.

DO NOT OPEN THIS EXAMINATION BOOKLET UNTIL THE SIGNAL IS GIVEN.

32. A 256-hertz vibrating tuning fork is brought near a nonvibrating 256-hertz tuning fork. The second tuning fork begins to vibrate. Which phenomenon causes the nonvibrating tuning fork to begin to vibrate?

- (1) resistance
- (2) resonance
- (3) refraction
- (4) reflection

33. On the atomic level, energy and matter exhibit the characteristics of

- (1) particles, only
- (2) waves, only
- (3) neither particles nor waves
- (4) both particles and waves

35. Which particles are not affected by the strong force?

- (1) hadrons
- (2) neutrons
- (3) protons
- (4) electrons

29. What is the wavelength of a 256-hertz sound wave in air at STP?

- (1) 1.17×10^6 m
- (2) 1.29 m
- (3) 0.773 m
- (4) 8.53×10^{-7} m

49. A deuterium nucleus consists of one proton and one neutron. The quark composition of a deuterium nucleus is

- (1) 2 up quarks and 2 down quarks
- (2) 2 up quarks and 4 down quarks
- (3) 3 up quarks and 3 down quarks
- (4) 4 up quarks and 2 down quarks

Even Research-Based Diagnostics Have Confounding Factors

- Post-test scores on FCI strongly correlated with pretest scores (“pre-instruction knowledge”): Hake, 1998.
- Normalized learning gains on CSE correlated with algebra skills: Meltzer, 2002.
- Normalized learning gains on FCI correlated with pretest scores, and with scores on Lawson test of scientific reasoning: Coletta and Phillips, 2005; Coletta, Phillips, and Steinert, 2007
- Anecdotal evidence (conversations with many high school physics teachers): Student reading levels can have significant impact on FCI scores.

Effects can be HIGHLY
Non-Linear:

From Thornton, “Uncommon Knowledge” (2004).

No students who were below 13% on pretest before instruction were able to reach 90% on post-test, while ALL students above 55% on pretest DID reach 90%

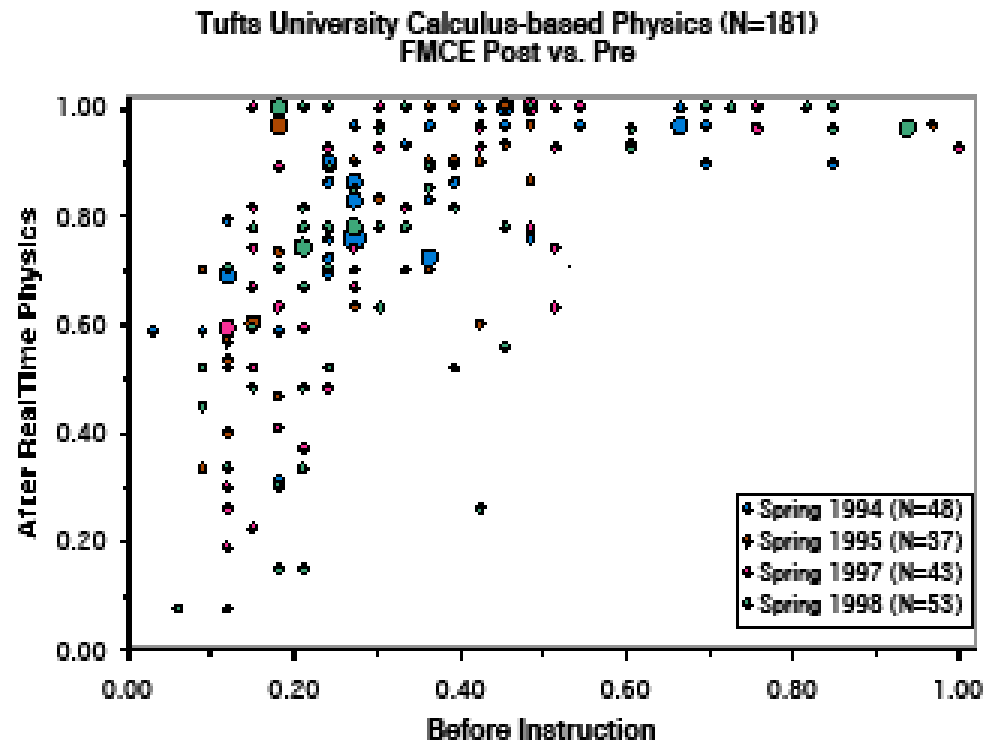


Fig. 1: Comparison of pre and post performance scores on the Force and Motion Conceptual Evaluation (FMCE) for 181 students in the course described in Section 3. Students who started low and finished low are in the lower left hand corner. Students who finish high are along the top edge of the figure.

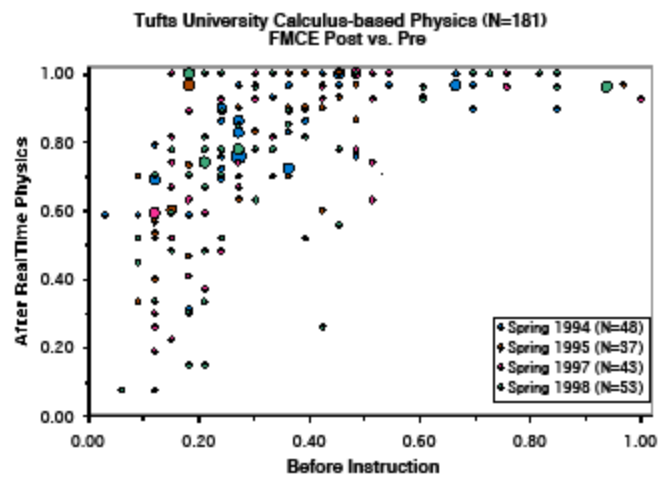


Fig. 1: Comparison of pre and post performance scores on the Force and Motion Conceptual Evaluation (FMCE) for 181 students in the course described in Section 3. Students who started low and finished low are in the lower left hand corner. Students who finish high are along the top edge of the figure.

Confounding Factors Prevent Extraction of Individual Instructor Effects

- Force Concept Inventory (FCI) is not well tuned to high school population (pretest responses are too close to random guessing) preventing reliable assessment of pre-instruction knowledge
- *Enormous* differences exist between classes and schools in math skills, reading ability, motivation, previous preparation, and learning resources
- Same instructor using same instructional materials and methods can get very high or very low FCI post-test scores, depending on the student make-up of the class

Assessment Challenges

- No currently available reliable and valid instruments for assessing most physics learning goals
- Reliable and valid instruments exist for assessing some areas of physics content knowledge, but increased validity (requiring explanations of reasoning) imposes increased logistical challenges
- Multiple confounding factors drastically reduce validity of comparisons of non-equivalent groups

What We *Do* Know

- Use of research-based active-learning instructional methods and curricular materials (“Interactive Engagement”) significantly improves student learning in physics.
- The improvements are tightly correlated with use of and practice with the materials and methods, but differences in outcomes with different instructors are relatively small.

Research-Based Active Learning Instruction

(Meltzer and Thornton, AJP, in press)

(“Interactive Engagement” [Hake, 1998])

- Instruction is informed and explicitly guided by students’ pre-instruction knowledge state and learning trajectory, including:
 - *Specific learning difficulties related to particular physics concepts*
 - *Specific ideas and knowledge elements that are potentially productive and useful*
 - *Students’ beliefs about what they need to do in order to learn*
 - *Specific learning behaviors*
 - *General reasoning processes*

1. Specific student ideas are elicited and addressed.
2. Students are encouraged to “figure things out for themselves.”
3. Students engage in a variety of problem-solving activities during class time.
4. Students express their reasoning explicitly.
5. Students often work together in small groups.

6. Students receive rapid feedback in the course of their investigative or problem-solving activity.
7. Qualitative reasoning and conceptual thinking is emphasized.
8. Problems are posed in a wide variety of contexts and representations.
9. Instruction frequently incorporates use of actual physical systems in problem solving.

10. Instruction recognizes the need to reflect on one's own problem-solving practice.
11. Instruction emphasizes linking of concepts into well-organized hierarchical structures.
12. Instruction integrates both appropriate content (based on knowledge of students' thinking) and appropriate behaviors (requiring active student engagement).

But...How can we Assess Teaching “Quality”?

- Determine learning goals
- Obtain many large student samples that are highly consistent and reproducible
- Control all factors while varying either instructional method or individual instructor
- Administer validated probes of student learning after instruction is complete
- Compare outcomes in order to ascertain methods or instructors most effective in realizing learning goals
- This has been done, thousands of times...in colleges and universities...to determine local effectiveness
- Most widely assessed goal is conceptual understanding

Findings for Well-Controlled Samples

- “Traditional Instruction”: 10-15% absolute score improvement, little instructor dependence
- “Active-Learning” [Interactive-Engagement] Instruction: 15-35% absolute score improvement
 - “Practice Effect”: ~ 10%
 - “Instructor Effect” ~ 5%

Note: numbers are approximate, based on my analysis of U. Minnesota and U. Colorado data bases

Summary: Outlook and Implications

- No near-term prospects for valid and reliable assessment of high school physics teachers *in situ*
- Overwhelming evidence shows that use of research-based methods and materials improves physics learning
- Most direct route to improved high school physics instruction is adaptation and use of research-based materials and methods, with continued mentoring and practice in use of these materials
- Assessment of use of physics teaching methods and materials is the key to *physics-specific* teaching quality