Formative Assessment: Effectiveness of ConcepTests and Conceptual Probes in Math and Science

Dr. George Collison &
Dr. Judith Collison
Cairo 03/28/09

Invited paper presented at the Cairo Grand Hyatt, April 28, 2009

First Annual Conference of Educational Research Center on Educational Measurement.

Authentic Assessment for Meaningful Learning and Informed Educational Policy
Abstract:

Formative Assessment: Effectiveness of ConcepTests and Conceptual Probes in Math and Science

Dr. George Collison &
Dr. Judith Collison
Cairo 03/28/09

In the mid 90’s Dr. Eric Mazur of Harvard invented devices called ConcepTests. ConcepTests are carefully crafted questions designed to engage directly students’ minds with content. Shocked by the ineffectiveness of his lecture technique, as measured by student responses to the Halloun-Hestenes Force Concept Inventory (FCI), Mazur radically changed his pedagogical technique. Mazur defines the lecture method as the ideal mechanism for knowledge to pass from the professor’s notes to the students’ notebooks without engaging the minds of either. There are no longer any lectures in introductory Physics classes at Harvard, MIT, Princeton, Stanford and other major American universities. In the volume Peer Instruction Mazur explains his technique. Peer instruction rests on a think-pair-share model that actively students’ with the material and with discussion with peers. Application of peer instruction techniques has achieved levels of understanding 50% higher than that of traditional instructional methods, as measured by the FCI. This talk engages participants in peer instruction using novel ConcepTests and invites audience discussion about the technique.

Parallel with Dr. Mazur’s work at Harvard the scientists and designers at the Concord Consortium in Concord, Massachusetts developed a complementary mechanism for formative assessment in web courses called conceptual probes. Conceptual probes, in some ways, resemble in construction the ConcepTests, but the administration, and pedagogical use and impact on participants differ significantly. The first conceptual probes were parts of online professional development for international groups of educators in mathematics and science. Designers or instructors administer conceptual probes via a CGI database. Participants invited students in their classes to vote anonymously with their votes and comments recorded for later viewing by a very wide audience. The data engine records thousands of responses, viewable in select groups or as a whole. Instructors, curriculum designers, or individual instructors and students can view and use these materials to inform design, assess effectiveness, or include as part of instruction. Participants engage in response to some conceptual probes and dialogue about potential use for their constituencies.
We will make concrete a set of effective links between high quality assessment and effective pedagogy in physics and mathematics education.

Too often in educational practice assessment trails instruction. This practice leads to production of data much too late in the process to serve useful purpose for the student.

We offer an alternative, research validated model that integrates formative assessment into educational practice.

The core of the presentation highlights a strongly validated method of formative assessment called “Peer Instruction” (PI). PI was developed by Dr. Eric Mazur at Harvard in the 1980’s.

In parallel with Dr. Mazur, researchers at the Concord Consortium employed formative assessment techniques in web-based courses. These assessments are called “Conceptual Probes.”

We present an interactive session using both techniques.
What is the Concord Consortium?

CC is a non-profit educational think tank located in Concord Massachusetts. The small group of scientists and educators at CC is dedicated to “realizing the educational potential of technology.”

Though small in size, CC has had a powerful impact on educational research and practice. CC is not an educational institution like a college or university, nor is it affiliated with any particular university of college. CC is an educational “think tank” whose charge is to think out of the box and bring educational innovations using technology to classrooms world-wide. CC is funded through grants from the Department of Education or the NSF.

I, Dr. George Collison was one of the founding members of the consortium. My interests lie in mathematics and physics education as well as in methods of initiating and sustaining dialog.

I, Dr. Judith Collison, also worked with CC since its beginning. I was trained in mathematics and philosophy. I have a long standing interest in authentic methods of assessment that inform instruction. This includes performance based and formative assessment, and testing that improves thinking skills.
The work and products of CC are widely respected.

CC has been in operation for 15 years. Dr. Bob Tinker founded the Concord Consortium. Previously he had been founder and Chief Science Officer of TERC (Technical Education Resource Center), another think tank in Cambridge, Massachusetts.

Currently there are about 30 members of CC involved in a wide variety of products. developing and using the best in educational technology.
All software and products of the Concord Consortium are free open-source code.

http://www.concord.org

Please feel free to visit the site and download our products.

All are open-source code.
Some of CC’s Products

Scientists at CC invented the MBL
The Sonic Ranger

Virtual High School
Global Consortium
The Virtual High School

Global Lab

CC’s Products:
In the 1980 members of the Concord Consortium invented the Sonic Ranger, the device that connects to a computer port and graphs motion in real time. The Sonic Ranger is now marketed by Vernier and others. It is widely used in mathematics and physical science education. CC also invented a wide variety of probes modeled after the ranger’s interface.

In 1993 CC invented the Virtual High School. VHS operates on a collaborative model that offers member schools 25 seats in any of its 300+ courses online for every teacher who teaches a VHS course. Seats can also be purchased. Students at schools may have need of offerings local staff cannot provide. Popular courses include: Advanced Placement Physics and Calculus, Celestial Mechanics, Advanced Economics, Engineering Design, C++ programming, advanced web design, Advanced Placement Statistics or Computer Science.

The Global Lab provided an environmentally based curriculum for secondary students from more than a dozen countries. Schools are matched by biomes. Participating schools include American, Spanish, Italian, and Portuguese schools, and many schools in the former Soviet Union.
More of CC’s Products

**Facilitating Online Learning**
A guide for moderators

*Seeing Math™* web-video case studies

**CAPA assessments**
in electronics.

*Facilitating Online Learning* summarizes CC’s experience in online moderation. It is currently the best selling text on online moderation.

*Seeing Math™* offers 23 web-based, video case studies for middle school and secondary mathematics teachers. It offers a scaleable path to quality professional development. Annenberg - Public Broadcasting currently offers *Seeing Math* to teachers world wide. There are elementary and secondary case studies. Teachscape, Inc. markets the elementary cases.

**CAPA**, a recent project, provides online instruction and assessments for individuals seeking licenses as electricians, systems controls workers. Dr. Paul Horwitz is the director of the CAPA project.
INTEC, CC’s first online course grant developed courses in mathematics, physical, biological and chemical sciences. The dialogues of INTEC proved the source of the moderation techniques captured in Facilitation Online Learning.

Hands on Physics brought Dr. King of MIT’s vision of “hands-on” physics learning to high school students. Using kits made from very inexpensive materials students build projects, at the same time learning physical principles.

Molecular Workbench, a recent CC project, offers a unique online tool which enables users to create shareable molecular models. MW includes capacities to author modeling environments for assessment in physical, biological and chemical studies.
Curricular Change Must be Motivated

For Physics Powerful motivation came from results the Force Concept Inventory (Halloun & Hestenes 1985)

FCI asked conceptual questions about Newtonian mechanics.

We will deal for this talk with the FCI developed by Drs. Halloun and Hestenes in the 80’s. There is another test, the Mechanics Baseline Inventory. The MBI makes use of computational probes.

The FCI proved to be a watershed in physics teaching. It gave a stunningly accurate, consistent, and depressing view of the effectiveness of physics education at all levels.

FCI poses only conceptual questions; there isn’t a number in sight. It targets the narrow, but very important field of Newtonian mechanics. The FCI revealed that traditional lecture delivery systems can improve student knowledge of physics by about 22% in one year. This improvement seems independent of lecturer or text. This figure is just not acceptable.

We cannot go into this next century with only 22% of the secondary and college population cognizant of basic physical principles.

Researchers at my university, the University of Massachusetts, coined the phrase “the 8 to 12 physicist.” If one asked conceptual questions from 8am-12noon, you got one answer - what students expected you wanted. If you asked the same questions (to engineers and science majors) after these hours one got totally different non-Newtonian responses. Misconceptions are exceedingly robust.
Dr. Mazur of Harvard heard of the FCI and tried it out with his students. “These are easy questions. They’ll have little or no problem with them.” he stated.

His interactions with the FCI questions and his students are documented in his 1985 book *Peer Instruction*. He gave well prepared and interesting lectures. They were, as measured by FCI responses, quite ineffective. The quote above came from the first day in the use of FCI questions in his classroom.

When asked a conceptual question (no numbers, no use of a formula needed) students are often very confused.

They commonly take numbers from a question, seek a formula with matching variables and plug them in and hammer out an answer, understanding little of the physics.

The conceptual question forces the student to reveal his or her more basic thought patterns.
In the spirit of Mazur’s Peer Instruction we will do a sample, not just talk about it. Above is a PI question, borrowed from the FCI (H&H 1982).

Two horses in situation 1 pull on a pair of jeans. In situation 2 one horse pulls on the jeans whose other side is tied to a tree.

**Using the index cards given please vote as an individual.** An ERC worker will tally responses visually. Technology can permit clickers to tally and display responses immediately. Research has shown that index cards are just as effective as infrared clickers. Lack of technology should be no barrier to this technique.

The first vote takes about 1-2 minutes.

**Vote again as a group. Take 3 minutes to talk with others at your table. Which answer is right and why?**

The 2nd vote takes about 3-5 minutes. The instructor explains the answer. Since the jeans are not moving the forces must sum to zero. By Newton’s third law the stump must be pulling back with the same force as the horse.

We will go into greater depth on the usage of conceptual questions in teaching following suggestions of Dr. Mazur.
This slide comes from one of Dr. Mazur’s talks.

First let’s try to understand the axes and the grey area.

The horizontal axis gives a normalized score 1-10 on a set of traditional quantitative physics questions. The vertical axis gives a normalized score on a set of 10 conceptual questions. The area of the circle represents the student population at each at each specific achievement.

The course was taught at Harvard by Dr. Mazur using traditional lecture format. If the traditional format taught computational and problem solving skills equally as it did conceptual skills one would expect all data to lie in the gray diagonal area.

It does not. About 40% of the students fell into a very disturbing pattern - computational prowess with no basic understanding of the physics. I have colored these scores orange. Or some conceptual understanding with limited capacity to apply knowledge.

As an assessment of effectiveness of the lecture/recitation system, the data above is a portrait of a pedagogical disaster.

Do you have any questions about the diagram?
The FCI Revealed Many Student had a High Degree of Computational Skill but a Low Measure of Conceptual Understanding

Physics (or any discipline) without conceptual understanding is not an acceptable educational outcome.

Mazur, Peer Instruction 1995

Mazur's results were replicated in many different colleges and universities.
FCI is a very stable measure of conceptual understanding.
Its application revealed deep problems in the teaching and learning of physics.

Researchers such as Richard Hake of the University of Maryland and R. Sokolof of Oregon, and Bill Gerace of the University of Massachusetts Physics Education Group amply demonstrated the effectiveness of peer moderated techniques.
In *Peer Instruction* Mazur outlines four essential changes in instructional practice.

The first, the most profound, involves the reconceptualizing of the role of the physics lecture or physics class. A physics class must not be conceived as some “delivery system” that pours the knowledge into willing (or unwilling) student minds. The process is measurably ineffective.

The physics lecture must be interactive. Short presentations of ideas are followed by short formative assessment pieces involving students’ voices.

Students must bring something to the classroom. Reading must be done before the lecture. The lecture does not “present new material”.

Short “ConcepTests” serve as formative assessment of material learned.
Aspects of Peer Instruction

This slide comes from a talk by Eric Mazur.
Please note the large lecture format. The PI technique serves the same 300+ students, but with a totally different mechanism of class/instructor involvement.

Students at Harvard use infra-red “clickers” to convey their voting to the professor.

They turn to work with another 2-3 other students in their group to share ideas for the group vote.

The book *Peer Instruction* gives a complete introduction to the technique. The DVD *Interactive Teaching* (published by Pearson) provides an extensive view of PI and *Just in Time Instruction*.

See http://galileo.harvard.edu/

If interested we strongly urge you to look at these materials at the Galileo Project.

At the above site one can purchase the DVD shown in this presentation.
Structure of a PI driven class

Peer Instruction gives 300+ concept tests as well as the complete FCI and MBL tests.

At Dr. Mazur’s site Project Galileo
http://galileo.harvard.edu/ one can find a vast variety of materials for PI for many disciplines.

Project Galileo offers an online community to share ConcepTests.
The Measure of Change: Normalized Gain

Normalized gain on the FCI test to be the average increase in students' scores on the FCI divided by the average increase that would have resulted if all students had perfect scores on the post-instruction test.

(R. Hake 1992 Am. J. of Phys, 66, 1, pp. 64-74)

\[
\text{Normal.Gain} = \frac{\text{AVG}(\text{FCI}_{\text{post}} - \text{FCI}_{\text{pre}})}{\text{AVG}(\text{FCI}_{\text{perfect}} - \text{FCI}_{\text{pre}})}
\]

Collaboration is great. We all like to talk. So what!!

What is the basis of all this activity and the rush to throw out centuries of finely honed physics lectures?

**The core, as it should be is data.**
As a scientist one must believe data.

Dr. Hake describes a measure called the “normalized gain” on the FCI. It is a POST/PRE measure of what the group did gain, compared to what it could have gained over the course of instruction.
Normalized gain Traditional vs. Peer Instruction teaching

\[
\text{Normalized Gain} = \frac{\text{AVG}(FCI_{post} - FCI_{pre})}{\text{AVG}(FCI_{perfect} - FCI_{pre})}
\]

N=6,000 college students

Traditional Instruction: gain = 0.23 ± 0.04
Peer Instruction: gain = 0.48 ± 0.14

Gains as high as 0.60 to 0.80 have consistently been measured with PI.
Similar results hold for secondary instruction.

(R. Hake, 1992 Am. J. of Phys, 66, 1, pp. 64-74)

This data comes from Dr. Hake’s 1992 study in the American Journal of Physics. The sample is large N=6,000.

The results have been remarkably consistent over a decade or more of experimentation since.

Traditional instruction, no matter how well presented, or how thrillingly it may be laced with elegant demonstrations and other juicy additions, consistently delivers consistently a normalized gain of about 23%. The range is quite narrow +/- 4%

In itself this data is quite depressing.

Peer instruction techniques deliver gains double that.

With skilled practitioners of PI normalized gains as high as 80% have been recorded.

How can one continue to lecture in the face of this data?
Our students, are measurably scientifically illiterate.

PI may serve as a way to address and conquer this problem

Please refer, if you wish, to two recent articles on Peer Instruction in the NY Times.


Also see

### Universities using Peer Instruction

- Harvard
- MIT
- Stanford
- Cornell
- UC Berkeley
- California Tech

- Plus hundreds of others

The NY Times articles document changes at MIT.

Other major institutions now include PI as a matter of course in their physics instruction.

Even Cal Tech, the home of the cherished *Feynman Lectures*, employs PI as its major vehicle.
After this lengthy introduction this slide now presents the backbone of the talk. You had to experience interactive teaching first with the FCI probe on horses and jeans to understand what PI actually means.

There are two threads:

1) the work of Dr. Mazur of Harvard in his highly innovative pedagogical design called *Peer Instruction* and

2) CC’s parallel effort to bring highly interactive formative assessment in online courses.

It may be chance that brought questions like those from the FCI together with techniques long advocated by teachers advocating cooperative learning. In any case, the results of the measures from FCI were unequivocal. Cooperative discussion is far more effective than any elegant lecture format.

The interplay between high quality assessment such as the FCI and innovative pedagogical designs using cooperative learning produced a demonstrably effective combination of techniques.
Sample ConcepTest

A piece of metal is heated 500 degrees Centigrade.

Please vote as individuals:

1. The hole diameter will expand
2. The hole diameter will contract.
3. The hole diameter will stay the same.

2 minutes

We will now do a 2nd concept test. Again this example comes from the work of Dr. Halloun and Dr. Mazur.

As individuals: Please read the ConcepTest.

Pick an answer. Hold up a card. Don’t look at others’ answers.
Sample ConcepTest Group Work

A piece of metal is heated 500 degrees Cent.

1. The hole’s diameter will expand
2. The hole’s diameter will contract.
3. The hole’s diameter will stay the same.

Please vote as a group:

4 minutes

At your table, discuss your answer.

Select, as a group, the one answer you think is correct.

Hold up a card.
Video from the ILT DVD

INSERT a LINK to the “hole in metal” clip from Interactive Learning and Teaching DVD

Discussion of PI 5 minutes

Please contact Pearson to get this clip.
Or make it in house and mount at the ERC site.

Eric does not include video in his lectures. He puts in a LINK to the video and invites the viewer to view it in a browser.

He supplies only the PDF of the talk.

Insert text.
Please go to the above site to view the video from Mazur’s DVD Interactive Teaching and Learning.
This slide summarizes PI as set out by Dr. Mazur.

One uses well-tested items, the ConcepTest, as an instrument of formative assessment.

The ConcepTest INFORMS instruction and REFORMS teaching to a more effective usage of student and instructor time.

Clearly high stakes testing does give much information but it can also be used to inform teaching and professional development.

An objection can be made that the questions are “not well formed”, “there is no crisp answer.” This unformed nature is deliberate. It forces students to add information from their own store of knowledge. It forces them to bring “something to the table” for the discussion. This active engagement often highlights misconceptions other techniques, such as summative assessment, can not identify.
Resources on Peer Instruction

- 2,000 + published articles
- Interactive Learning Technologies DVD + online databank of questions
- *Calculus Concept Inventory*, Epstein of Duke University, and ConcepTests by Hughes-Hallet from Arizona

*Project Galileo* offers many resources on Peer Instruction.

The effectiveness of Peer Instruction is very well documented in the research literature.

Some textbooks in physics and mathematics have actively included peer instruction techniques.

Astronomy, psychology, biology, physiology instructors have authored peer instruction questions for sharing through Galileo.
Conceptual Probes

- Independently developed at CC parallel to Dr. Mazur’s work

- Online “conceptual temperature taking” for CC’s online PD courses

The Concord Consortium developed Conceptual Probes for use with online professional development for math and science teachers.

The conceptual probes have some similarities with PI questions but their application and format is different.

Initially they were conceived as “conceptual temperature takers” for online participants in professional development courses, to see if participants were willing to examine their conceptions, uncover and correct errors in thinking.

The participants, and their students took the probes.
Conceptual Probes

Characteristics

- Anonymous, anywhere, anytime
- Thought provoking interesting questions.
- Responses reveal misconceptions
- Possibly dynamic with video or simulations
- Very wide administration possible + database of responses.
- Nations or regions can administer conceptual probes.

Conceptual probes offer a more general application of the principles in ConcepTests.

- Thousands of responses are possible.
- All responses are anonymous. They can be sorted by ID used for entrance. We did not list names with the ID.

- It is possible to administer conceptual probes to large student or teacher populations anonymously.

- They can serve as follow up and discussion pieces locally or online to any professional development or regional meetings.

- Teachers or administrators can author their own probes.
Take a piece of paper.
Roll it into a cylindrical shape with the long edge as the cylinder’s end and imagine it capped at both ends.

Take the same paper into a cylindrical shape, this time with the short side as the cylinder’s end.

SAMPLE CONCEPTUAL PROBE:

Author: Dr. Judith Collison, Concord Consortium

Take a piece of paper, roll it into a cylinder in two different ways:
Case ONE: Long side vertical as the height of the cylinder.
Case TWO: Long side horizontal as the circumference of the cylinder.

Which has more volume: Case One or Case Two?
Please take 2 minutes to think about this as individuals:

Hold up a card.
Sample Conceptual Probe

Consider the two cylinders formed.
Please vote as individuals:
1. If the VOLUMES are the same
2. If the VOLUMES are different

1 min.

Please Vote:

We will tally responses.
Sample Conceptual Probe

Consider the two cylinders formed.

Please vote as a group:
1. If the VOLUMES are the same
2. If the VOLUMES are different

Discuss for 4 minutes

Please take 4 minutes to discuss your answer with your table.

You must convince your table to pick one answer.

Please vote.

The class instructor invites several tables to give a reason for their answer.
Above is a screen shot of a sample conceptual probe.

The scroll bar permits viewers to look at responses. There are no names with the response, only the reasons are captured.

The software captures 30-50 random responses for display at each viewing.

This probe is assessing mathematical dispositions, jumping to conclusions, and reluctance to give up the “common sense answer”. Researchers can look at all responses to assess responses sorted by school or by region.

Teachers can use a print out to invite students to look at a wide variety of reasons and assess them for deeper thinking.

Is there any sense of responders using knowledge of volume of a cylinder? Are they reluctant or at all hesitant to think quantitatively?
Conceptual Probe Discussion

What did you learn from the exercise?
Conceptual probes can be administered anytime, anyplace.
A database holds all responses, including reasons for answers.

The question can be modified to include a follow up on limits. What if the paper were cut in half the long way and a new wider but fatter cylinder were made? Would the volume be the same if the cut were made on the shorter side?

What happens to the volume of the cylinder if one continues the halving?

What happens to the surface area if one continues to halving? Is it the same result for halving the long and the short side, or is it different?
Facilities in the Molecular Workbench enable web-authoring and collecting of data on conceptual probes.

The conceptual probes can include animations or movies of molecular models built in the Molecular workbench. External video or images may also be used.

Sorts on tallies inform the instructor.
Where is the innovation?

- Assessment is an opportunity for learning not outside of instruction.
- Assessment informs pedagogy
- Assessment is ongoing

Both ConcepTests and conceptual probes assess the willingness of students to engage with problems, see the need for using the skills taught in solving problems, and push them towards confronting and correcting misconceptions.

They also help the teacher see where misconceptions have a stronghold, and where instruction failed to bring about understanding. Assessment becomes diagnostic and an opportunity for learning and reaching a deeper understanding of major concepts.
Formative Assessment: Effectiveness of ConcepTests and Conceptual Probes in Math and Science

Dr. George Collison & Dr. Judith Collison

Please do not hesitate to write us with any questions or inquiries about projects.

George@concord.org

Judith@concord.org