Measuring the teaching and learning that matters

I. Identifying what learning matters
II. How to best measure this learning
III. Assessing teaching
<table>
<thead>
<tr>
<th>Assessment</th>
<th>meaningful</th>
<th>useless</th>
<th>useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLA, NSSE</td>
<td>4 yr grad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>most teaching</td>
<td>eval.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>program eval.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

But Easy!

“useful” = directly involves people responsible/can change outcome shows how to improve (individually & collectively) practical for widespread regular use allows peer comparisons
Rest of talk—meaningful assessment *(that is also useful)*

1. Learning—in Courses & Programs
2. Teaching
Assessing meaningful learning—

Programs—desired overall expertise

What students are able to do.
What decisions they will make.

Courses—broken down into pieces, detailed learning goals
Fit together to achieve program goals

NOT!!
“Understand x, y, z…”
“Know…”
“Think critically”
Measuring learning—

Goal-- read student’s mind.  
How much better have they gotten at thinking & making decisions like expert?

**Use proxies to measure**

Give task. From performance on task, thinking like expert?

Attitudinal

Give choice & see response (next course, choice of major, ...)

More explicit the expert thinking required by the task (more authentic), more accurate the proxy.
Common ways assessments of learning go wrong --

- There are alternative (nonexpert) ways to come up with the correct answer *(memorize without understanding procedures that apply to a specific situation, and tested only on that situation)*

- measuring things expert never needs

- solution requires some subtle trick *(requires “figuring out instructor, not subject”)*

What is the nature of expertise/expert thinking that want students to develop?
Expertise research*

Historians, scientists, chess players, doctors,...

Expert competence =

• Factual knowledge

• **Mental organizational framework** ⇒ retrieval and application patterns, relationships, scientific concepts,

• Ability to monitor own thinking and learning

*Cambridge Handbook on Expertise and Expert Performance*
Some components of S & E expertise

- concepts and mental models + selection criteria
- recognizing relevant & irrelevant information
- specialized procedures & criteria for when useful
- does answer/conclusion make sense- ways to test
- **model** development, testing, and use
- moving between specialized representations

Only make sense in context of topics.
Knowledge essential, but only as integrated part- *how to use/make-decisions with that knowledge.*
Compare with typical HW & exam problems that provide practice and assessment

- Provide all information needed, and only that information, to solve the problem
- Say what to neglect
- Not ask for argument for why answer reasonable
- Only call for use of one representation
- *Possible* to solve quickly and easily by plugging into equation/procedure

- concepts and mental models + selection criteria
- recognizing relevant & irrelevant information
- what information is needed to solve
- How I know this conclusion correct (or not)
- *model* development, testing, and use
- moving between specialized representations (graphs, equations, physical motions, etc.)
How to do better assessment

1. Identify expert thinking/problem solving - “cognitive task analysis”

example-
“We want students to learn to be good field geologists.”

Assessment?

A. Send them out to work, see what employers say.
    Authentic, but slow & nonspecific feedback.

B. Better--What do good field geologists do?
    Spend lots of time making geological maps of an area to determine the types and distributions of rocks.
    Assessment-- send student out to make geo map of an area.

But:
• Just tells you if student can do it or not.
• Does not tell much about detailed strengths and weaknesses, how to teach to improve outcomes.
• Lots of time and effort required
Better-- what are the cognitive processes that experts use in making a geological map?
Cognitive task analysis of making geological field map

• **Planning**
  1. Examine the topographic map and plan a hiking path for your day.
  2. Identify key areas that you plan to cover.
  3. Think about criteria you will use to decide when to modify plan when in the field.

• **Systematic collection of relevant data as plan is followed**
  1. Identify minerals and rocks in the field at the outcrops you encounter, mapping them and taking careful notes as you go.
  2. Assess change from last outcrop (measure sufficiently to check), characterize change;

• **Interpret data by formulating and testing alternative models.**
  1. Interpret data you have collected. How different is current outcrop from previous outcrop? If difference is large, take additional data, revise interpretation as needed.
  2. Identify structural trends/relationships in the data you have collected.
  3. How well does the data match your “preferred” and alternative tentative models?
  4. How big is the extrapolation from data to your model?
  5. Do the geologic contacts and structural relationships in models “make sense?”

• **Predictive testing and refinement of model/map**
  1. Identify other areas in the field area where the model(s) can be tested
  2. Include these areas in plans for future field days.

Repeat these steps and/or iterate as needed.  

**take 2 minutes. apply to own field.**
Good *(practical and authentic)* Assessment-- break task down into the expert steps.

- Here are set of outcroppings, create model of the geological structure. Explain reasoning.
- Which are the regions where model is most uncertain? Why? How could you get data to improve?
- Here is a map. Here is a set of outcroppings that are observed. Is this map correct? Why or why not?

What if 200 students, 2 hours of TA time?
Here are set of rock outcroppings, which of the following 6 models/map images are consistent with them? Which of the following 6 new models are inconsistent? For this map, geology of which of points A-H is the most uncertain?
Physics problem solving example of cognitive task analysis
• visualize problem and appropriate model (make sketch, label relevant physical quantities)
• what concepts apply?. Criteria for deciding.
• what is plan for solving?
• what strategies useful or not? what procedures will work? (what are similar or analogous problems?)
• what information is needed to solve? what is irrelevant?
• how to convert model to mathematical representation?
• what are appropriate calculation procedures? how execute to get answer?
• does answer make sense? (come up with criteria: units correct, limiting cases, symmetry, related cases where know answer)
Best course level assessments of learning- Departmentally owned, **limited** set of test questions/specific guidelines that are used in **every** offering of course. Represent key learning outcomes. Pretested for validity (student interviews) Periodically revisit to ensure are good proxies.
II. Meaningful and Useful assessment of teaching

Meaningful—how effective it is at achieving the desired learning and student success?

What do we know about effective teaching and learning?
Learning expertise*--

Challenging but doable tasks/questions
Practice all the elements of expertise with feedback and reflection.

Requires brain “exercise”—intense engagement

Effective teacher- “cognitive coach”
• designing practice tasks
  (what is expertise, how to practice, proper level)
• feedback/guidance on learner performance
• motivation

* “Deliberate Practice”, A. Ericsson research accurate, readable summary in “Talent is over-rated”, by Colvin
Number of different teaching methods implementing these principles. Many hundreds of research studies showing effectiveness (usually relative to lectures).

Strong correlation between specific teaching practices used and amount of learning. *(my talk yesterday)*
II. Meaningful and Useful assessment of teaching

Meaningful—how effective at producing the desired learning?
Consideration of the options.

1. Student course/faculty evaluations—predominant method.
   Most recent meta-analysis:
   • Correlations with learning very small, zero in STEM.
   • Large correlations with factors not under instructor control: course level, subject, class size, instructor appearance.
   Say nothing about how to Improve.
   • Poor at recognizing relative effectiveness of teaching practices.
     = meaningless and useless

2. Classroom observations by random/untrained faculty.
3. Validated classroom observation protocols with well trained observers. RTOP
Meaningful, but takes 10-20 hours of training+ 2 hours observing/course, and only captures in-class part of teaching.

**Impractical for widespread regular assessment—too time consuming.**

4. Teaching portfolios—too varied to say much about. Impossible to do comparisons.
5. Teaching Practices Inventory*
2.5 years developing at UBC. Completely characterizes all elements of teaching a course. Tested with ~ 250 course offerings. 8 categories, 51 items. 
~10 minutes per course to complete

Allows meaningful comparison of all STEM courses except labs and project/seminar courses. *(likely works for other subjects as well, but not yet validated)*

Allows measure of use of effective teaching practices *(research shows result in greater learning)*

*http://www.cwsei.ubc.ca/Files/CWSEI_TeachingPracticesSurvey.pdf*
<table>
<thead>
<tr>
<th>I.</th>
<th>Course information provided</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Information about the courses, such as list of topics and organization of the course, and learning goals/objectives.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II.</th>
<th>Supporting materials provided</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Materials provided that support learning of the course material, such as notes, video, and targeted references or readings.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III.</th>
<th>In class features and activities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>What is done in the classroom, including the range of different types of activities that the instructor might do or have the students do.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV.</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nature and frequency of the homework assignments in the course.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>V.</th>
<th>Feedback and testing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Testing and grading in the course, and the feedback to students and feedback from students to instructor.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VI.</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assorted items covering diagnostics, assessment, new methods, and student choice and reflection.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VII.</th>
<th>Training and guidance of teaching assistants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>What selection criteria and training is used for course teaching assistants, and how their efforts are coordinated with other aspects of the course.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VIII.</th>
<th>Collaboration or sharing, use of research, in teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Collaboration with other faculty, use of relevant education research literature, and use of educational materials from other sources.</td>
</tr>
</tbody>
</table>
Effective teaching practices scores for 5 math & sci depts

High degree of discrimination within and across departments
Shows **how** individuals and dept’s can improve.
Measure of improvement.
Department that had major program to improve how they were teaching.

**D3**

![Bar chart showing extent of use of research-based effective teaching practices across different courses.](chart.png)
TPI— not perfect way to assess quality of teaching. **Just much better than anything else available.**

“useful” = *directly involves people responsible/can change outcome shows how to improve (individually & collectively) practical for widespread regular use allows peer comparisons*

“meaningful” = proxy that research shows correlates with desired outcomes. Amount of learning and student course completion. Particularly learning of expert-like thinking.
extras with details below
Learning Measurement Design Concepts

1. To change teaching, must be specific to particular courses. NESSE (student overall experience and engagements) and CLA (learning in college) useful, but too general to guide change.

2. At university level, external assessments not going to work. Faculty have so much expertise, will reject something they do not see as valid.

3. Must be specific to discipline, consensus as to value, specific enough to measure single instructor impacts.
Tips for developing assessment tools.

1. Interview largest possible range of people. Patterns and expert-novice differences more obvious.

2. 100+ student classes in large university don’t vary year-to-year.

3. Best questions: a)measure important aspect of student thinking and learning. b) measure aspect instructors care about & shocked at poor result.
What to Measure?

- Conceptual understanding (= expert model or mental framework for subject, can make good argument = transfer)
- Beliefs about science (framework of knowledge, how learned and used, interest and relevance)
- Problem solving – many components, includes above and others, some discipline specific, many other general.
- Communication skills
- Teamwork skills

Often don’t assess these, but instead assess factual knowledge (memorization) and mastery of processes.
How to Measure?

General approach:

1. Establish what the characteristics of experts are.
2. Observe and analyze in detail student learning—path of novice to expert development.
3. Establish questions that distinguish students location along path.
4. Give tests pre and post instruction (individual class, start and end of course—learning gain, longer term—what sticks)

(may or may not be connected with marks)
How to Measure – Conceptual Understanding

1. Choose things that faculty in the department agree are important; faculty establish learning goals.

2. Extensive student interviews: identify basic concepts that students get wrong, and why.

3. Design questions; include common student misconceptions as possible answers.

Avoid problems that can be solved by simple memorization. Require seeing overall framework of ideas, how they can be applied generally, understanding of deeper structure.

a. Apply to new context.

b. “Redesign problem” e.g. “What are possible ways I could change an incandescent light bulb to make it brighter?”

c. “Troubleshooting problem”— something stopped working, what are possible things that could have happened to cause this change?
(Very helpful to test assessment before using officially. Give to a few students, have them do while thinking aloud.)

Are they solving using intended methods or strategies or finding “non-expert” shortcuts. If they are unsuccessful, is it because they are not using expert approach, or some other reason like not understanding wording of the question?
tip--If homework, use open ended short answer questions including asking for explanations, but only mark 1/3 of the problems.

Computer grading of expert thinking (makes it hard to just memorize without conceptual understanding and problem solving approaches:
“Of this list of 12 choices, select which information is relevant (or irrelevant to solve the problem).”
“What information do you need to solve this problem that is missing?”
“If X stops working as well as before, list which things might have changed that would explain this behaviour.” (can also make quantitative if desire, not just what changed, but by how much)
“What changes would make x work better?” (faster, hotter, more radiation out, ...)
What Does It Mean To Achieve In Science?

- **Declarative knowledge**: knowing that--facts and concepts in the domain
- **Procedural knowledge**: routine procedures and some aspects of problem solving
- **“Schematic” (analytic) knowledge**: conceptual models of how the natural world works
- **Strategic (“transfer”) knowledge**: knowing when, where and how knowledge applies
- **“Epistemic” knowledge**: knowing how we know—knowing how scientific knowledge is built and justified
- **Communication & social skills**: ability to communicate ideas clearly and concisely in the genre of science, team work

from Rich Shavelson talk
Remember not to leave out sense-making steps. Any missing step that can be a failure point will result in less meaningful assessment. If student fails, but you never assessed particular failure mode, you will not know they failed, if it was because of a step that is never measured.
Steps (6) in test development
(~ 6 months post-doc)

1. Interview faculty-- establish learning goals.
2. Interview students-- understand thinking on topic ⇒ patterns emerge. Find what traditional exams are missing.

“Expert-like thinking” best target.

I. Way knowledge in subject is organized and applied = “Concepts”
II. Way experts approach learning and problem solving
Failures of standard exams

**electricity**

End of course.
70% can calculate currents and voltages in this circuit.

learning to solve problems by memorizing recipes-- not using conceptual understanding

only 40% correctly predict change in brightness of bulbs when switch closed!

* from E. Mazur
Steps in test development

1. Interview faculty--
2. Interview students-- understand student thinking

3. Open-ended survey questions to probe.
4. Create multiple choice test-- answer choices reflect actual student thinking.
5. Validation interviews on test-- experts agree on questions and answers; students interpret correctly and pick answers for intended reasons.
6. Administer to classes-- run statistical tests on results. (some standard psychometric criteria do not apply. measure independent concepts & responses worse than random)

Publish, use at multiple institutions!
Validated Concept Inventories

FCI (intro mechanics)
BEMA (intro electricity and magnetism)
FMCE (intro mechanics)
QMCI Quantum mechanics concept inventory (intro quantum)
CUSE (3rd year electricity)

Concept inventory tests under development or in early use in geology, chem, biology, physiology, ...
II. Measuring expert thinking beyond content
(nature of subject, how learn)

attitudes and beliefs, epistemology

same process
1. interviews⇒ expert & novice thinking
2. create measurement instruments
3. validate
4. use to measure impact of instruction (pre-post)
### Beliefs about physics/chem

<table>
<thead>
<tr>
<th><strong>Novice</strong></th>
<th><strong>Expert</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Content: isolated pieces of information to be memorized.</td>
<td>Content: coherent structure of concepts.</td>
</tr>
<tr>
<td>Handed down by an authority. Unrelated to world.</td>
<td>Describes nature, established by experiment.</td>
</tr>
</tbody>
</table>

*adapted from D. Hammer*
The CLASS Survey  (CLASS.Colorado.edu)  
(Colorado Learning Attitudes about Science Survey)

- **Main Goals:**
  - Focus on *beliefs about the discipline and learning the discipline*
  - Valid/Reliable across university populations (non-sci to majors)

- **CLASS-Phys (42 statements) & CLASS-Chem (50 statements)**
  (7-11 minutes to complete)

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

*I think about the physics I experience in everyday life.*

*It is possible to explain physics ideas without mathematical formulas.*

*Understanding physics basically means being able to recall something you've read or been shown.*

Builds on previous work in physics by (MPEX & VASS)
Distribution of Beliefs

- Alg-based Phys I (N=309)
- Calc-based Phys I (N=389)
- 3rd semester for phys majors (N=61)

'Overall' % Favorable (PRE)
(Percentage of statements for which student agrees w/ expert)
Beliefs about physics/chem

**Novice**

Content: isolated pieces of information to be memorized.

Handed down by an authority. Unrelated to world.

Problem solving: pattern matching to memorized recipes.

**Expert**

Content: coherent structure of concepts.

Describes nature, established by experiment.


CLASS survey % shift?

nearly all intro physics courses ⇒ *more* novice

ref. Redish et al, CU work--Adams, Perkins, MD, NF, SP, CW

chem as bad or worse

*adapted from D. Hammer
Impact of teaching on students’ beliefs

• Maryland MPEX survey:
  Students’ expectations shift to be *more* novice
  (decline of ~5-8% in ‘Overall’ %fav)

• CLASS-Phys results similar:
  (example data from Univ. of Colorado)

![Graph showing shift in % Favorable (Post-Pre)]

- **Calc-based Phys I** (N = 389, 348, 398)
- **Calc-based Phys II** (N = 218)
- **Alg-based Phys I** (N = 128, 312, 306)

minor changes addressing beliefs
Retention in the major
Who from Calc-based Phys I, ends up as successful physics major 3 yrs later?

Calc-based Phys I (Fa04-Fa05): 1306 students

Beliefs at **START** of Phys I

~ all students who stay in physics, arrive at university with expert-like beliefs
Student attitudes & beliefs important!

Correlate with

- retention
- **interest**
- content mastery

Affected by instruction (as measured by CLASS)

- negatively with conventional instruction
- positively if properly addressed

(recent UM & FIU results)

CLASS- physics, CLASS-chem done
CLASS -Biology, math, earth sciences in development
Of the following 10 possible segments marked on map, which would be the 5 most important to walk?

For this segment, which of the 10 possible reasons would be relevant to why you would want to walk it?

For this segment, rank the 5 most important reasons for walking, in order of priority.

For this new segment, pick all the reasons that would make it less desirable a path to walk.
Back to 10 students, lots of TA time

- Here is topo map and what is known about geology in this region, draw route you would walk to collect data for map. Explain reasons for different segments of route.

200 students, 2 hrs TA time  How would you do?

Suggestions?