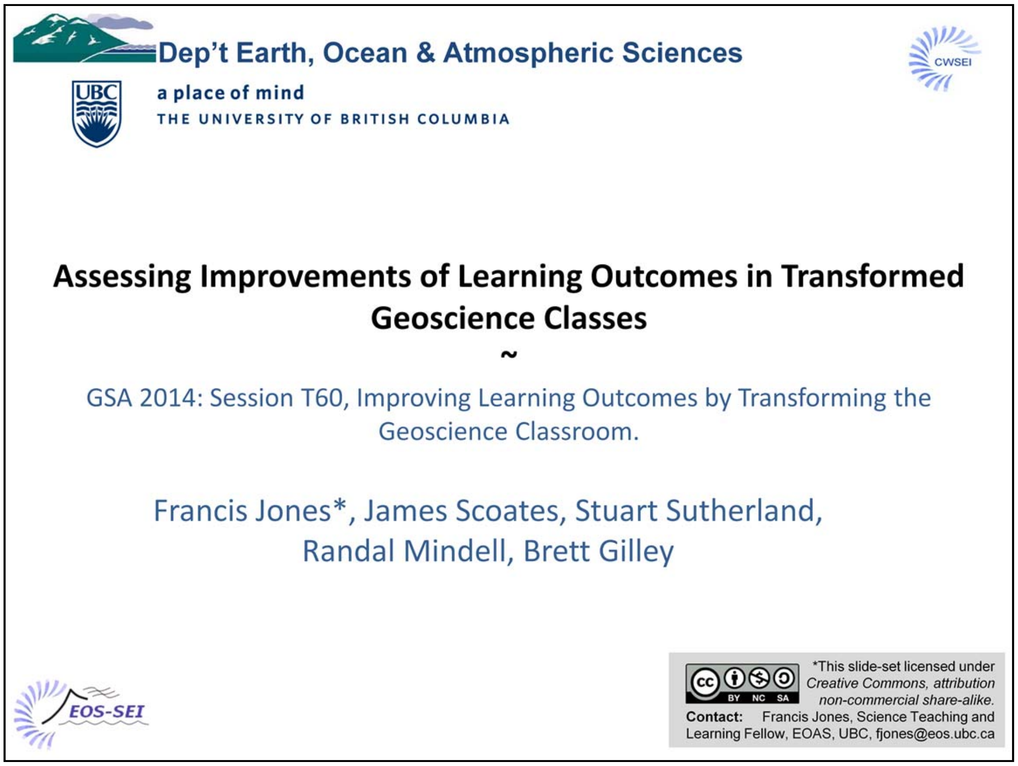
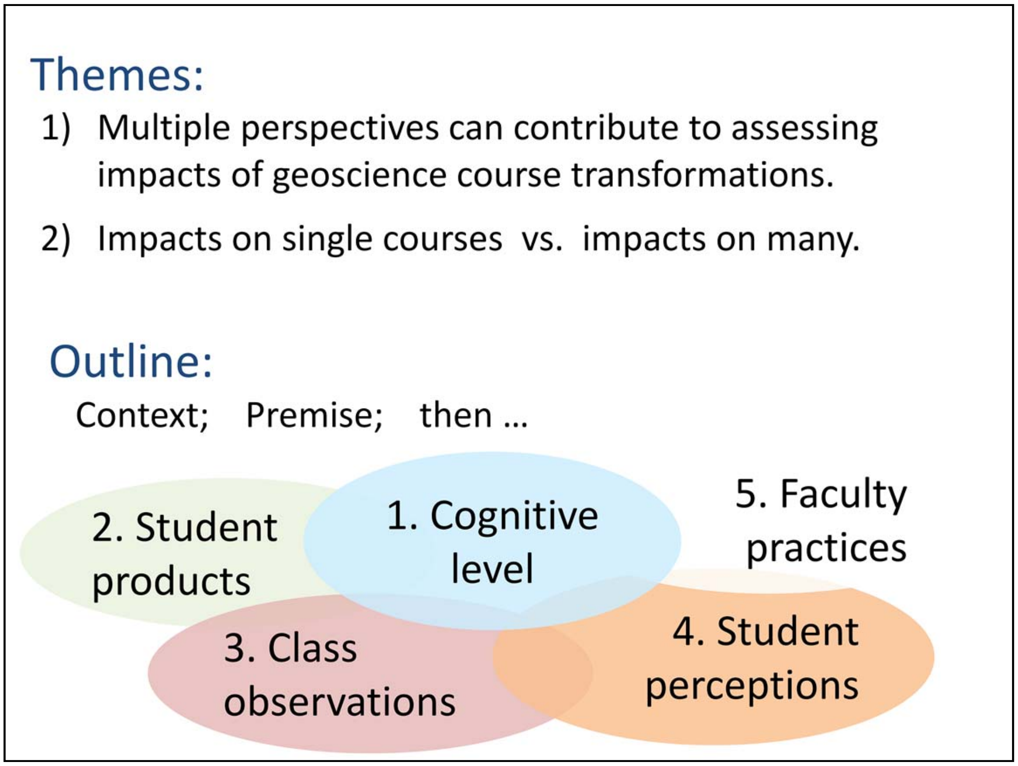



Hi, in this last presentation of the session, I'm going to step back a bit and consider ways of assessing the impact of transformations on many courses within a single department.

The other authors are Professors James Scoates & Stuart Sutherland who teach the courses from which examples were drawn and Randal Mindell & Brett Gilley who supported improvements in these courses.




In my abstract I ambitiously identified six aspects – I think here I'll touch on these five, and will be happy to chat about any of these, or others after the talk.

## Context

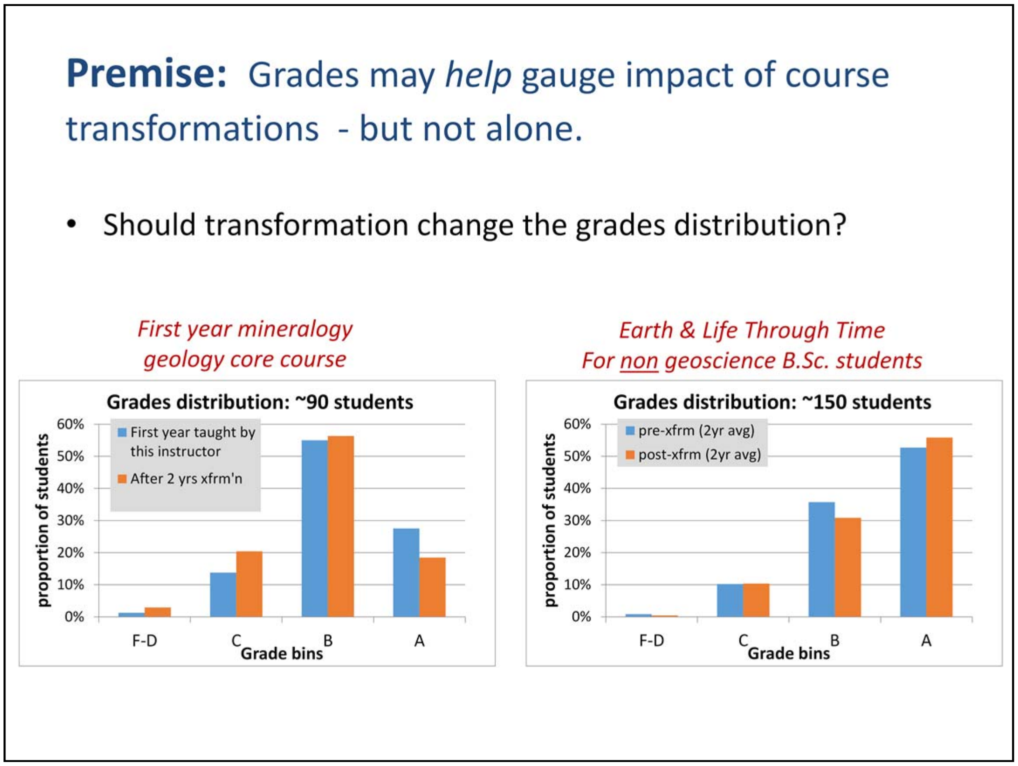
EOS-SEI

- UBC's Dep't Earth, Ocean and Atmospheric Sciences – EOAS, with ~50 research and teaching faculty
- 7 year funded science education initiative;
  – the Carl Wieman Science Education Initiative, CWSEI.
- Since 2007, between 2-4 full-time "*Science Teaching and Learning Fellows*" = geoscientist with education commitement.
- Affecting 38 / ~60 courses with ~80% faculty participation.
- Details: http://eos.ubc.ca/research/cwsei/ & http://cwsei.ubc.ca/

We teach in the Dep't of EOAS at UBC. This is a mid-to-large research focused department which was fortunate to benefit from the Carl Wieman Science Education Initiative between 2007 and 2014.

Over that period we have been able to significantly impact many of our 1st, 2nd and 3rd year courses, and some of our 4th year courses.

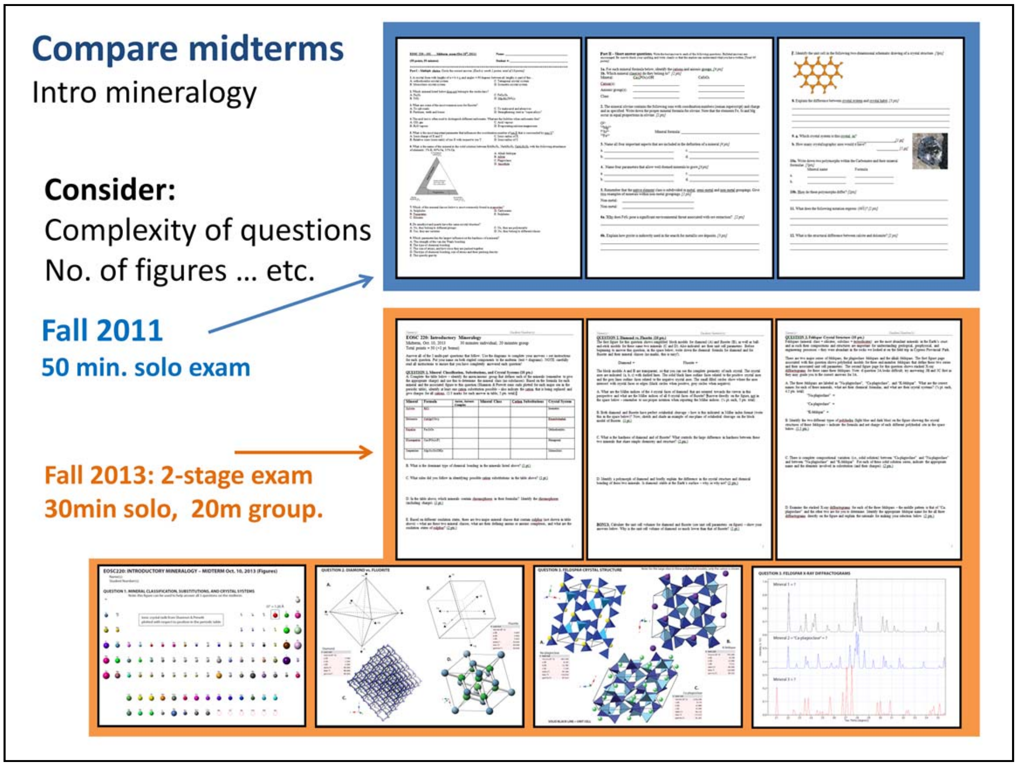
As many of been saying in this and other sessions, improving teaching seems to benefit more from emphasizing faculty improvements rather than course adjustments. This is the approach we've been taking since the beginning, and our participation rate suggests it has worked quite will in our setting.




So – as a starting point, assessing improvements in learning outcomes is not just a matter of measuring changes in grades or grade distribution. The reason is of course that geoscience course transformations will inevitably lead to students learning and doing different stuff and in being assessed in ways that are different from before – hopefully in more sophisticated ways that are based on evidence about how learning works.

These two figures show grade distributions for two courses both before and after their respective 2 year “transformation” efforts. One is a core course for geology majors – class averages are in the high B’s – and the other is a service course with averages in the A’s.

We could argue about whether these grade distributions are good, bad or indifferent but that’s a discussion for another day ☺ The point is – taken alone, these distributions have no information on improved learning outcomes.




A first order approach can be to simply compare pre- and post-transformation exams. The ones shown here are 2 years apart.

The difference in how students demonstrate more sophisticated capabilities (learning outcomes) is fairly evident.

But can this be quantified or at least characterized in repeatable, reliable ways?

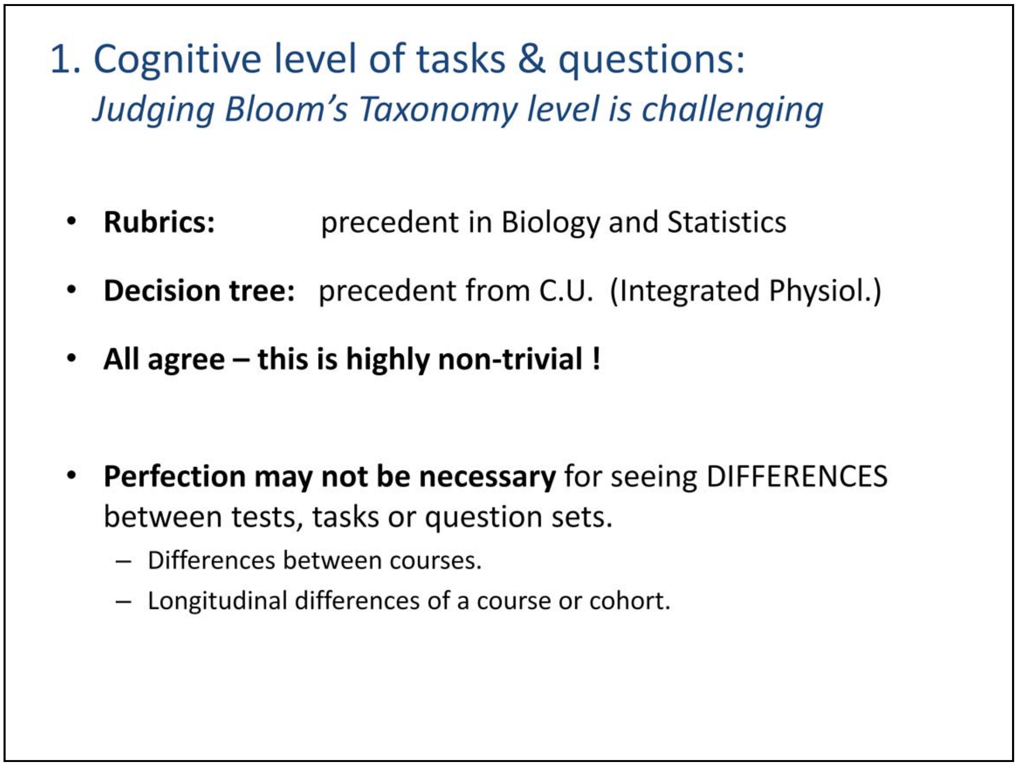
## Exam format and sophistication

| Aspect of midterm exams: | 2011 | 2013 |
|---|---|---|
| Time for individuals | 50 mins | 30 mins |
| 2-stage time for groups | 0 | 20 mins |
| No. qns | 25 | 19 |
| Avg. parts per qn | 1.3 | 2.6 |
| No. of figures | 3 | 7 |
| Proportion mult. choice | 35% | 0% |

“Improvement” is suggested but learning outcomes are not visible.

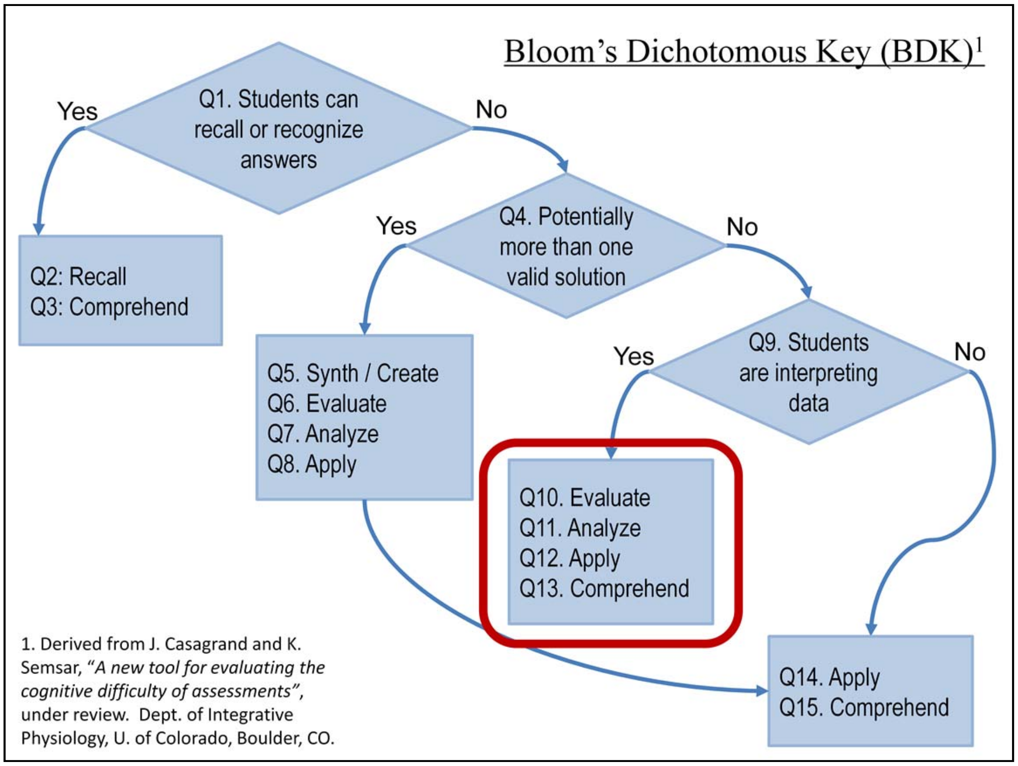
Difference in sophistication of these two exams is fairly evident. The example is from introductory mineralogy illustrates several aspects:

- The 2013 exam was a two stage exam.
- There are fewer questions but they are more involved and employ many more figures.
- There are now no MC questions.




But we would like a more quantitative comparison of learning outcomes before and after transforming a course. We want to assess in a reliable, repeatable manner, the sophistication of questions or tasks students are being assessed on.

There is significant precedent involving ad-hoc judgments of Blooms level, use of discipline-specific rubrics, and decision tree approaches. All agree this is a task that is difficult to make entirely reliable and repeatable.



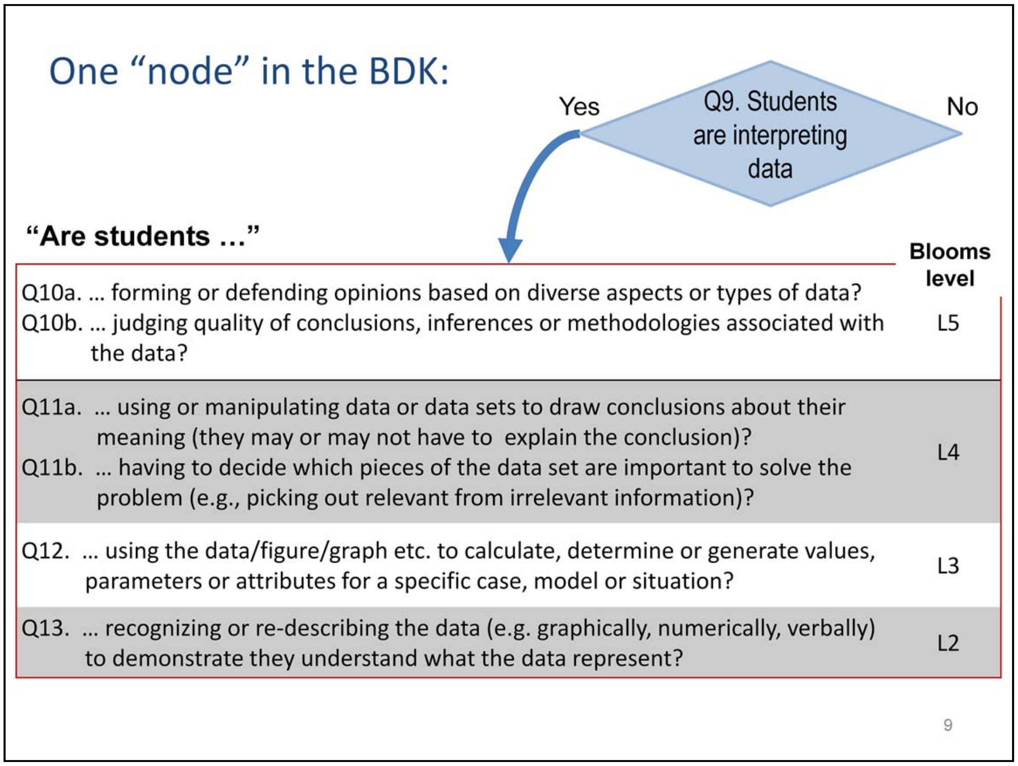

Here is an outline of the decision tree approach we have been working on, which is based on work of Casagrand and Semsar at U. Colorado (Physiology). Benefits are

1. Each judgement is made using a consistent sequence of decisions.

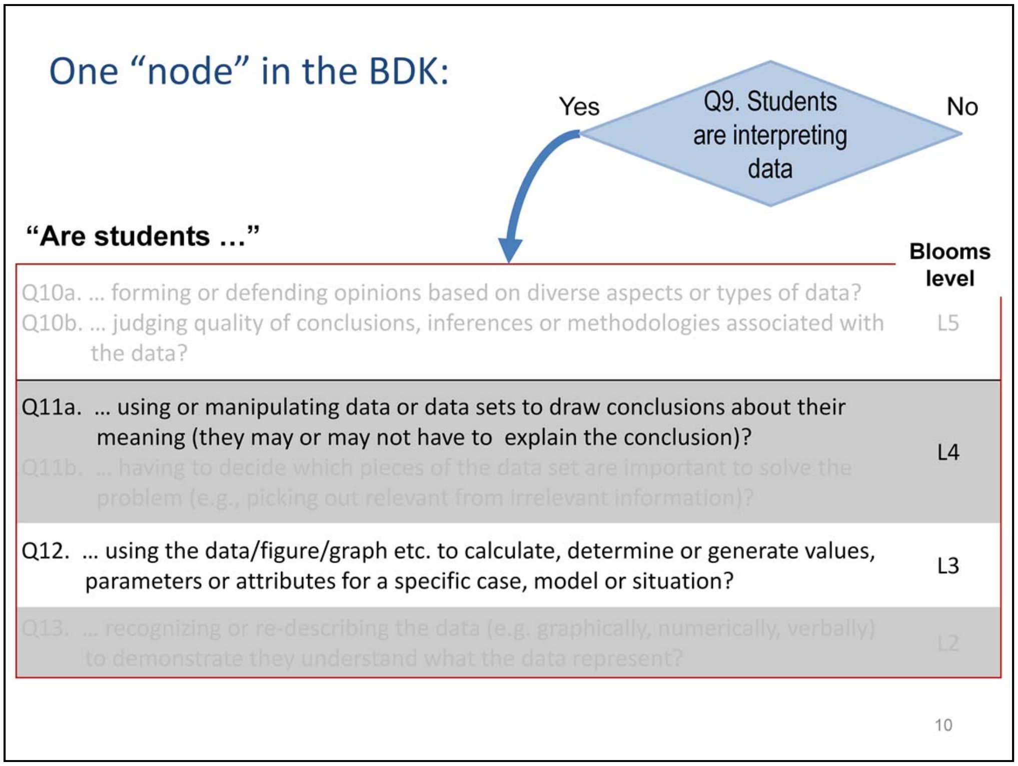
2. We have found these three key decisions points are appropriate for the first two settings we've used (mineralogy and Earth History), and we will be evaluating their relevance in other geoscience courses.

3. Crucially, labeling each decision like Q5 etc. facilitates comparison of how different rankers made decisions about Blooms level.

We find this process attractive but that the devil is in the details. Let's expand one node – the "*students are interpreting data*" node.




We found we had to very carefully refine the wording for each decision point. Also it was helpful to NOT include words indicating which of Bloom's cognitive levels are associated with each decision to help prevent "colouring" our decisions with those words.



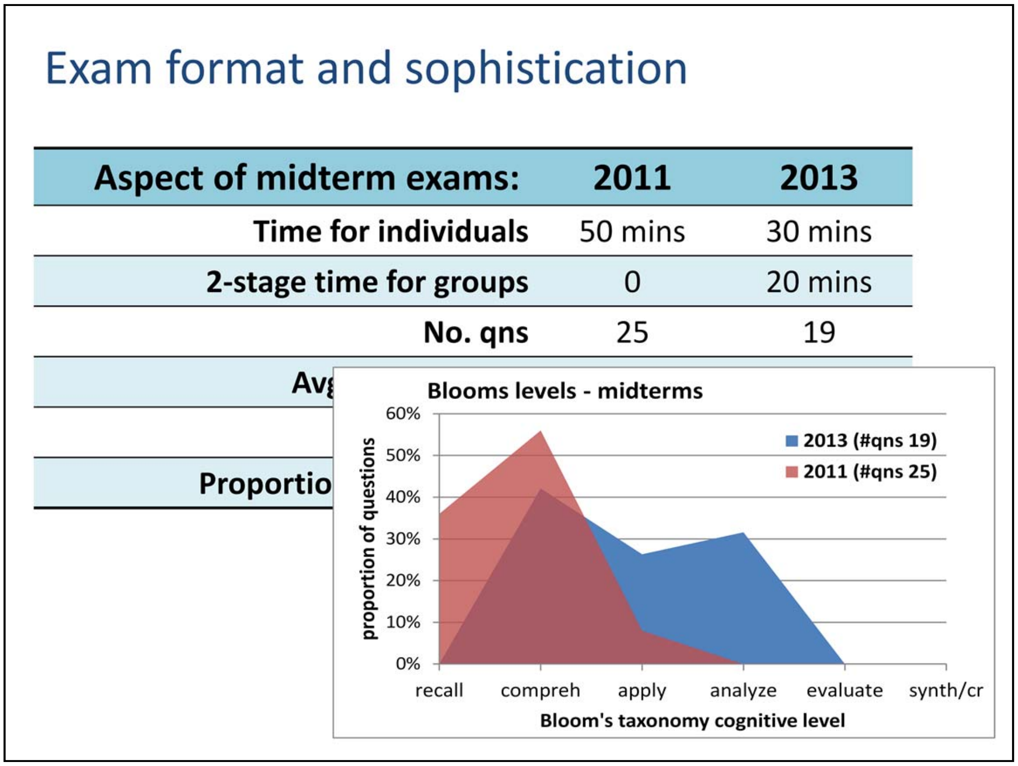

As one example, the distinction between "apply" - Q12 and "analyze" - Q11 is important but subtle and frequently debatable.

## We find that repeatable results depend upon

1. Knowing what the instructor intended students to actually do – eg. solutions & corresponding thinking.
2. Knowing what students saw / heard and could have memorized.
3. Very careful wording at each decision node.
4. Training to recognize how question *verbs* relate to specific cognitive levels.
5. Discussion to resolve discrepancies.

11

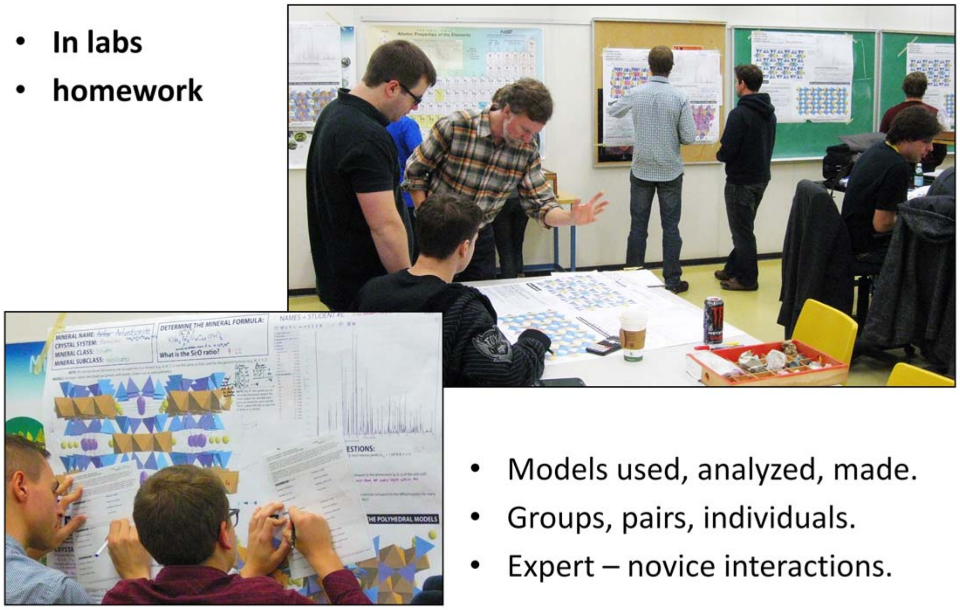
So far (and this is very much still a work in progress), we have found that repeatable results depend on these 5 points.




Applying results of this cognitive level analysis to the earlier comparison of midterm exams now helps demonstrate that learning outcomes being assessed have shifted from mainly "*recall and comprehend*" cognitive skills in 2011 to more "*apply and analyze*" tasks in 2013.

## 2. Student products; What do they make or use?

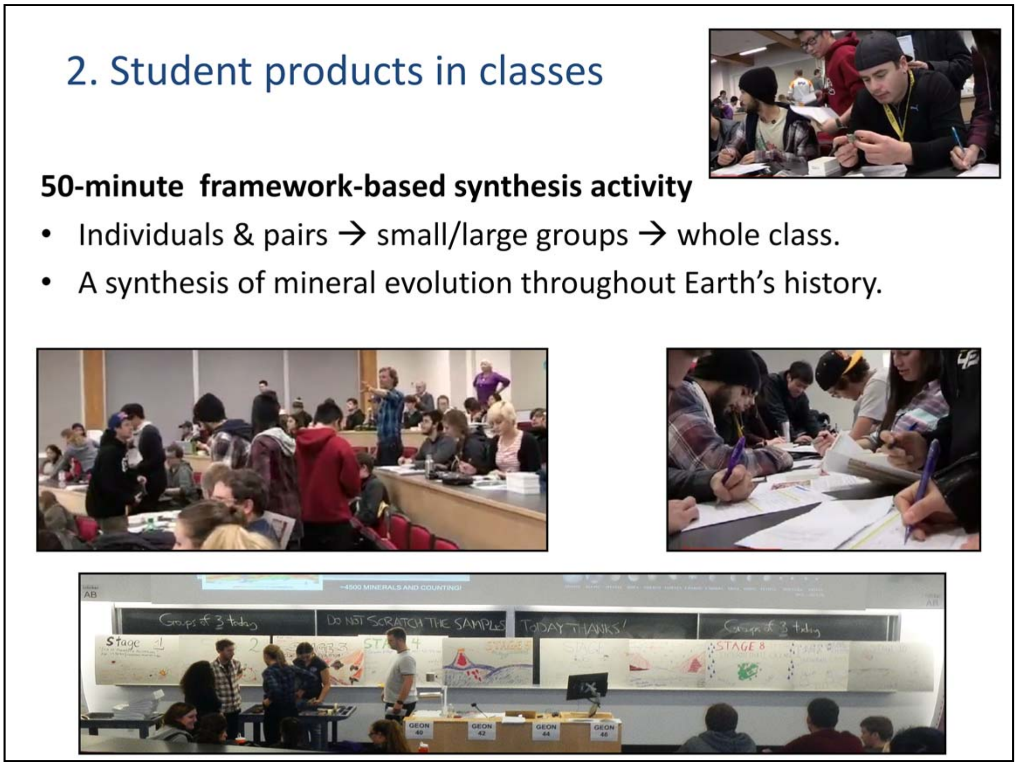
- **In labs**
- **homework**




- Models used, analyzed, made.
- Groups, pairs, individuals.
- Expert – novice interactions.

Another aspect that speaks to how learning outcomes change due to transformations involves the actual STUFF students use and produce. Here we see sophisticated crystallographic models used throughout this course by instructors and students, in class, labs and assessments. You can see discussion, argumentation, articulation of concepts and all that great cognition that promotes deliberate practice and learning.

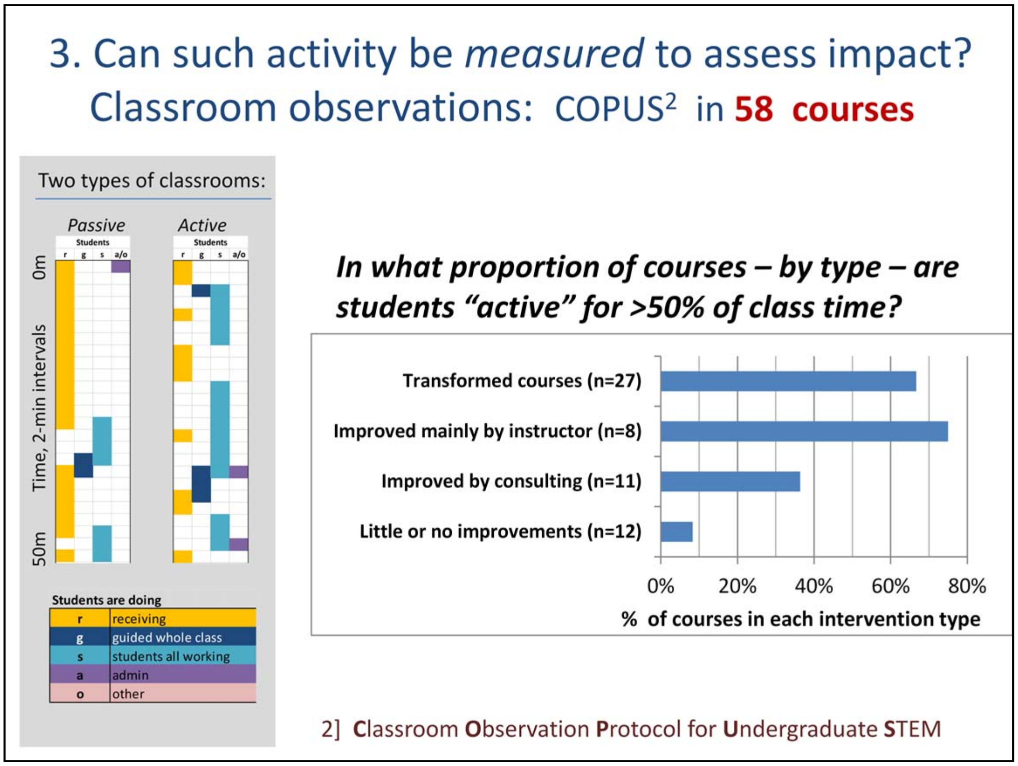
These were never part of this introductory course before.




In the classroom, this course is now highly interactive and student-centric, with plenty of peer-peer and expert-novice interaction.

The course is also well-grounded in a persistent set of framework concepts that students relate to from beginning to end.

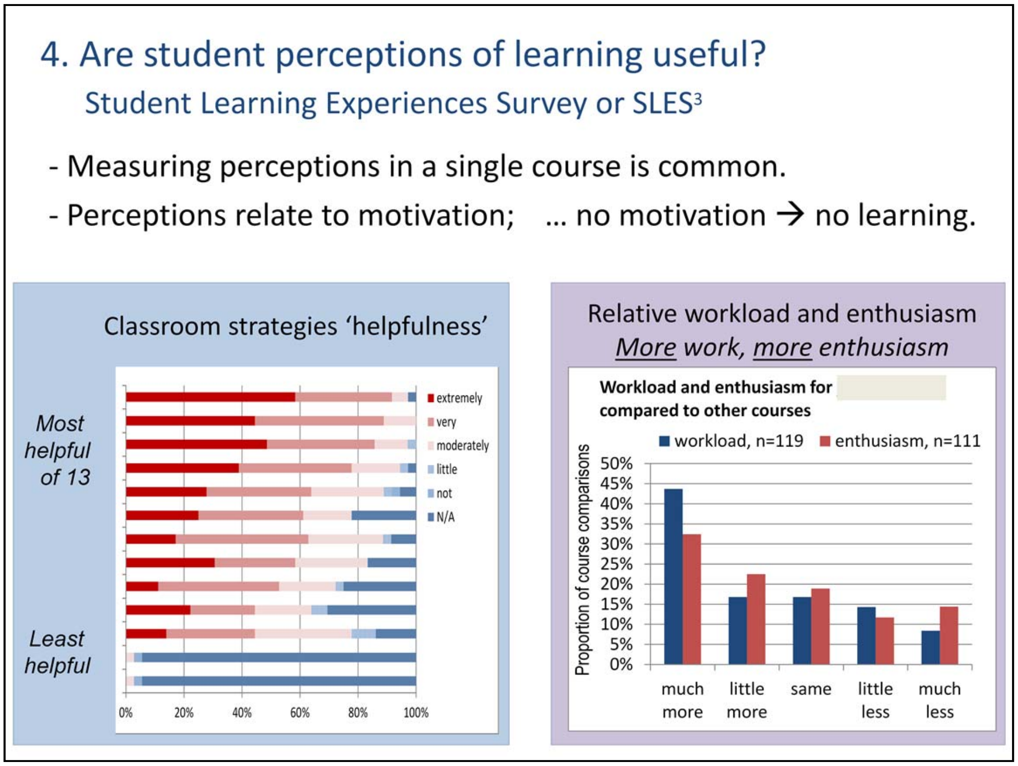
Here we see a capstone activity that takes a full 50 minutes and results in an illustrated timeline of mineral evolution.




Can such changes in pedagogy be "measured" and associated with improvements in learning outcomes? Certainly – as others have mentioned in this and other sessions there are several classroom observation protocols such as RTOP and TDOP. We use the COPUS because it is quick and easy to use, and easy to implement at a moment's notice. So far we have recorded observations in one class for at least 58 different courses.

Results for a **single** class are interesting and provide valuable feedback for instructors. Two examples are shown comparing a rather passive class to one with significant variation in "receiving" and "active working" during the 50 minutes.

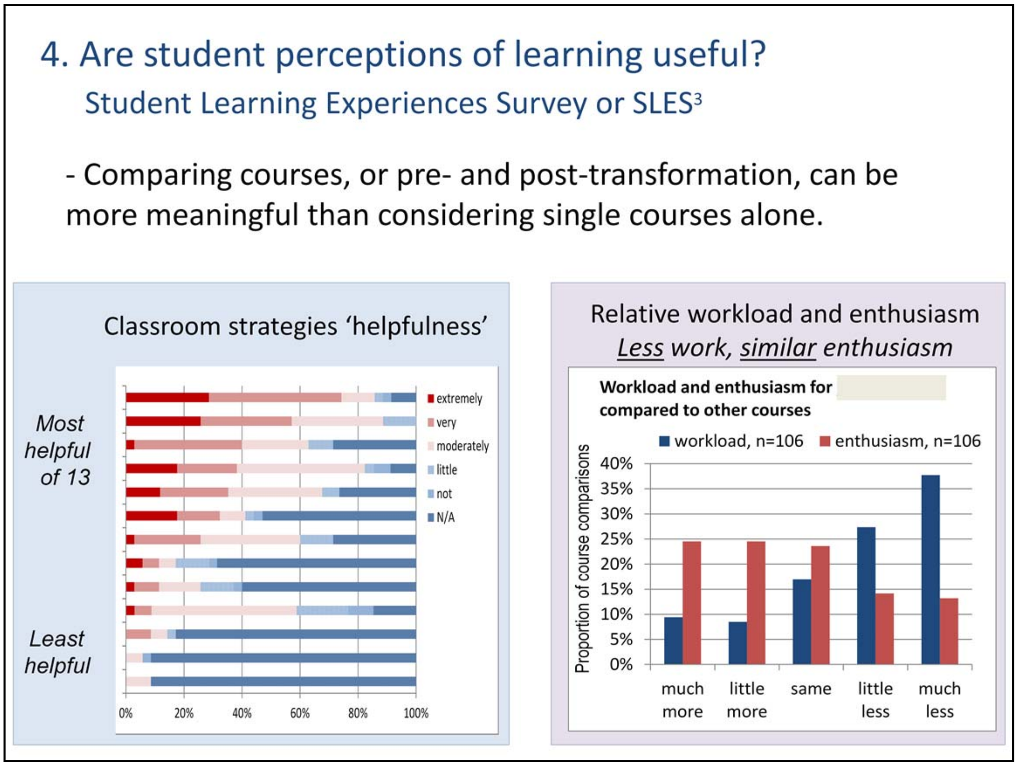
Associating observations with improvement or transformation efforts is best done by comparing across many courses. The bar graph shows proportions of courses that have active classes, grouped by type of transformation effort. Transformed courses are indeed more active than the less-extensively modified courses.




Yet another important perspective is that of students themselves. Surveys are commonly used but perceptions can be important if questions are carefully designed because they tell us something about how motivated students are in the course.

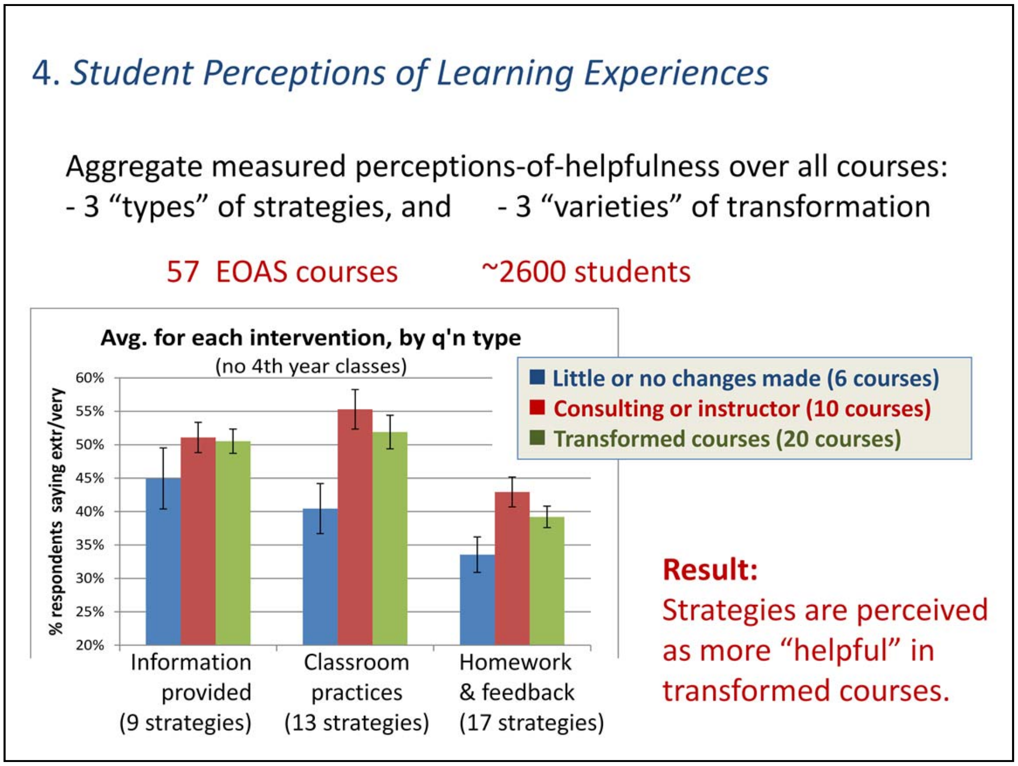
The left graph shows that 7 of 13 classroom teaching strategies are considered extremely or very helpful by students.

Also – student perceptions of workload and enthusiasm for this course – RELATIVE to each other course they are taking – are informative. Here, students feel they work much harder in this course AND they are much more enthusiastic about it.




In this “un-transformed” course, classroom strategies are less highly “endorsed”, students work less, and there is more variable enthusiasm.

It should be clear that comparing two courses can be more informative than considering self-reported perceptions data only in one setting.

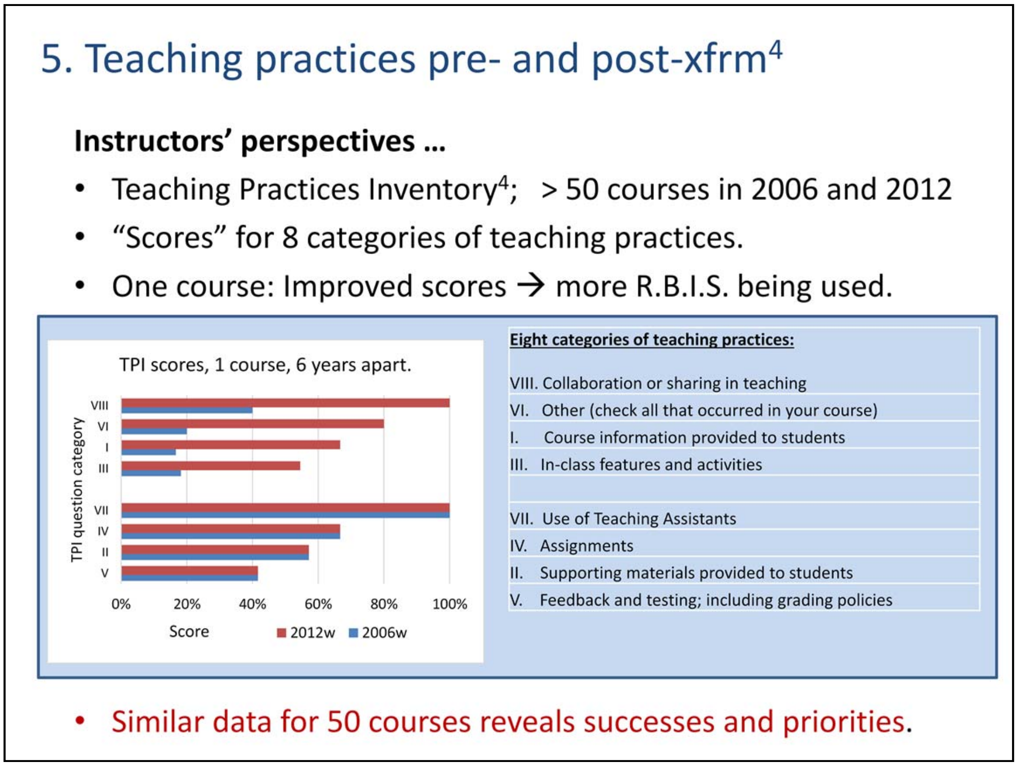



Measuring perceptions in one course is common; it's easy, inexpensive, and can help confirm or refute instructor's own assumptions.

However, comparing perceptions across a whole department yields more convincing results about whether strategies in transformed courses are better appreciated than those in un-affected courses.

This figure shows that students consider "information delivery", "classroom", and "homework/feedback" strategies are more helpful in transformed courses than non-transformed courses.

Courses transformed by consulting or instructor appear slightly more highly endorsed, however these data are cumulative over 7 years and this presentation does not show that many instructors who were involved in transformation projects found they could not teach "the old way" in their other courses, so they almost "had" to improve their other courses. It would be interesting to plot a time line of course transformations, instructor participation, and improvements made using individual and consulting models. Perhaps for another talk ☺.




Finally (for this talk), teachers' perceptions are also worth comparing. We used the "Teaching Practices Inventory", a published instrument that essentially asks instructors to check boxes indicating which Research based instructional practices they use, or how much they use them. Like other data sets, this has yielded the most interesting results when used widely across a department.

Here we show results from one course comparing teaching practices before and after transformation. The result shows increased use of Research Based Instructional Strategies – another indicator that student outcomes are likely better now than before.

## Other Approaches to Impact Assessment

**Done but not presented here**

- Compare changes in program objectives & course learning goals;
- Concept inventory tests (several but used to varying degrees);
- Assess changing metacognitive and study skills;
- 3rd party interviews with faculty (Bay view Alliance);
- Graduating exit surveys since 2009;
- Controlled studies (2-3 done; "expensive").

**Rarely done in our case**

- Pre-post student interviews & focus groups;
- Compare student entry and exit surveys;
- Others ???

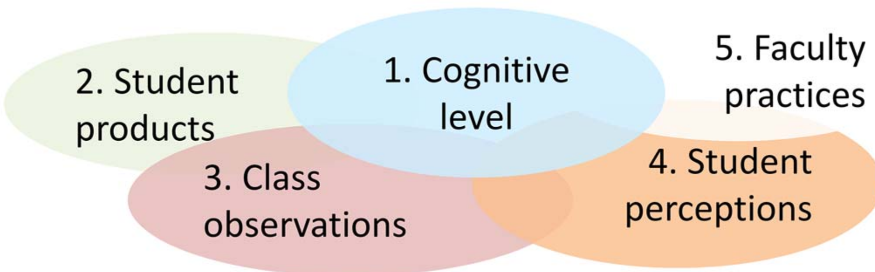
There are many quantitative and qualitative measures that contribute to a complete picture either of actual changes in learning outcomes or at least in an enhanced likelihood of improved outcomes because teaching and learning practices are aligned with Research Based Instructional Evidence.

For example, we could discuss pre- and post-transformation program & course objectives and learning goals. In fact, quite often, learning goals didn't even exist in the past so just their presence ends up being an improvement!

## Take home messages:

Individual evaluation perspectives are OK … but …

1) Multiple “dimensions” yield stronger evidence of impacts on learning outcomes of geoscience course transformations.
2) Data from many courses enables comparisons that are more informative than single settings alone.




But I’ve talked enough. Thank you for your interest and attention, and it would be great to address any questions.

Dep't Earth, Ocean & Atmospheric Sciences

# Assessing Improvements of Learning Outcomes in Transformed Geoscience Classes

~

**THANKS TO ORGANIZERS AND CONVENERS OF THIS SESSION**

Francis Jones*, James Scoates, Stuart Sutherland,
Randal Mindell, Brett Gilley

EOS-SEI

CWSEI

# Sources

1. Derived from Casagrand , J. and K. Semsar, *"A new tool for evaluating the cognitive difficulty of assessments"*, under review. Dept. of Integrative Physiology, U. of Colorado, Boulder, CO.
2. *The Classroom Observation Protocol for Undergraduate STEM (COPUS): a New Instrument to Characterize University STEM Classroom Practices*, M. Smith, F. Jones, S. Gilbert, and C. Wieman, CBE-Life Sciences Education, Vol 12(4), pp. 618-627 (2013)
3. *Comparing Student, Instructor and Observer Data to Assess a 7-Year Department-wide Education Initiative*, F. Jones, S. Harris, B. Gilley, the International Improving University Teaching conference, Vancouver BC, July, 2014. http://eos.ubc.ca/research/cwsei/resources/research/FJones-poster-140720.pdf
4. *The Teaching Practices Inventory: A New Tool for Characterizing College and University Teaching in Mathematics and Science*, C. Wieman and S. Gilbert, CBE-Life Sciences Education, Vol 13(3), pp. 552–569 (2014)
5. See also http://eos.ubc.ca/research/cwsei/ & http://cwsei.ubc.ca/