2007 National AAPT Meeting – Summer
Notes from July 30-August 2
Greensboro, NC

The following notes were taken by Dr. Van Domelen during the regular sessions of the 135th National Meeting of the American Association of Physics Teachers.

Contents | Page

Table of Contents 1
Glossary of Abbreviations 2
Personal Note 3
A Note On Posters 3

Monday, July 30
- AG: PER – Instructional Reform 4
- BA: Klopsteg Medal Ceremony – Neil deGrasse Tyson 9
- BE: PER – Observing and Modeling Cognition (partial) 10
- BN: PER – Surveys and Other Assessments 13
- BW: PER – Student Understanding of Topics in Physics, Teaching Assistants (partial) 19

Tuesday, July 31
- CB: PER – Conceptual Understanding: Models and Measurement I (inv.) 24
- Labs Committee Meeting 29
- DF: PER – Conceptual Understanding: Models and Measurement II (inv.) 31

Wednesday, August 1
- ED: Introductory College Physics Textbooks: Current Role and Future Possibilities (Panel) 36
- FC: PER Faculty Crackerbarrel 43
- FG: Laboratories 45
- PERC Bridging Session (guest notes) 49
- CP: PERC Poster Session (minor notes) 52

Thursday, August 2
- WS-B: PERC Workshop B: Physics Learning in the Context of Scaffolded Diagnostic Tests 54
- IT: PERC Invited Talks 56
- LUNCH: PERC Lunch Talk 62
- TP-B: PERC Targeted Poster Session B: A Conversation About Models, Modeling, Representations and Cognitive Science 63

Abbreviations

There are a few concepts that came up frequently enough in the talks I attended that I refer to them solely by abbreviation, making up my own where necessary. (Not all will appear this time; I just pasted in the 2004 glossary and added to it as necessary.)

ADT – Astronomy Diagnostic Test

BCI – Biology Concept Inventory

BEMA – Brief Electricity and Magnetism Assessment

CLASS – Colorado Learning Attitudes about Science Survey

"Clicker" – Not an abbreviation grammatically, but it covers a LOT of other terms, such as PRS (see below), and it seems to be the semi-generic term of common usage now. It's also apparently a word I can't consistently write correctly when I'm taking notes at high speed, which occasionally causes me problems. ☺

CSEM – Conceptual Survey of Electricity and Magnetism

EBAPS – Epistemological Beliefs Assessment for Physical Science.

ECCE – Electronic Circuits Concept Evaluation, developed by Thornton and Sokoloff.

ECR – Elicit, Confront, Resolve. The UWash method.

FCI – Force Concepts Inventory, a common mechanics test.

FMCE – Force and Motion Concept Exam, a common mechanics test. I may sometimes type it as FCME.

GG – Gender Gap, a disparity between male and female results, usually to the detriment of the women.

IE – Interactive Engagement, one of many strategies that get students more involved in their own learning.

ISLE – Investigative Science Learning Environment, developed at Rutgers.

MBL – Microcomputer Based Labs

MPEX – Maryland Physics Expectations Survey, a test of student expectations about physics and physics courses. MPEX2 is a recent revision.

N1L, N2L, N3L – Newton's First Law, Second, Third.

PBI – Physics By Inquiry (or {PbI sometimes), frequently used conceptual physics course using a workshop setting, developed at University of Washington.

PI – Peer Instruction

PRS – Personal Response System, a means of letting students answer multiple choice questions in class using a remote control. See Clickers above.

SPS – Society of Physics Students

TIMSS – Originally "Third International Math and Science Survey", but they changed what the acronym means in order to keep using all the logo stuff as they move on. ☺

TIPER – Tasks Inspired by Physics Education Research (subsets include eTIPERs and mTIPERs for electrostatic and magnetic activities respectively)

ZPD – Zone of Proximal Development, the space between "too hard to learn even with help" and "easy enough to learn without help".

I will sometimes add personal comments in [brackets]. I will try to withhold value judgments most of the time...if you feel the need to argue a point about a paper, please contact the author, not me (email addresses provided). However, if you are the author and feel I have missed the point of your paper, let me know and I will attempt to remedy the situation.

Personal Note

Most people who know me did get a chance to talk about my "new look", but in case you were one of the few who didn't, here's a short explanation.

In October 2006, I was diagnosed with Type 2 Diabetes (which was pretty far along, and part of the reason I was so tired and out of sorts at the Syracuse meeting, on top of all the other problems there!). Over the next several months I changed my diet and exercise habits and went on various medications, dropping a significant amount of weight and getting my condition under control. In July, I had gotten things under control enough to stop taking insulin and start cutting back on my other prescriptions, and had lost a significant (but not dangerous) amount of weight. As a result, by the time of the Greensboro meeting, I was under 200 lbs for the first time since, oh, 1984.

And that is why I looked different. ☺ On the other hand, I now have a lot less padding in general, which made sitting for some of the back-to-back sessions a little painful, and I left early at least once as a result.

A Note On Posters

This doesn't really go with any session that I took notes at, but is based on discussions during the Tuesday midday poster session with Pamela Gay of Southern Illinois University at Edwardsville.

As we shift to a larger emphasis on posters, it might be worthwhile to consider what style we present. A lot of posters, for instance, are simply the PowerPoint slides from someone's contributed talk, tacked up on the board. Others are more elaborate, complementing talks or taking the place of them. Pamela comes from the Astronomy field, where posters are used in a rather different way, and she had a number of criticisms to offer in terms of both content and form (pgay@siue.edu if you want to ask her advice specifically, check her portfolio at starwalkerdesigns.com).

Now, I'm not saying we should shift to the Astronomy way of doing posters, but I do think we should consider how we want to do posters, as a field. If we just keep going as we are, I'm sure some sort of standards will evolve eventually, but before that happens we're likely to have a lot of scattershot posters. It would probably be a good idea to come up with at least a direction to start nudging things, so that we have some conscious control over the evolution of AAPT poster style.

How to go about this? Well, a good start would be to take advantage of those of us who are dual-field and have more experience with poster sessions at other meetings, like DAMOP or AAS or whatnot. Personally, I've been to very few non-AAPT poster sessions, so I really don't have a strong view of how else these things can be done. But a lot of people who attend AAPT meetings are dual-field, and could help in establishing some sort of rough standard for future poster sessions.

Monday, July 30, 2007

AG: PER - Instructional Reform

AG01: Incorporating Philosophy of Science in Physics for Nonscience Majors – Todd Timberlake, Berry College, timberlake@berry.edu

Looked at an "Introduction to the Physical World" course, about 24 students per section. Activity-based course, with 23 hands-on activities, 10 labs, and various discussion group sessions. Student attitudes were measured using the EBAPS.

Prior to introducing philosophical elements, the active learning course showed minimal changes on the EBAPS. Content was conveyed successfully as shown in other measures, but the null hypothesis of "no student change in student attitudes" was not refuted.

Explicit instruction in the various philosophies of science (i.e. Newtonian worldview, various quantum interpretations, looking at science versus pseudoscience) was added to the course, along with reflective activities where the students tied their activities into these philosophies. There was also a new group project involving philosophy of science.

EBAPS scores improved somewhat overall (about 0.1 σ_e improvement, statistically significant), with most of the actual improvements happening in the axes concerning structure of knowledge and evolution of knowledge. The conclusion is that there is some benefit to explicit instruction on the philosophy of science.

AG02: Assessing Adaptations of Physics By Inquiry: Learning Outcomes – Leon Hsu, University of Minnesota, lhsu@umn.edu (work with Cummings and Taylor)

PBI has generally been used in its intended environment, small classes with trained teaching assistants in the room at the same time as the main instructors and a generally high student-to-staff ratio. This project seeks to adapt PBI to situations where this environment cannot be obtained for whatever reason, and also seeks to address state educational standards (since PBI courses are generally aimed at pre-service teachers).

Usually, according to instruments like the EBAPS and MPEX, students come out of introductory courses with worse attitudes and beliefs, even in courses with high conceptual gains. Using the CLASS instrument, typical shifts are insignificant or even negative.

Adapted PBI courses in situations where the usual environment didn't hold (three different cases, each "violating" the precepts in one or more ways) got large positive shifts (about 5-6 σ_e) in all trial institutions. However, the results look erratic over time. Further work will try to figure out which parts of PBI promote these shifts.

AG03: Assessing Adaptations of Physics By Inquiry: Learning Outcomes – Karen Cummings, Southern Connecticut State University, cummingsk2@southernct.edu

Continuation of the work presented in AG02. This looked at the use of the circuits components of PBI volume 2 combined with the Science Technology for Children (STC) kits, and evaluated with the ECCE. The course had small enrollment, and 64 total students were aggregated over several terms.

Pretest scores were low (under 30%) and posttest scores were very high (over 80%) on simple circuit items. Current relating to bulb brightness in PBI was shown to be a very effective modeling strategy. Students generally did very well on concept assessment, although the course only really covered current with very little coverage of the conceptually more difficult voltage.

AG04: How To Teach With Analogy – A Research-Based Approach – Noah Podolefsky, University of Colorado at Boulder, noah.podolefsky@colorado.edu (work done with Noah Finklestein)

This study looked at the difference between using straight analogies and using conceptual blending.

An analogy is a mapping from an unfamiliar base to a familiar target, preserving structure as well as possible. However, this mapping can carry along detritus that hinders learning.

Conceptual blending takes elements of two representations and carries them into a new one, rather than a one to one mapping.

A standard mapping might be to look at the progression to an EM wave. Start with waves in a string, analogize to sound waves, and then analogize further to EM waves, getting more abstract in each step but trying to preserve the structure learned in more concrete situations.

Two tutorials were created. One used only a sine wave representation for a single analogical mapping, while another one blended both concrete and abstract representations of a wave in three dimensions. The blending tutorial was much more successful in posttesting (75% versus 25%).

In any case, analogical scaffolding is a useful tool.

AG05: Instructional Reform and the Importance of Promoting Coherence in Physics Knowledge – Mel Sabella, Chicago State University, msabella@csu.edu

Specific case study of taking numerous educational reforms (including tutorials, web homework, interactive PowerPoint slides, TIPERs and clicker tasks) and merging them in a small enrollment class environment. Students are all in one room for a three hour stretch, which is broken up into lecture, lab and other activities using a workbook format. The goal is to create a nice portable and linear workbook for export to other instructors, digesting a great deal of PER results for ready use.

An example was presented of the coherent structure used to teach kinematics, a situation where acceleration is positive but velocity is negative. Students lacked a robust understanding of +/-, disliked drawing graphs in the lower right quadrant and going negative in general was bad for them.

So far, the educational issue has been well-identified, but they don't yet have a workbook section that fixes it. [I think that was the point, the second half of the talk seemed a bit disconnected from the first.]

Students in general lack coherence between representations.

AG06: Not All Interactive Engagement Is The Same: Variation in Faculty Use of Peer Instruction – Chandra Turpen, University of Colorado at Boulder, Chandra.Turpen@colorado.edu

Motivation: it doesn't matter how good a curriculum is if we don't know how it actually gets used, or even if it can be used as intended. Therefore, this study looked at difference in implementation of Peer Instruction in various courses. Faculty interviews, student surveys, ethnographic observations, daily clicker records and other measures [that were covered too quickly for me to copy down] were used to analyze use of Peer Instruction.

A large-enrollment introductory physics course was studied in this case, with six sections (3 taught by PER-versed instructors, 2 by novice users of Peer Instruction, and 1 by an experienced but non-PER faculty member). There was no consistent link of practice/usage across PER/non-PER, although with only 6 instructors it's hard to draw much of a conclusion there.

Some instructors would stay "on stage" after posing a clicker question, others would get off the stage. Those who left the stage were more likely to talk to individual students after posing the question.

Novices were more likely to adopt an all-or-nothing approach to methods, either using them every time or never. Experienced instructors seemed more likely to vary their approach.

AG07: Towards a Set of Research-Based Best Practices for Clicker Use – Christopher Keller, University of Colorado at Boulder, Christopher.keller@colorado.edu

Many disciplines have researched clicker use, resulting in over thirty names for clickers that the author could find. The author attempted to perform a broad study of clicker use: how are they actually used by faculty across disciplines, and what do students think and do?

From the study, an attempt is being made to generate broad guidelines for clicker use. They seem to be used far more frequently in large lecture classes than in smaller sections [which makes sense, but just because something makes sense doesn’t mean it's true, especially in educational practices].

Students were 56% favorable and 20% unfavorable to clickers across all departments at UCB that used them. This positive attitude correlates with effective use of clickers, such as the type of activity, level of discussion, etc. The more appropriately clickers were used, the more popular they were.

Therefore, one broad guideline we should follow in using clickers is to always encourage and promote discussion during clicker use.

[Aside: due to an unpleasant day of travel and late arrival, I slept poorly, and started to nod off at this point. Apologies to any speakers after this point if I missed your point by a wide margin.]

AG08: Transfer of Scientific Abilities: PER Research Design Project – Eugenia Etkina, Rutgers University, Etkina@rci.rutgers.edu

This is the first in a trilogy of related talks.

Looking at workplace needs, it was seen that employees needs to understand and participate in the practices of science (rather than the results of science). Therefore, it's important to focus education on the design and development end of things, not just the content. The ISLE curriculum attempts to model this iterative process.

The "Scientific Abilities" project provides students with a rubric to help them self-evaluate on the following criteria for every project: representation of physical processes, design of experiments, collection and analysis of data, testing relationships and finding explanations, modification of the explanations and relationships, evaluation of procedures and outcomes, and communication of results.

Once students are taught to develop these abilities in one context, do they transfer to another?

Both the treatment and control groups were in ISLE cycles, but the treatment group performed design-focused labs while the control used PER-developed prescriptive labs. The treatment group was given scaffolding help and used the self-assessment rubric on their designs, plus had different homework assignments (for instance, the treatment group analyzed historical passages to look at how scientists designed their experiments). To keep the time-on-task similar, the control group was given additional problems to work on.

At the end of the semester, both groups performed the same two labs. First, a design-based physics lab (in part to see if content knowledge alone from the control group was enough to design a good experiment), and then a biology lab. In the biology lab, they were given enough information to design the experiment, making this a truer test of the transferability of design skills. In both cases, the design labs weren't completely open-ended and they had some guidance, but they were far from prescriptive.

Observers watched student activity during the labs and classified it according to a previously designed scheme. The treatment group spent more time on writing and sense-making, and about the same amount of time in all other categories. The control group spent less total time on task, and the treatment group frequently ran out of time before finishing the lab. The treatment groups never got the "right answer".

On a lab practical exam, the treatment group outperformed the control (effect size 0.3). They were slower to pick things up, but did better on the final exams and during the second semester of the course than the control group.

AG09: Design and Nondesign Labs: Does Transfer Occur? – Anna Karelina, Rutgers University, anna.karelina@gmail.com, http://www.islephysics.net

Continued from AG08.

In week 12, both groups were asked to design a lab on aerodynamic drag, a topic not covered in lecture but still within the physics "frame". Treatment and control got slightly different versions of the lab to avoid cross-group direct copying of lab reports.

The control group spent much less time on sensemaking activities, and more time asking the instructor for help. The treatment group spends more total time on task, and had better lab reports. The treatment group was more likely to include multiple representations, explanations, justifications and explicit assumptions in their reports.

Looking at the elements of the Scientific Abilities rubric presented in AG08, the treatment group was better able to identify assumptions, evaluate and validate the effect of assumptions, evaluate uncertainty and evaluate results by independent methods. They also had better-quality free body diagrams and more consistency between representations (i.e. what they drew was more likely to match their equations) than the control group.

Conclusion: if students are explicitly taught to consciously plan, monitor, evaluate and reflect on their actions, these good habits transfer to multiple contexts.

AG10: From Physics to Biology: Helping Students Attain All-Terrain Knowledge – Maria Ruibal-Villasenor, Rutgers University, mruibal@eden.rutgers.edu

Continued from AG09.

In week 13, students capped the lab course with a biology experiment to see how their general skills improved. The goal was to determine the transpiration rate of plants using stem cuttings. Resources were provided and internet access was allowed so that they could look up whatever they needed. About 200 students were studied, 5 of which did not give permission to have their data used. There was IRB-approved deception involved [If I recall correctly, the IRB suggested it, but I may be mixing up recollections here].

As with the drag force lab, the treatment group spent more time sensemaking and the control group more time asking the instructor questions, otherwise both groups spent about the same amount of time on other types of activity.

The treatment group outperformed control in the following areas, all with very high chi-squared scores [the lowest I caught was 28, and some were around 100]: Identifying and evaluating assumptions, dealing with uncertainties, evaluating results via an independent method, recording, representing and analyzing data, and communication. The treatment group also was judged to be having to think more.

AG11: How Upper-Division Physics Students Respond to Studio Laboratory Activity – Fran Mateycik, Kansas State University, mateyf@phys.ksu.edu

Recently, the upper-division Optics course was reformed along studio lines. The students were a mix of upper division physics majors and physics graduate students. The course involved lab activities of a more exploratory nature, rather than confirmatory.

In two teaching interviews performed in the summer of 2006, the studio activities dealing with diffraction were examined. The subjects were a mix of Kansas State University students and various visiting REU students, and backgrounds were widely mixed. A sheet of butcher paper served as a "lab notebook", and the instructor/interviewer made minimal comments aside from cueing.

Students were required to defend their explanations, and seemed to understand the writeup of the lab and its goals. However, some had questions about the experimental equipment, and they were generally unfamiliar with the "messing around" style of lab that studios try to promote. Instead, they tried to use "plug and chug" approaches, relying too much on equations and having difficulty with the conceptual physics.

There were no trends found to be tied to student backgrounds, although performance was uniformly low, which could have masked any differences. Despite being upper level students, they relied on equation-centered approaches rather than attempting to apply any conceptual understanding. Student expectations of what lab will be like seems to constrain exploration, we need to open their minds up.

AG12: Evaluation of the Physics and Astronomy New Faculty Workshop – Charles Henderson, Western Michigan University, Charles.Henderson@wmich.edu

[NFW will henceforth be used for "New Faculty Workshop.]

The NFW is a four day conference/workshop that tries to reach a large fraction of new Physics and Astronomy faculty. About 70 people attend each year, but the numbers are rising. It's short and transmission-oriented, so by what we know of education it shouldn't really work.

Yet it does. NFW increases knowledge and changes practices, according to a survey of about 500 past participants. About 200 now use Peer Instruction, most of which use some aspects of it one or more times in every meeting of their class, and about a third use all aspects. It's an impure form, but an improvement over nothing at all.

NFW is a gateway experience, it puts the instructors on a useful path which they then follow on their own later. It's a seed, not a full-grown tree. Sponsorship and support by organizations like AAPT, AAS and APS may help explain why this is more successful than other gateway experiences. It's also a broader exposure, rather than the system-specific sales pitch that many gateways are.

Conclusion: it may not do much in those four days directly, but NFW is clearly doing good things.

BA: Klopsteg Medal Ceremony

BA01: Adventures In Science Illiteracy – Neil deGrasse Tyson

[Note: this talk was various anecdotes, so there really wasn't much of an overarching structure, and my scattered notes are reasonably reflective of the way the talk went. I left before Q&A since the session ran very long and I had hoped to make it to most of the talks in session BE.]

Tyson is People Magazine's "Sexiest Astrophysicist Alive" and frequent guest on talk shows, including the Daily Show. He hosts Nova Science Now and various other outreach efforts. This makes him a natural recipient of the Klopsteg, which is for promoting science in the public eye.

His recent book is titled Death By Black Hole (and Other Cosmic Quandaries).

Old astronomy trick: the fist at arm's length subtends about 10 degrees each direction unless you've got a very oddly-proportioned body.

"Not paying attention to the world around you" is one type of science illiteracy, and is very common.

A few years ago, he was teaching a seminar on standards of evidence and the unreliability of eyewitness testimony and got summoned for jury duty. The fact he was teaching about this got him booted from the jury during voir dire.

The Scientific Method – do whatever it takes to avoid fooling yourself into picking based on bias. Everything is fair game, and has to stay fair game.

If you don't watch TV, you're going to have trouble communicating with the public. Television shapes the common language of analogy and experience, and you have to meet them halfway. Maintaining an aloof "TV is trash" attitude can keep you from communicating with people, so suck it up and watch some boob tube. "If you can't meet them halfway, how dare you call yourself an **educator** rather than a lecturer?"

In the movie Titanic, there's one scene where the night sky is not only wrong for the latitude and time of year, it's also only half the sky mirror-flipped to fill! This seemed odd, given how obsessive Cameron was about researching all the other details in the movie. Tyson had a few occasions to bring this to Cameron's attention, and while it was blamed on an error in post-production (out of Cameron's hands), Cameron once replied, "Imagine how much more it would have grossed if I'd gotten the sky correct!" However, the 10th Anniversary release has some restored scenes that had never been through post-production, so they will get the correct sky added in.

Tyson on the "Pluto Controversy": "Get over it!" Naming doesn't impact the named, especially when the named is an inanimate object. Science is not dogmatism, we shouldn't get so het up about this sort of thing.

Realizing that the Intelligent Design debate seemed to be omitting an important point, Tyson stepped into the issue, and found it not unlike stepping into a cow pasture after heavy rains. But the point he wanted to make was that ID invokes the God of the Gaps, something that is an automatic endpoint. He used an example of how our knowledge of planetary motion progressed from Ptolemy's Almagest (waxing poetic in awe of the universe that Zeus had set in motion) through to Newton inventing calculus on a dare, and how anywhere along the line just accepting "here's where God/gods stepped in" would have stopped that next discovery. [Aside: actually, argument against ID does lean heavily on the God of Gaps thing, but I suppose that side of things doesn't make it into the popular press much.]

[At this point, session BE was about to start, but since I'd been ushered into the middle of a row I couldn't get out until Q&A started. My notes are sparser because I kept looking around for a way out.]

A philosophical change happened in the Islamic world around 1100 CE that poisoned the well for Islamic scientific advances. [In fact, recent issues of Skeptic and Skeptical Inquirer have a lot to say about the state of science in Islamic countries. Engineering, primarily reverse-engineering of advances made in other countries, is acceptable, but pure scientific research is heavily blocked by religious considerations.] We seem to be headed towards a similar poisoning of the well here in America.

Innumeracy tends to lead to distorted values, such as a news item about Michael Jackson's latest court case getting more time on the news than when thousands die in an earthquake.

The universe itself has become a pitchman for advertisers.

BE: PER - Observing and Modeling Cognition

[I got to this late, and had to sit on the floor. It turns out that this hurt a LOT, but I couldn't really take notes while standing either. So I stuck around for a few talks and then left. Apologies to those whose talks I missed.]

BE04: Beyond Confusion: Alternative Accounts of Students' Failure to Differentiate – Brian Frank, University of Maryland, bwfrank@mail.umd.edu

[Differentiate meant in the sense of telling apart concepts and quantities, not in the dx/dt sense.]

The classic view is that students have a static, undifferentiated blob of concepts that get separated during instruction. For instance, in one example presented, a student conflated volume and mass in one context, but was able to separate them out in another. An alternative dynamic perspective was offered that is context-dependent.

[I don't recall if the talk was this thin, or if I was just having trouble taking notes while sitting on the floor...I don't have any marginal comments relating to the talk going too quickly to follow.]

BE05: Priming Epistemological Framing: "Answermaking" and "Sensemaking" in Introductory Physics Courses – Paul Hutchison, University of Maryland, hooch@umd.edu (Renee Guertzen presenting)

Students are able to shift into a mode of thinking that bypasses common sense, in order to get at what they think is the Right Answer. This behavior is referred to as "answermaking" as opposed to the commonly used term "sensemaking". They get into a frame of mind where they simply decide that nothing has to make any sense, and just try to play by what they perceive to be the rules of the game. For instance, when seeing something demonstrated and then later asked a question that's similar but not identical, they may pick an answer that bears a surface similarity to what they saw in class, because they don't understand how it might be different (and may not care).

It's necessary to prime students for sensemaking behavior, and breaking out of the expected "textbook" frame of mind can help. Sensemaking questions are often best taken from topics that are covered lightly or not at all in class, so that there's no perceived "book answer" to fall back on.

Students were presented with a two-ball problem, in which two balls start with the same starting speed, and height, but different directions (one is moving down, the other horizontally), and asked which would hit the ground first. This problem followed either a textbook problem that would prime for incorrect answermaking (i.e. the classic problem where one ball is dropped and the other shot horizontally), or a problem designed to prime sensemaking behavior.

The conclusion was that priming for sensemaking was possible, but not trivial. It didn't always happen in the study, and instruction details may matter. Explicitly encouraging sensemaking did seem to make priming for it work better, though.

BE06: Explicit Reflection in Introductory Physics – Michael Scott, University of Illinois at Urbana-Champaign, mscott1@uiuc.edu

Students were asked to reflect on matters, looking back on their studies to find meaning and commonalities while integrating knowledge. In the treatment group, quizzes were altered by the addition of explicitly reflective questions in place of standard quantitative items.

"Will this pair of problems be solved similarly?" is an example of the sort of surface versus structure reflective items that were added.

Both treatment and control groups scored too well on pretests of the FCI and sorting tasks (a la M. Chi) for there to be much room for differential improvement. However, looking at the final exam scores, the treatment group outperformed the control group by 1.7 standard deviations (2.4σ on the conceptual items, 1.4σ on the quantitative ones).

BE07: Effect of Viewing Order on Students' Judgments of Realistic Motion – Adam Feil, University of Illinois at Urbana-Champaign, adamfeil@uiuc.edu

This study involved the two-track animations from Mestre's work, in which students see animations of balls moving on one or two tracks, one of which has a V-shaped dip in the middle. The "FST" state, or "Fast-Slow Tie" in which both balls reach the end of the track at the same time (which is not the actual physical result) is favored by those with prior instruction but not by those without. Also, the unphysical motion of the ball on the V-dipped track is not picked as realistic in one-ball simulations, only when it's matched up against motion on the non-dipped track.

An eye-tracking study presented at a previous meeting showed differences in eye motion between non-physics students and physics students. The eye tends to overshoot the motion of the ball in the unphysical FST-style motion when only the one ball is shown. However, oddly, in the eye-tracking study the FST situation is also no longer favored in two-ball animations.

Non-dipped track animations shown earlier to the participants seem to have primed them, and the order of showing the animations is now recognized to have been important. The eye-tracking study did drop some animations used in the original study, which could affect priming even if the specific dropped animations didn't seem to be interesting study subjects on their own. Depending on order, either FST is favored, or the even more non-physical situation of the dipped-track ball losing is favored, and they're not yet sure why. It's possible that inconsistencies in numbering the options may be involved.

The current plan is to modify the eye-tracking experiment to recapture the presentation order of the original study, and see if more information can be teased out.

BE08: Studying Fine Structures of Mixed Mental Model States – Steve Stonebraker, the Ohio State University, sstoneb@mps.ohio-state.edu

When repeatedly probed on a single concept, students often give inconsistent responses, as if each question triggers a different model, a mixed state as per Lei Bao's Model Analysis framework.

Mixed models can be explicit (the student knows they're mixed) or implicit (student is unaware of any contradictions). The hope is to develop a diagnostic test to separate these. Students were allowed to pick a second answer to a multiple choice test when they were uncertain about their main answer, making explicit contradictions a little more likely. An updated FCME and BEMA were used for this, with students assigning confidence ratings to their responses.

On the FMCE pretest, higher confidence correlated with lower rates of multiple answers ($r = -0.47$), but no effect was found on the posttest. Overall, however, no correlation was found between confidence and multiple answers, although few students exercised the option of picking a second response in any case on the FCME.

On the BEMA, there were conflicting results. On the pretest, there was no relation between the overall score and a tendency to give multiple answers (which was, as in the FCME, rare). However, in the posttest, those with low to medium confidence were more likely to select multiple answers.

[Left the session at this point.]

BN: PER - Surveys and Other Assessments

BN01: Does Feedback Improve Scientific Writing Quality Independent of Grade Motivation? – Dedra Demaree, College of the Holy Cross, demaree.2@osu.edu

This work was a collaboration between the Ohio State University and Rochester Polytechnic Institute, across Physics and English Departments.

Premise: When giving feedback on writing as formative assessment, there should be a measurable effect on performance of this feedback. However, the literature is silent on the matter,

"World of Energy" (Physics 104 at OSU, a non-majors conceptual physics course) was used as a testbed in the Winter 2006 quarter. Students were shown videos on various topics and had to write summaries of the videos. Scores for these summaries were purely participation points, a bad summary got the same points as a good summary.

There were three sections. Control 1 got no feedback at all on the summaries. Control 2 got "warm fuzzies" feedback from the instructor (positive and encouraging comments regardless of actual quality of work). The treatment group got minimal feedback from the instructor and additional feedback from the researcher. This additional feedback was in the form of commentary aimed at improving writing style in general, rather than content-based.

All essays were evaluated using an ISLE-style rubric for purposes of the study, but all received the same course points for participation. Also, participation points for students in all three groups were tracked over the course of the quarter. Those in the treatment group improved on participation points over the course of the quarter, while those in the control groups actually got worse, turning in fewer assignments. The differences were small, but statistically significant. Control 2 was not significantly different from Control 1 in any measure. Surveys of attitude showed a noticeable positive difference between treatment and controls.

Conclusion: the title question is answered in the affirmative. Even if the feedback has no direct impact on grades, it still improves student performance and attitudes in measurable ways.

BN02: Scientific Reasoning and Metacognition: Results of an Exam Reflection Survey – Craig Wiegert, University of Georgia, wiegert@physasst.uga.edu

Cancelled.

BN03: Using Clickers in Upper-division Physics Courses: What Do Students Think? – Katherine Perkins, University of Colorado, Katherine.Perkins@colorado.edu

The use of clickers has spread to all introductory courses at the University of Colorado, and is starting to spread into the upper division courses as well, including those taught by non-PER faculty. One graduate physics course is also using them. Conceptual tests and questions are used with clickers, rather than quantitative ones.

Five upper division or graduate courses were surveyed, with good response rates (2/3 or more in all courses). Four of the five courses reported the encouragement of peer discussion.

Responses on a Likert scale showed that when compared to standard lecturing, clicker use garnered 80% favorable ratings. 76% expressed a desire to use clickers in other upper-level courses. Other responses recommended 2-5 clicker items per lecture (distribution skewed towards the low end), clicker items interspersed with lecture rather than all in a chunk, and more encouragement of peer discussion (both as part of clicker use and in general). The items found most useful, at 91% "useful" rating, were the challenging conceptual questions. Recall-based or plug-and-chug items were not favored (less than 40% finding any of these useful, with plug-and-chug coming in at 18%).

Students in upper division like the clickers, and want to use them more, but with conceptual items.

BN04: Peer Instruction: Clickers vs. Flashcards, Is There A Difference? – Nathaniel Lasry, John Abbott College (Montreal), lasry@johnabbott.qc.ca

In class, a brief lecture would be followed by a clicker test to determine where the class should go based on the success rate on the items.

Questions asked: is PI effective in a small college, as opposed to a university? How does it help those with less background knowledge? Does it help with retention (as in, not dropping out of the class)?

In the study, students were effectively randomly assigned by registrar to one of three sections of a course. In the control group, traditional methods were used. In treatment 1, the speaker used clicker-based PI. In treatment 2, the speaker used flashcard-based PI. All three sections were evaluated using the FCI and a centrally-graded common final exam. Student retention was also examined in all sections.

<g> on the FCI for both PI groups was about 0.5, with no difference between treatment groups. <g> for the control group was about 0.33. p=0.008 for the statistics. There was no significant difference in exam scores.

In the control group, there was little difference in scores between the high group [I think it was a quartile] and the low group. Both treatment groups had much larger discrimination between high and low performers.

The treatment groups had a 4.8% dropout rate, the control group had 20.5% dropping, consistent with Mazur's results.

There seemed to be no difference between using clickers and using flashcards in this study, but either PI method seemed to work well in the small college setting.

BN05: A New Clicker Methodology for Introductory Physics Lectures – Bill Reay, the Ohio State University, reay@mps.ohio-state.edu

Early work on clickers focused on the technology, which wasn't always reliable or easy to use. Fortunately, those hardware and software hurdles have now been pretty much jumped, so it's time to focus on the pedagogical methodology we use with the clickers.

Since learning is context-dependent, we should present numerous items that have the same deep structure but different surface features. The trial in question took place over three consecutive quarters (Fall 2005 through Spring 2006) in the calculus-based introductory physics sequence at tOSU. The CSEM was used for evaluation, and had previously seen gains of about 0.19 in this course.

The survey responses were generally positive when not graded. [I'm no longer sure what I meant by this line in my notes, sorry.]

The clicker sections averaged 1.3 items better on posttesting than control sections. The clickers helped female students more than they did male students, closing the gender gap. It's thought that this is because clicker use is something where improvement is correlated with class attendance (use the clickers more, get more benefit) and female students in this course generally have better attendance rates than male students. However, attendance in general was poor, introducing a lot of noise to the signal. Even pointing out to students that attendance was correlating positively with performance did not improve attendance.

BN06: Designs and Evaluations of Two Types of Clicker Question Sequences – Pengfei Li, the Ohio State University, li.427@osu.edu [Trivia note: yes, that means that this is the 427th person with the family name of Li to attend Ohio State since they instituted the @osu.edu email system. By contrast, I'm vandomelen.1@osu.edu, although I don't read that account anymore as it's almost entirely spam.]

This is a sequel to BN05.

Two sequences of clicker questions were used in the sequence described previously. Either "Easy-Hard-Hard" or "Rapid Fire". EHH used a simple question to get the ball rolling, then two harder ones to elicit misconceptions and ensure that a correct answer on the first hard question wasn't a fluke. RF had three items of roughly equal and medium difficulty. Each meeting of the class saw one or two sequences.

So, is one type of sequence better than the other?

Yes and no. Higher performing students do better with EHH than with RF, although they do well with both kinds. Lower performing students benefit from RF, but not from EHH.

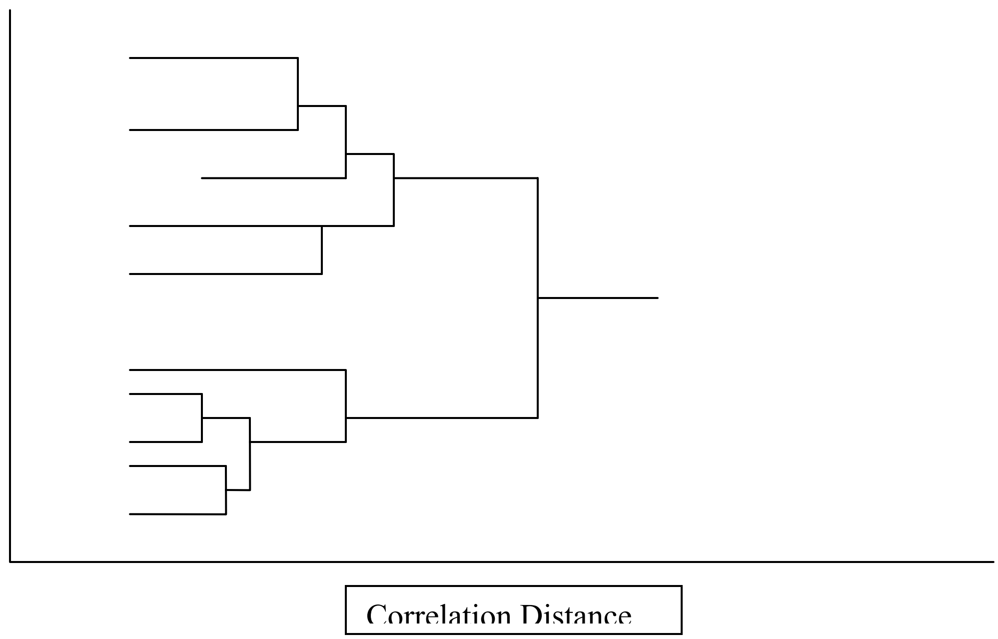
If limited to just one type, RF would probably be better, since it covers everyone. But since the higher performers shouldn't be completely neglected in favor of the lower performers, a ratio of 7:3 is suggested for RF:EHH.

BN07: Using Cluster Analysis to Group Written 2-D Kinematics Responses – R. Padraic Springuel, University of Maine, r.springuel@umit.maine.edu

[Okay, cluster analysis is pretty neat, but there really isn't time in an 8 minute talk to properly explain it. A game attempt was made, but I'm going to take a digression here and explain it as far as I understand it, rather than limiting to just what was covered in the talk. Keep in mind that this is itself based only on attending a few talks about it, so I may be a bit off.

Cluster Analysis is a way of taking a bunch of correlated items and diagramming the relationships between them in an attempt to clarify which ones are more closely related to each other, finding clusters like grapes on a bunch. It has similar goals to Factor Analysis, but different methodologies, assumptions, etc.

In the very crude drawing below, you see several items in a cluster analysis. Normally there would be a label for each point at the leftmost end, but Word Draw is a royal pain, so I'm forgoing that part. The horizontal axis is increasing correlational distance from left to right, so the far right end would be $r = 0$ and the far left end would be $r = 1$.



The point where two lines connect with a vertical branch is a correlation* between those two items, be they tests, questions, etc. The farther to the right this connection, the more weakly the items are correlated, and just follow the lines until you get a connection. So, for instance, the top item and the bottom item on my cluster graph aren't connected until you hit the farthest-to-the-right vertical line, so their correlation is low. And everything in the top cluster is more closely related to everything else in the top cluster than to anything in the bottom cluster. In essence, this is really just a representation for raw correlational data that makes it clearer what stuff is close together and what stuff isn't.

The utility of this diagram comes not only in a purely exploratory sense, but also when you look at the difference between the actual clustering and your a priori groupings. Say, for instance, out of the ten items in my graph I had four intended to test Newton's Third Law and six intended to test Newton's Second Law. But I don't have clusters of 4 and 6, which means that at least one of my N2L items clusters better with N3L than with other N2L...and it might be that both of my clusters are mixed in terms of N3L versus N2L! This would tell me that there might be some other element underlying my items that is stronger than N2L/N3L in determining how students see them. I can't remember if I thought this up or someone else did, but it's basically "looking for Romeo and Juliet". Try to find people in the wrong "family" in other words, a Montague among the Capulets or vice versa.

* Strictly speaking, it's a little more complicated than just an r score. If you were to graph student data on just two items, the "correlation" would be the average distance on the graph between students in each clump of scores. This is why you need specific software packages to perform this analysis...I simplify greatly here in order to get across what I think is the core idea of Cluster Analysis.]

In this study, students were given a path line of motion (x-y diagram) and asked to draw the acceleration and velocity vectors onto this diagram at various points. The actual materials come from a large University of Washington project aimed at calculus-based physics courses, and 104 algebra-based students were looked at in this particular case.

Student responses were sorted into descriptive categories via binary scoring (0 means the category doesn't describe the student's work, 1 means it does). After discounting items where student responses were uncategorizable, 313 distinct data points remained. Cluster analysis was performed on this data.

One cluster that was found contained responses pertaining to acceleration always being tangential (never having a centripetal component). This split, rather high in the tree (i.e. low correlation), into two distinct clusters representing different reasons for there being only tangential acceleration.

Cluster analysis is very good at teasing out differences in seemingly similar groups of students.

BN08: Revised Methods for Analyzing the Force and Motion Concept Evaluation – Trevor Smith, University of Maine, Trevor.I.Smith@umit.maine.edu

The FCME items are currently clustered by face-value topic, rather than by underlying factors. A new clustering scheme is suggested based not just on the content of the items, but also on the presentation of the item (i.e. cases of horizontal motion should be considered distinct from vertical), the use of natural language versus a graphical representation, and so forth. The diagnostic utility of the incorrect responses is also considered in this scheme. It results in a much larger number of chunks than the normal clustering.

BN09: Critical Thinking and the Learning of Force and Motion Concepts – Susan Ramlo, University of Akron, sramlo@uakron.edu

Linear models of many variables have never accounted for more than 30% of variance in the FMCE post-test scores, so it was decided to look for some other variable that might account for some of the missing variance. Rubrics of critical thinking were considered as this new variable, and a multiple linear regression was performed. These rubrics included the Holistic model and the SOLO (Structure of the Observer Learning Outcomes) taxonomy. The rubrics were applied to activity in the lab portion of the course.

After analysis, it did seem that critical thinking accounted for a significant chunk of variance in FMCE posttest scores. The complete model with SOLO rubric has an R^2 of 0.52, while taking SOLO out drops this to 0.22. The Holistic rubric also improved variance, but not as well as SOLO. SOLO also provides better feedback to the students.

In conclusion, critical thinking skills are a big part of success on the FMCE, and SOLO is a good way to measure them.

BN10: Constructivism and Science Anxiety I: Questionnaires and Interviews – Jeffry Mallow, Loyola University Chicago, jmallow@luc.edu

This was a cooperative effort with a Danish school, see BN11.

Do constructivist views correlate with science anxiety (SA)? Does this correlation vary with gender, age, or other factors?

A Science Anxiety questionnaire was developed, comparing subject anxiety on paired tasks. In each pair, the tasks were similar in structure, but one was overtly scientific in surface details while the other was not. Both general anxiety levels and SA-specific scores were evaluated. [This was your basic "15 minute talk in 8 minutes" situation, so I had trouble figuring out what was going on after this point, all notes should be taken with a grain of salt.]

Personal constructivism, social constructivism and educational constructivism were all considered.

Binned by nationality and gender, American women were the most anxious, then American men, Danish women, and Danish men. There were clear differences in the groups, suggesting that anxiety cannot be purely a genetic or biological effect. At least, not a simple one.

Mention was made of a British study that disproved the old saw that children are born scientists and have it drummed out of them.

BN11: Constructivism and Science Anxiety II: Study Results – Helge Kastrup, Copenhagen Day and Evening College of Education, Helge.Kastrup@skolekom.dk

[Significant problems getting PowerPoint to work, plus time issues again and a generally scattered talk, so same caveat as in BN10.]

Danish subjects in the study were from both Gymnasium (college-prep high school) and university, while the American subjects were Loyola students and in-service teachers.

For Danish men, negativity about science was strongly related to both SA and general anxiety. For Danish women, fears about the negative effects of scientific endeavors were moderately related to SA. There were no significant gender differences in the American results presented in this talk. Some subjects were interviewed in depth.

Intelligent Design views were more likely to be held by American subjects than Danish ones, agreeing with general studies on the matter. Denmark is facing a teacher training disaster, having revamped the national curriculum in such a way that discourages science educator training.

BN12: Influencing Students' Attitudes and Expectations in Introductory Physics – Jeff Marx, McDaniel College, jmarx@mcdaniel.edu

In a calculus-based introductory course, problem solving skill sections were added in 2004, in which students explicitly model problem solving techniques such as using multiple representations, stating assumptions, etc. About 20-25 percent of the scores in grading for the course are for correct process, so grading reinforces the problem solving sessions.

The Attitudes about Problem Solving Survey (APSS) was administered at the beginning, middle and end of the sequence (see the PERC 2003 proceedings for more about the APSS).

Three groups were examined. A treatment group at McDaniel was put through the PS sessions, while a control group there was not. Additionally, students at RPI were tested with the APSS while taking a similar course without PS sessions.

Between beginning and midpoint, both McDaniel groups improved somewhat in attitude, with little difference between their APSS scores, while the RPI group's scores declined. Between middle and end of the course, the treatment group continued to improve in attitude, the control group declined, and the RPI group continued its decline. However, the control group's attitudes still improved between pretest and final administration, they seem to be a happy lot in general at McDaniel.

BW: PER – Student Understanding of Topics in Physics and Teaching Assistants

[Note: this is not a session about papers that cover both topics, rather it's about 2/3 of a session on student understanding followed by 1/3 of a session on TAs.]

BW01: "The Concept of Color" of Middle School Science Teachers – Seongeun Lee, Seoul National University, aqua0202@snu.ac.kr
Cancelled

BW02: Student Understanding of Kinematics Graphs in Algebra- and Calculus-Based Mechanics Courses – Michael Loverude, CSU Fullerton, mloverude@fullerton.edu

The previous CSU lab manual was...scattered, to be charitable, leading to the creation of the new "Direction of Motion" manual, which was at least more coherent.

Student understanding was assessed with a graphical motion problem based on three moving students: Alonzo, Beth and Cho. "ABC" problem. Students using the Direction of Motion course made little progress on ABC problems, however.

So, it was asked, to what extent do underlying "what is slope?" difficulties cause poor student progress on ABC problems? The "Simple Slope Problems" were developed to answer this question. The first two turned out to be easy for all groups. However, context (math versus physics) makes a big difference. When asked explicitly about slope of a graph, students had no trouble. When given a graph and asked for the acceleration, however, the same problem became hard.

The results do suggest that building a conceptual (rather than procedural) understanding of slope may be a promising direction for future work.

BW03: Will Students Construct A Free-Body Diagram For This Problem? – David Rosengrant, State University of New Jersey/Kennesaw State University, rosengra@eden.rutgers.edu

Rosengrant's dissertation work was on the relationship between the features of a problem and how likely it was that students would draw a Free-Body Diagram (FBD). If the problem had no picture and asked for force, they did tend to draw FBDs. Both asking for force and requiring the students draw their own diagram correlated with providing FBDs...if they had to draw a picture anyway, they were going to make it a FBD.

A testing experiment was constructed with two special problems. Both problems had the same setup and physics, but while one asked for a force, the other would ask for the acceleration from the same data. Half the students were provided with a diagram, the other half were not. Thus, there were four groups: Force or acceleration, diagram or not.

The results were not encouraging, suggesting that there might have been an unaccounted-for confounding variable in there. A post-hoc analysis broke responses down by course grade, and found that course grade did correlate with likelihood of providing a FBD, but even within each grade bin there was no correlation between FBDs and which of the four groups the student had been assigned. No compelling feature other than overall course performance was found that suggested why students did or did not provide FBDs in this study.

BW04: An Investigation Into Student Understanding of Longitudinal Standing Waves – Jack Dostal, Montana State University, dostal@physics.montana.edu

The Standing Wave Diagnostic Test (SWDT) was developed at Montana State to look into student understanding of the topic.

With transverse waves, the graphical representation looks much like the physical situation (the graph = the picture), but with longitudinal waves that's not the case. Longitudinal waves are also generally more abstract, and you can meaningfully do things like draw amplitude lines for waves in a pipe that "violate" the pipe boundaries. Longitudinal representations generally necessitate a hybrid of graph and picture, and ideas taken from transverse waves can infect thinking about longitudinal waves in ways we don't intend or want.

The SWDT has 22 multiple choice items, covering waves in pipes (particle representation, pressure and change in pressure graphs, resonant frequency, etc), and is intended to take 15-20 minutes to administer. Instruction was modified with specific tutorials, which seems to help a lot in terms of student performance on the SWDT.

BW05: Facilitating Student Understanding of Motors in an Everyday Context – Jacquelyn Haynicz, Kansas State University, haynicz@ksu.edu

The overall goal of this work is to find ways to enhance learning experiences and broaden the appeal of physics courses via real life applications. The specific case under consideration involves the common kitchen blender, chosen because it's something familiar to college students and the physics involved (i.e. electric motors) is already part of the curriculum.

15 teaching interviews were conducted with students in the algebra-based sequence (9 from the mechanics semester, 6 from the electricity/magnetism semester). A phenomenographic [http://en.wikipedia.org/wiki/Phenomenography for a general view, although I think we use it a little differently] approach was used.

Two blenders were used in the interviews. A working model, and one of the same type that had been cut open to display the inner workings. Students were asked to talk about what they knew about the blender at the start of the interview. Then they were exposed to (and asked about) a progression of apparatuses related to the blender: a magnetic "railgun", a simple permanent magnet canister motor, and an electromagnetic coil motor. All of these could be taken apart and their pieces more closely examined. At each step, subjects were asked to compare and contrast with the blender.

Students tended to be in the "knowledge is self-constructed" epistemological mode during the interviews. Several phenomenological primitives (p-prims) were in evidence, such as "closer is stronger".

During the interview, attunement to affordances was used. In other words, subjects tended to look for ways to make things do stuff...if a part was movable, they'd look to see what might make it move. Subjects tended to focus on structures over function in their comparisons. They also tended to confuse or conflate charge and magnetic field, and their level of prior exposure (mechanics versus E&M semester, prior to or after the lecture on motors) seemed to have no effect.

From the interviews, a spectrum of ideas was developed that seemed to cover most student views of how electric motors work. Keep in mind that this is not a trajectory, students do not necessarily start at one end and move to the other...it's simply a description from "least expertlike" to "most expertlike".

At one end of the spectrum are mechanical explanations, such as "charges bounce off the motor and make it move." In the middle are magnetostatic explanations, where the magnets and rotor experience magnetic attraction or repulsion. Finally, at the fully electromagnetic end of the spectrum, students recognize that currents interact with magnetic fields to result in attraction and repulsion.

Future work will seek a greater variety of subjects.

BW06: Studying Student Reasoning About Energy Diagrams – Jeff Morgan, University of Northern Iowa, jeff.morgan@uni.edu

Given various quantum mechanical energy diagrams (i.e. the infinite well, potential steps, wells and barriers, etc), how do students reason about energy? How do they concretize abstract diagrams?

Students in the second semester of sophomore year taking a modern physics course were studied here. Three students in particular were interviewed in depth over the semester, looking at various energy diagrams. This talk will concentrate on student responses regarding the diagram of Bohr energy levels in a Hydrogen atom.

None of the students could explain why all of the energy levels had values that were negative. They also had trouble with confusing the height of the n^{th} energy level and the n^{th} orbital mean radius, space-energy conflation. When mapping this onto experimental results, students understood that the energies had been determined, but couldn't explain how. They tended to reason about the energy of objects rather than the energy of systems.

BW07: Interpretations of Entropy Among Advanced Undergraduates Across Disciplines – Brandon Bucy, University of Maine, brandon.bucy@umit.maine.edu

While considering various thermodynamics concepts in the research, this particular talk will focus on entropy, comparing and contrasting the methods used by students in different disciplines to deal with entropy.

For instance, in thermodynamics, $\Delta S \geq \int dQ/T$, but in Statistical Mechanics they define $S = k \ln(W)$ (where W is proportional to $V^{N/2}$ and $T^{3N/2}$).

Ideal gas processes were used for context in this study, and students asked for the change in entropy of the system and of the surroundings. Six courses were examined: four physics courses, one physical chemistry course and one chemical engineering course.

Entropy related to heat transfer (the thermodynamical approach) was used widely when ΔQ could easily be found. It was generally oversimplified in use, however, with students only considering ΔQ and ignoring the 1/T part.

Entropy related to the change in volume (statistical mechanics approach) worked well if volume was the only thing changing, and not so well when temperature was also allowed to vary. It was also not very good at finding the change of entropy of the surroundings.

Students in statistical mechanics and chemical engineering courses preferred the thermodynamics approach (ironically enough in the former case).

Each approach has its strengths and weaknesses, but a combined approach generally works best. It won't be as strong in the best situations as a one-method approach, but neither will it be as weak in any situation.

BW08: Why We Should Teach the Bohr Model and How To Teach It Effectively – Sam McKagan, University of Colorado, mckagan@colorado.edu

[Original title: Should We Teach the Bohr Model?]

Teaching atomic models helps with content, beliefs/attitudes and model-building skills, all of which are relevant to students from kindergarten through college.

However, teaching the Bohr model in specific is controversial. There's a common view that Bohr is an obstacle to student understanding, a historical step that is just as soon skipped in education, and that students should go right to the Schrödinger model.

McKagan feels that the research supporting the "Bohr is an obstacle" position is flawed, however, and points out that working physicists use the Bohr model. Besides, odds are that our students will be exposed to it in high school anyway, so simply skipping it isn't an option. Better to address it and its role in the development of models. That, and getting faculty to omit Bohr at any level might be a functional impossibility.

Students are instead asked to compare and contrast various models both in class discussions and in homework assignments. They're given not just Bohr, but also the billiard ball model of atoms, the plum pudding model, the planetary Rutherford model, the deBroglie standing wave and the Schrödinger wavefunction. Then they're asked in an essay question format to describe Hydrogen in its lowest state, without specifying which model the students should use.

About ¼ of them use Bohr exclusively, and only 2/3 use the wavefunction model in any form. There's a 5 week gap between presentation of the deBroglie model and the wavefunction model, so the sequence got broken apart during the time needed to build up student background and allow understanding of wavefunctions.

The course has since been redesigned to focus on the contact between the models. In the revamped course, only about 1/10 of the students used Bohr exclusively, and 4/5 used the wavefunction model.

For more, see http://arxiv.org/abs/0707.1541

BW09: Transfer of Learning in Medical Imaging: Analogies and Computer Simulations – Spartak Kalita, Kansas State University, spartak@phys.ksu.edu (title does not match program)

Part of the Modern Miracle Medical Machines (MMMM or 4M) project, focusing on the application of modern physics to medical technology.

In this talk, the focus is on Group Teaching Interviews (2-3 students in each group) to look at the social interaction component of transfer. A cyclic framework involving learning cycles and modeling cycles was used.

Stage I of the interview comes in two parts. Part 1 shows the students pictures of various imaging technologies, like CAT scans and X-Rays, and has them discuss what they know. Part 2 uses translucent Lego™ bricks to simulate soft tissue and has students measure light absorption to build up an image of what's inside a structure.

Stage II focuses on what students see, and how they see it.

Students were found to trigger and reinforce each other's transfer process. A peer instruction module developed based on these materials looks promising.

BW10: Learning About Physics Graduate Student TAs Using Case Studies – Renee Goertzen, University of Maryland, goertzen@physics.umd.edu

GTAs in an algebra-based introductory course are responsible for running tutorials and teaching labs. Case studies were used to characterize these GTAs and figure out what their jobs really are, as opposed to what they are on paper. "How do TAs listen to student ideas?" is the aspect of these case studies focused on in this talk.

Different GTAs follow instructions to "listen" in very different ways. A good listener is active, trying to figure out what does and doesn't make sense in what the students are saying and helping draw out their conceptions. A bad listener just keys on targeted answers and jumps in as soon as one shows up, encouraging students to parrot book responses rather than revealing what they actually think. The bad listener filters too much, and is too quick to assume based on flimsy evidence that students understand what they're talking about.

[There were two more talks in this session, but at this point I'd been taking notes nonstop for four hours and could literally not write any more. My apologies to the speakers in BW11 and BW12.]

Tuesday, July 31, 2006

CB – Conceptual Understanding: Models and Measurement I (invited)

CB01: The Dynamics of Coherence in Conceptual Understanding – David Hammer, University of Maryland at College Park, davidham@umd.edu

Based on a chapter in Handbook of Research on Conceptual Change.

Coherence means that something holds together. It doesn't mean correct or consistent with an expert view, just that the subject sticks to their story.

Concepts and misconceptions are recurring, stable patterns of sensible reasoning, misconceptions being those that contrast with the canon. Both have recognized exceptions, for instance, "gloves keep hands warm" but "an oven mitt is a glove that keeps hands cool".

Now, consider coherence of the "theories of science" variety. Theories -

- Are understood as structural. i.e. "blankets and gloves are warm."
- Have domains of validity. i.e. "except oven mitts." Any theory can be pushed into inconsistency at the periphery of its domain.

Learning requires structural change on the part of the learner. The implications for the instructor are things like ECR/accommodation, and it's easy to generate research ideas with this model. Just find a topic and poke at it for coherence!

However, this comes with a number of concerns.

- It's too rational, and doesn't distinguish between stable long-term theories and ad hoc constructions put together on the spot.
- "Domains of validity" explanation is inconsistent with robustness of student conceptions. If domains of validity are really what the theory says they are, it should be relatively easy to introduce exceptions into student concepts, no ECR needed.
- There's no mechanism in the theory for change, or of productive resources. We know people change theories, this model has no explanation for how that happens.
- The zealous hunt for misconceptions to shatter has failed to focus on how student ideas are often quite sensible in their own ways.

Instead of this model, we should consider generation of ideas from implicit presuppositions and/or context-sensitive primitives. These are dynamic accounts.

Presuppositions are knowledge structures that underlie and constrain reasoning, but do not directly determine reasoning. For instance, students presuppose that "hotness" is a transferable property. Coherence in this case is a result of the cognitive system adjusting to constraints, such as developing "blankets are warm" from the "hotness is a property" presupposition.

P-prims are knowledge structures underlying reasoning that get activated by the particular context. They aren't presupposed (i.e. they're not always lurking at the surface), they're dynamically cued. For instance, blankets could cue a "covering" p-prim, while oven mitts cue a "blocking bad stuff" primitive. Coherence and systematicity arise from the dynamics of cueing. A p-prim with high cueing probability can act like a presupposition.

In other words, presuppositions are always there and always active, while p-prims may be dormant and never get involved in a particular situation, requiring some sort of cue to bring them out.

Three mechanisms of coherence were presented.

- Structural: conceptions, presuppositions and p-prims, coherence arising from the way the brain puts things together.
- Contextual: viewing order, question order, and other cases where changing the situation affects coherence.
- Epistemological: paying attention to what you're saying, some deliberate choice to try to establish coherence, such as addressing first principles.

In general, we want to temper our tendency to attribute things to fixed knowledge, as it's more fluid than that. We want to identify multiple stable dynamics and the transitions between them, then figure out how to trigger those transitions on purpose.

Dynamic accounts come with their own concerns, however. Presuppositions also lack a mechanism for change and don't describe the nature of coherence. P-prims lack an explanation for how they're generated, they come in multiple varieties, and we're not clear on their substructures. All of these are soft constraints.

Thelen and Smith's 1994 "A Dynamic Systems Approach to the Development of Cognition and Action" was recommended as useful.

Arising in Q&A: Might p-prims and concepts be the same sort of thing, just with different grain sizes? Non-linear situations mean that larger scale dynamics aren't simply the sum of component properties. Our tendency to describe fixed properties is a "stable attractor".

CB02: Measuring Understanding: Dependence of Student Answering on the Question Task – Andy Heckler, the Ohio State University, heckler@mps.ohio-state.edu

What is the meaning of "understanding"? Commonly, it's considered the ability to answer correctly in multiple contexts. These contexts will seem similar to the expert, but dissimilar to novices.

Understanding can be thought of as knowledge that is context-independent, so we want our instruction to help students overcome context-dependence. To overcome it, we must study it.

Context dependence includes the content of the question (i.e. impetus only used sometimes), the format of the question (i.e. diagrams versus words, variants of the Tower of Hanoi problem) and the question task ("what do we want them to do?"). One can compare consistency of tasks within a question.

Equations and diagrams have a dual role. Not only are they tools that help the students solve a problem, they give the instructor insight into the student's mind and aid in assessment.

An example of the effect of question task would be to ask if requiring the answer in diagram format (task = "draw a diagram to answer the following...") affect performance? Question tasks can trigger incorrect answers, though. For instance, if you specifically ask for a Free Body Diagram, students may add in some inappropriate forces just to have a really snazzy FBD.

Does the ability to draw diagrams imply the ability to solve problems? Is it necessary, sufficient, or neither? Trends seem to indicate that with this particular task, a correct diagram is associated with a correct solution and an incorrect diagram is associated with an incorrect solution, but these are not absolutes. Some work with incorrect diagrams still has correct solutions (i.e. diagrams in which Force follows velocity frequently have correct work otherwise). Correct with correct happens more often than incorrect with incorrect, suggesting that diagrams are sufficient but not necessary.

However, it was also found that requiring a diagram can degrade performance on conceptual problems! Requiring a diagram often cues students onto a more formalistic path that trips them up, when the same student might have been correct had they relied on their intuition instead of trying to do it the "book way". This was found in numerous investigated problems...when no diagram was required, students were more likely to rely on correct intuition than on poorly-understood formalism. So, are diagrams more of a good thing or more of a bad thing?

In cases where FBDs were not required, such drawings as students did make tended to be more concrete, pictures rather than diagrams. In cases where the FBD is required, drawings were very abstracted, all points and arrows.

Conclusions of this study:

- The question task does matter.
- Inconsistencies within a question task can illuminate student issues.
- Task-dependence does make analysis tricky.
- Open-ended problems are not equivalent to ones that specify more, in cases where you wish to compare performance on multiple items.
- Students can learn to solve force problems before they learn to use FBDs. [Anecdotal: I knew someone once whose teacher mistaught FBDs badly, and despite having a malformed tool in their toolbox, the student was still able to solve most statics problems.]
- Requiring students to use FBDs both helps and hinders novices, depending on the situation.

CB03: Conceptual Change and Transfer: Consolidating Various Viewpoints – Sanjay Rebello, Kansas State University, srebello@phys.ksu.edu

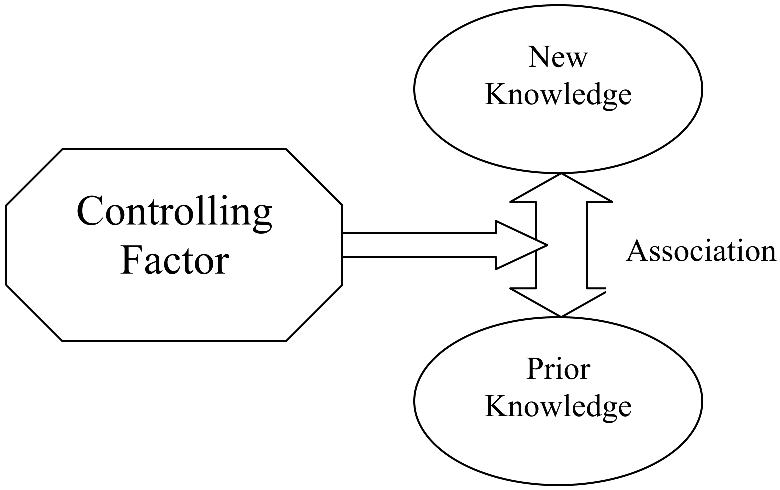
Transfer, in general, is the ability to use stuff learned in one situation in a different situation.

[Digression for some vocabulary. I've coined a couple of terms to describe the difference between traditional views of transfer and some of the new ideas that have come into currency: paleotransfer and neotransfer. These terms are not used in this talk, but I will use them to refer to the classical and newer ideas respectively.]

Paleotransfer requires that identical elements must exist between situations and a coherent model is needed...it all goes or nothing goes. In researching paleotransfer, the research decides what the target element is, uses a one-shot assessment, and focuses on internal knowledge. In this style, transfer is rare, almost to the point of never happening.

Neotransfer involves the transfer of pieces, reconstructing them in a new context. Not all pieces necessarily make the trip. Researchers use dynamic assessment such as interviews, have no pre-chosen elements (they look to see if anything transfers, not just topic X), and focus on mediating factors like motivation. Neotransfer is ubiquitous, we're always doing it. Transfer in this case may involve learning.

The KSU model of learning is that it is the creation of associations between new information and prior knowledge, with controlling factors mediating that association.



The vertical double arrow in the diagram is an association. There are two types of association:

A) Assignment of a new case to an existing knowledge element (see how X fits into Y)

B) Construct a link between two existing knowledge elements (see how X relates to Y)

Models consist of elements connected by associations, and any element could potentially be broken down into sub-elements connected by their own associations.

Horizontal learning is a type A association. The model already exists, you're simply fleshing out the cases to which the model applies. Thorndykian drills are a form of horizontal learning, for instance, where you just keep practicing with new examples to solidify understanding of a given model.

Vertical learning happens when you just can't hammer that new idea into an existing model. Either the model lacks a necessary element, or the elements are not configured in a way that works. So you create or suppress associations in order to generate a new model. This requires conceptual change.

These two types of learning can be associated with efficiency (horizontal) and innovation (vertical), or with model deployment (horizontal) versus model development (vertical), or with assimilation (horizontal) versus accommodation (vertical).

Horizontal and vertical learning experiences are **not mutually exclusive**, and cannot be universally labeled. A single activity may be horizontal for one person and vertical for another, it depends on whether their existing models can handle the experience or not.

Learning is facilitated by cognitive conflict, happens in the Zone of Proximal Development (ZPD, the range between "too hard to learn even with help" and "easy enough to learn without help"), and results in model building. It follows a general cycle of exploration/concept construction/application. Concept construction can be seen as vertical learning, or model development. Application is a horizontal style of learning, or model deployment.

The idea is to strike a balance between horizontal and vertical learning, along what is called the Optimal Adaptability Corridor (OAC). If graphing learning in two dimensions, the OAC is like a diagonal path, with student learning zigzagging vertically and horizontally within the path (for ease of display, the OAC is usually shown as a path of constant width, but it may vary, with excesses of vertical or horizontal allowed at various points along the way).

Too much of one kind of learning in a single sitting can be bad. Too much time spent trying to learn vertically can lead to the "frustrated novice" who stalls out, lacking sufficient experience with existing models to know how to usefully modify them. Too much time on horizontal activities can create a "routine expert" who is very good at defined tasks but has little ability to beyond their boundaries. Too much vertical leads to confusion, too much horizontal to boredom. Someone who stays in the OAC is considered an "adaptive expert".

The modeling cycle as well as Vygotsky's Zone of Proximal Development can be adapted to trying to keep a student in the OAC.

Figuring out where the boundaries of the OAC are can be difficult, but there are some guidelines. Excessive horizontal learning will start to reach cognitive dissonance at the "floor" of the OAC, as the student tries to perform more tasks that don't work within their current model, and just hammers them to fit. And one can scaffold activities upward until the ZPD runs out to get the "roof" of the OAC. (And yes, finding the ZPD is also a non-trivial task.)

So, how do we design instruction with the OAC in mind? The standard development cycle runs from clinical interviews through curriculum development to field testing and then back to more clinical interviews. It is suggested that a step be added between clinical interviews and curriculum development in order to better gauge how the OAC can be held to. This step is the Teaching and Learning Interview.

Teaching/Learning interviews are "mock instruction" in a one-on-one or one-on-small-group setting of up to 3 students. This lets you not only see where the students are, as a clinical interview does, it also tries to get the students to develop from that starting point, seeing how successful various activities might be. The research is not just an observer in this case, but an active participant (although a second researcher may be around as a pure observer). This method grants more insight into the learning process.

As seen in these interviews, students deploy strategies based on what they **think** they need to do in order to succeed. When they meet with failure, this generates cognitive dissonance, inspiring students to develop a new model to meet the current needs, and encouraging metalearning.

Similarly, we can use metateaching, such as taking the OAC as a tool to help us develop our teaching model. For instance, Mazur's FCI results were a "failure-inspired cognitive dissonance" for him.

Neotransfer may be indistinguishable from learning in general.

[Points that came up in Q&A]

OAC walls are not very solid. You have to impose consequences for cognitive dissonance, for instance, or they will just plow through with their horizontal training, or keep hammering on a vertical task without thinking that maybe they need to refine their existing models first.

Simply following the OAC may not make one self-adaptable, there may be a third axis that charts such flexibility.

Hestenes added that there are teaching/learning interview elements built into the modeling cycle as part of its dynamic process.

Vertical learning has the affective "challenge" aspect to it.

Labs Committee Meeting

[I suppose I should take more careful notes here from now on...my Seattle notes were used to supplement the meeting minutes when we tried to go back and figure out some of the points. Unfortunately, I had very little information on the topic of interest written down.]

Various sponsored sessions at Greensboro were brought up, but no one at the committee meeting had attended any of the ones that had yet happened. Session BB, sponsored by Labs, was reported to have good attendance (about 30 people) and was well received, according to a later arrival at the committee meeting.

Baltimore, Winter 2008

Discussing the Baltimore meeting, it was pointed out that two counties in the area use a Physics First curriculum in secondary school, so a session covering labs for Physics First would be a good idea. It would be best scheduled for Monday, though, since that's the day that high school teachers get into the conference for free. Leon Letterman was raised as a good anchor speaker for such a sessions, if a way could be found to invite him. (Letterman is a Nobel laureate with an interest in Physics First.)

The submitted session on "Laboratories with Biomedical Applications" for the Baltimore meeting lost the "bio" part somewhere along the line, and it needs to be put back on in the AAPT system.

Labs committee is co-sponsoring a Daytime Astronomy Labs workshop, but no one could remember how that happened.

The topic of "Why is the Baltimore meeting scheduled so late in January?" was brought up. A lot of us [myself included] have schedules that make it difficult if not impossible to attend that week. The best guess at reasoning was that being later put it past the beginning-of-term rush and would let people more easily find substitutes and attend, and for some reason doing it before the term started wouldn't have worked.

Edmonton, Summer 2008

Considering various types of workshops for Edmonton. A "Photon Quantum Mechanics Laboratory" workshop was proposed, although such labs are very expensive to set up, a minimum of $15,000 in equipment. It was agreed that they're worthwhile, though. "Falsification Labs" workshop was also proposed. It's a workshop that's been done in the past, but not very recently, and it bears repeating. Such labs probe misapplied resources. (Of course, some students in falsification labs manage to "confirm" the falsities anyway.) Tim Erickson is involved as is Ayars.

We do not yet seem to have a theme for Edmonton, if there is to be one.

There should be an Advanced Labs workshop at Edmonton, although international shipping will make the apparatus end of things a bit trickier than usual. As far as we could tell, though, Advanced and Intro Lab workshops are actually run by the Apparatus Committee, so it probably wouldn't be our concern.

A paper session on labs for Physics First was proposed for Edmonton. If the workshop in Baltimore doesn't work well, the paper session can always be dropped before the deadline.

A "Writing In Labs" or even a "Different Forms Of Communication In Labs" session was proposed, to solicit papers on both intro and advanced labs and get some communication going between those specializing on both types. Marsha Hobbs will sponsor it as an Invited/Contributed mixed session, possibly seeking David Haas as an invited speaker. Try to focus on writing rather than other forms of communication, though.

There was demand for an Advanced Lab paper session, and there was a lot of back and forth on the relative merits of this.

Advanced Labs Goals Statement

A draft document of the Labs Committee Advanced Lab goals statement now exists. It's based on the existing statements for high school and introductory labs, mainly in format. Feedback has been sought from the members of the advlabs-l mailing list.

[Meeting had to break at this point, a lot of time was spent on the Advanced Lab paper session, leaving not much for the goals statement.]

DF – Conceptual Understanding: Models and Measurement II (invited)

DF01: Facets and Facet Clusters as a Framework for Organizing, Interpreting and Analyzing Student Responses – Jim Minstrell, FACET Innovations, jimminstrell@facetinnovations.com (with Pamela Kraus)

The NRC (National Research Council, not Nuclear Regulatory Commission) Assessment Triangle has the vertices of Observation, Interpretation and Cognition.

Learner responses often make perfect sense from their own perspective, or from the perspective of fellow learners, but not from ours.

A facet is a unit of analysis, one or more pieces of knowledge or reasoning not believed to exist in learners' heads as a separate object. Like p-prims, but at a larger grain size, they're more of an "observable" in the quantum sense, existing when you measure for it. Facets are meant as a usable teacher tool at the K-12 level, implemented in classrooms and used as a translator between teachers and researchers.

Goal facets exist in clusters, each cluster has sub-pieces, and each sub-piece has levels. Within a topic, all the sub-pieces and levels are labeled with two digit numbers, with the tens digit being a genus and the ones digit a species, more or less. So, facets 01-09 are generally correct or nearly correct student views, while 91-99 are pretty awful. The genus label has 0 in the ones place, so "50" might code for the general idea that the magnitude of a force is tied to the amount of effect something has, while "51" could be "force is proportional to impact damage".

The free site http://www.Diagnoser.com is meant to help identify facets. It can be used by students, but needs instructor guidance! It's mainly intended for instructors and researchers, and helps them find the likely facets that a student is displaying. Diagnoser kicks out a summary report (including, against Minstrell's preferences, a "number correct" sort of datum) with feedback that may be of help.

The class summary from Diagnoser is presented by genera, and provides suggestions for activities that might help a class that is dominated by a particular genus of facet. The report can also be broken down by student, useful in a smaller class or where one or two students stand out as having problems that need attention.

Not all items have all possible codes attached to them, just due to limitations in how many options you can meaningfully provide on a question. A certain question simply may not be amenable to code 62, or code 62 may have come up too infrequently in the open development to merit being one of the multiple choice options. Therefore, a student who really believes consistently with code 62 might pick their "second best" choice and be misdiagnosed on that item.

To get around this, Diagnoser's report lets you select only items where a particular code is available, and then see if the student consistently picks a response tied to that code or not, determining if there's a true mixed state or not. It helps that many item choices come with multiple codes, as there can be different reasons for a single response (i.e. 62 and 91 both attached to an item). If an inconsistency in responses persists, you can query the database to check for context-based issues.

In Q&A, Michael Wittmann endorsed Diagnoser as a good tool to help one change teaching methods quickly, but warned that it doesn't tie in too well with outside resources like the University of Washington tutorials.

DF02: Comparing the Probabilistic Frameworks of Popular Quantitative Education Measurement Methods – Lei Bao, the Ohio State University, lbao@mps.ohio-state.edu

PER tends to use assessment methods and models that are not the same as those used in general education research, for various reasons, and not all of our measures are good ones. For instance, the flaws of <g> have been brought up repeatedly in the past few years (ceiling effect, lack of theoretical grounding, etc).

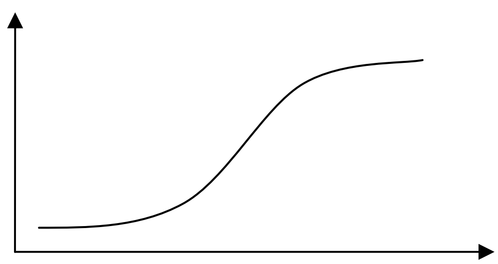
Therefore, we need a coherent theoretical framework for assessment.

Depending on the goals, we have various types of measurement available:

- When assigning scores/grades, we can perform a one-time measurement and get a score.
- When evaluating the effectiveness of a method, we take two or more measurements over a relatively short timespan.
- When trying to find the patterns of associative probabilities, we can use Model Analysis.

Regardless of what method we use, there are a number of issues that need to be considered, including the dynamic range, the measurement window, test equivalence, measurement of change, and the multidimensionality of our knowledge structure. Some tools address some of these issues and not others.

Item Response Theory is commonly seen in education research. It's a nonlinear probability model that adequately addresses the floor and ceiling effects by essentially unfolding the bell curve. Ability is tracked on the horizontal axis, and score as a fraction of 1.00 on the vertical:



Different instruments may give differently shaped sigmoidal curves, with the bends in different places for skewed distributions.

<g> is nonlinear like Item Response Theory, but lacks an underlying probability theory. Model Analysis, for its part, completely lacks an "ability" dimension.

To go into more math [and strain my talents with Word], for the horizontal axis being labeled as θ (with 0 at the center) and P(θ) being the score:

$$P(\theta) = \frac{1}{2\pi}\int_{\theta}^{\infty} e^{-\frac{y^2}{2}} dy \qquad \text{or} \qquad P_{kj}(\theta_i) = \frac{1}{1+e^{-a_j(\theta_k - b_j)}}$$

Essentially what those equations do is look at the area of a bell curve in y that is above the defined success level θ, and then chart that as P. So at the right end of the IRT curve, the entire group at θ successful, while at the left end no one is successful. The index a_j is the discrimination of item j, b_j is the difficulty of item j. There was also a c_i involved, but I didn't catch the equation that used it. In any case, defining "ability" is very tricky, no matter how many subscripts you tack onto it.

Normalized gain <g> can be fit, post-hoc, to a probability framework like IRT's. In this case, the main parameters are student ability (or efficiency of learning) α_k and knowledge not yet possessed by the student, s_k. IRT doesn't have these parameters per se, so gain extends our ability to measure things. Consider a two-level system, where wrong is 0 and right is 1. We have s_k at 1 and $(1-s_k)$ at 0. [Okay, I think I may have swapped some things here.]

A basic learning theory would say that $ds_k/dt = (1-s_k)\alpha_k$ [Similar to Pritchard talk at the Seattle 2007 meeting]. The more they don't know, and the more efficiently they learn, the faster they learn. This assumes an independence between what they know and what they don't know.

To bring in the idea of connected knowledge, where understanding part of the domain helps in learning more, change that to $ds_k/dt = \gamma_k(1-s_k)s_k$, a where gamma measures ability to apply what you know to figuring out what you don't know. A beta coefficient represents the rate at which what you already know can impede learning.

For a purely alpha process, in which rate of learning does not depend on the nature of already-gained knowledge, gain on item k should be a fairly simple exponential equation, $g_k = 1-e^{-\alpha kT}$ where the pretest scores end up canceling out. When you bring in the gamma process of connected knowledge, you get a much more complex exponential equation.

Graphing with pretest scores (out of 1.00) on the horizontal axis versus absolute gain (post-pre) on the vertical axis should give a pretty much straight line graph with slope of -1 for a pure alpha process, and more of a ^ shape for a purely gamma process. The actual data seems to lie somewhere between these graphs, suggesting a mixed alpha/gamma process, where the ratio of how much is which type is a characteristic of a particular class.

The alpha process is a characteristic of good students, who are not bothered when they hit a hole in their knowledge, and learn what they need to. The gamma process seems to actually represent where prior understanding hurts students [contradicting my earlier note about beta], slowing them down when something doesn't fit what they already know. So you want a high alpha to gamma ratio in your class.

Gain does seem to measure the extent to which you have an alpha class, and has an unintended but robust framework. It should not be abandoned, simply used in conjunction with Item Response Theory.

DF03: Defining "Conceptual Understanding" Through Appropriate Constraints on a Knowledge Domain – David Meltzer, University of Washington, dmeltzer@u.washington.edu

This talk overlaps with work presented in Michael Wittmann's recent paper on resource graphs. "Concepts" defined here as per Karplus, DiSessa and Reif.

Concepts are diverse, including things like specific measureables, specialized descriptives, and general or philosophical things (like explanations or proofs). Concepts only have meaning within a system that includes them as particular cases (i.e. Vygotsky's view of concepts as cells in a multicellular organism). Karplusian concept clusters have empirical or theoretical interrelationships, and one must deal with the entire cluster as an entity for purposes of instruction. This instruction, according to Reif, must promote knowledge organization.

Concept clusters can be assembled in multiple "correct" ways, but a flawed concept can poison the whole cluster, and inadequate organization plays a key role in hindering understanding. These two problems, of bad concepts and bad organization, can be hard to disentangle.

Redish proposed a model of a concept cluster that looks like a bull's-eye target. In the central circle, the subject's understanding is solid and difficult to change. In the outermost ring, only scattered bits of knowledge exist, and there's not much to latch onto for purposes of building on the subject's existing structures. But the middle ring, equivalent to the ZPD, things are connected enough to build on but loose enough to change if need be. We need to figure out a way to map out this middle ring more completely in order to bring students into the bull's-eye.

Meltzer takes Redish's model and fleshes out the middle ring with a simplified version of a much more complex concept mapping model borrowed from Chemistry Education [email him if you want the full details] keyed the following way.

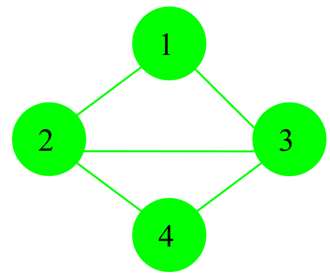
Green ovals represent good, stable concepts. Green links between ovals are good, stable associations between concepts. Red ovals are concepts that are either bad but coherent, or good but unstable. Red dotted lines are similarly bad/coherent or good/unstable associations. Gray ovals are both bad and unstable concepts, and grayed out links are those that should/could be there but are missing.

Once a student's middle ring has been mapped out this way, it's possible to try to bring gray and red elements up to green. Note that any arbitrarily chosen concept cluster (ovals and links) will evolve over time, even if it's mostly in the green. There are no permanent intrinsic qualities.

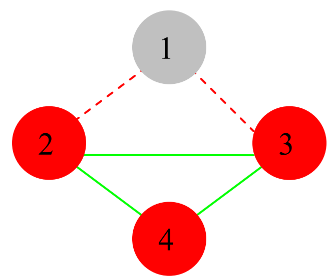
An empirical example was presented based on research done at Iowa State, the "Entropy-Increase" Cluster. It is composed, in "correct" form, by four elements:

1. $\Delta S_{universe} > 0$ for any real process
2. $\Delta S_{arbitrary\ system}$ is indeterminate
3. $\Delta S_{surroundings}$ is indeterminate
4. Any entity can be chosen as part of either the system or surroundings

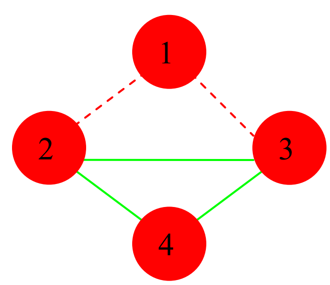
In a correct cluster, these elements are arranged as follows:



A common problem with students has them assuming that the entropy of the system must always increase, which mapped out as:



After undergoing a tutorial on entropy, student responses were mapped again, yielding the slightly better cluster below:



After this, a more focused tutorial on slow heat transfer was used to specifically address this cluster, and things were brought nearly all into the green (a few ovals and links were on the low side of green, mind you, such as 53% of the class being correct).

A similar example was given for a cluster involving electric potential and fields, which brought some red elements into green, but didn't affect a gray element that wasn't specifically addressed in the tutorial.

A "concept" can be seen in this model as an arbitrarily circumscribed portion of an interlinked array of these ovals. It's necessary to probe a specific set of elements and links in order to assess conceptual understanding.

When asked to what extent a class-wide "all green" cluster actually reflected the views of any one student (i.e. if 53% correct meant a green circle, what about the other 47% of the students?), Meltzer replied that more detailed information is needed in order to see what turns any given oval red or gray.

We can use this structure to follow the progress of a student towards an all-green model, but there are other useful interpretations, such as using it to follow a student wherever he wanders and see what structures arise, rather than herding them in a particular direction.

[Q&A]

Words like "concept" have hidden assumptions, and we don't all agree on the meanings...a glossary of terms probably wouldn't ever be accepted fully. That's why some people, like Minstrell, simply sidestep the issue by making up new terms. In any case, we need to negotiate the meanings rather than freeze them in a glossary too quickly.

[In later discussion back home, the issue was brought up, "How can you have a green connection to a red or gray oval?" While possibly just an artifact of the compressed notation, we did come up with a few ideas that might work. For instance, you could be pretty sure a concept is there, and be properly connected to it, but be hazy on the details of the concept. The example given was Bessel Functions...it's been over a decade since I had to use one, and my oval labeled "Bessel Function" is red or even gray, but the connection it has to what it's **used for** is pretty solidly green even today.]

Wednesday, August 1, 2007

ED – Introductory College Physics Textbooks: Current Role and Future Possibilities (Panel Discussion)

Moderator: Peter Lindenfeld, Rutgers University. The format is four 15 minute talks followed by an hour of panel discussion.

ED01: How and Why are Textbooks Used in a Physics Course? – Suzanne Brahmia, Rutgers University, brahmia@rci.rutgers.edu

[Speaker is co-author of a textbook with the moderator.]

Textbooks as we know them were set in their form by Millikan, and haven't really changed much since then.

High school classes commonly use a lab manual, study guide and test bank as ancillaries to the main text, all provided by the publisher. The authorial focus is similar to what's important to the teachers and students, with material covered in the order it is to be presented.

In college, students and teachers want different things. Students want lots of sample problems and a way to help expand on the lectures, while teachers want just a general outline and a bank of homework problems. Neither student nor teacher really reads the text, nor are the chapters necessarily followed in order or without skipping. However, the author still writes the same sort of book as for high school students, assuming that not only will the book be read, it'll be read in order with no skipping. As a result, the book can end up meeting the needs of neither student nor instructor.

In college, the instructors are generally free to pick the text they want to use, but spend little time on the choice, as texts seem to be interchangeable. Since the instructors control the choice, though, publishers market to these uncaring faculty rather than to the students who will have to actually pay for the thing. Fewer than half of these paying customers read anything but the problem-solving sections, only one in seven read the book before the material is covered in lecture, and there seems to be no correlation between reading the book and doing well in class. 2/3 of the students use the book as a problem-solving manual and that's about it...and for this they pay $200?

Rather than address the existing value for money question, though, the rest of the talk focuses on whether students really are missing out by not fully reading the text. Consider the metacognitive process of **critical reading**. It goes beyond just the decoding of words, and leads to understanding. We process on the fly, then reflect afterwards. Non-critical readers, however, are satisfied with recognition and restating (i.e. parroting) while critical readers reflect on the meaning.

Problem solving and critical reading use a lot of the same metacognitive processes, differing only by the fact that critical reading uses decoding while problem solving uses computing...all other processes are the same. Additionally, poor critical reading skills can impede problem solving, such as trouble with pronouns (what does "Them" mean in this case?[1]) or difficulty in spotting hidden assumptions.

We want students to display critical thinking in general, and critical reading is part of that. The better a text is at transmitting the canon, the worse it is at fostering critical reading (as per P. Richardson, 2004, research on Economics texts).

Unfortunately, just knowing that reading is fundamental doesn't ensure that students will learn from the book. They will try to "do the reading" but frequently fail to read effectively. If we really value critical thinking and reading, we need to have textbooks that help it rather than hindering.

[[1] When in doubt, "Them" always means giant ants.]

ED02: An Editor's Perspective – Stuart Johnson, Wiley & Sons, sjohnson@wiley.com

Briefly broke from prepared talk to assuage audience concerns, stating that customization of texts (i.e. omit chapter 14, swap order of chapters, etc, so that your text follows the syllabus) is possible.

An editor, in the speaker's case, does not mean proofreader. Rather, he's a Managing Editor, involved with finding good authors and giving them work. However, the job is entering new territory of late. Such as...

The Introductory Large-Class Textbook of 2012! The remainder of the talk addresses how Text2012 will be put together, both from the instructor's and the student's viewpoint.

The instructor will experience the least change, should they desire to avoid change. A hardcopy textbook will still be made available to instructors, more on a Print On Demand basis, using either 2-color or even black and white printing to save on costs since this will be a much lower printing volume. The lifetime of an edition is expected to increase to about 8 years. Publisher research indicates that the typical instructor is, sadly, very conservative, unwilling to accept change or innovation, and wants to be able to keep using the same lecture notes that they first wrote a decade ago.

Students will mainly get some sort of electronic text, whether on a disc/chip or via a password-protected webpage (or both), although they will be able to buy the dead tree edition should they really want one (few are expected to, but some will always prefer a hardcopy). The e-text will be dynamic and interactive with multimedia and search options.

Publishers will be focusing on the e-texts. The instructors have been satisfied with The Way Things Are Now aside from the inevitable minor grousing, but the students really want change. Especially spare change...textbooks cost a lot in hardcopy form, but comparatively little in electronic format. And cost is the deciding factor in making the move to e-texts, although not the sole benefit. The efficiency of e-texts is also expected to be a big selling point...Google/Wikipedia versus Encyclopedia Britannica, basically.

The 19-23 age group is much more comfortable reading online than the 35-60 age group [although there are, of course, exceptions], and publishers need to pounce on that demographic before Wikipedia or something similar comes along and eats their market.

Additionally, the used book market is becoming a huge headache for publishers. Rather than keep books in stock, campus bookstores are increasingly shipping buybacks to the publisher, then turning around and ordering used books from the publisher in order to save on inventory costs. But this results in shipping charges making up about 15% of the cost of textbooks!

And, of course, campus bookstores haven't exactly been the best pals of the publishers lately anyway, so the publishers are not shedding many tears over the thought that e-texts will effectively cut bookstores out of the picture. Especially since big chains like Barnes & Noble now own most campus bookstores, so it's not like they'll be putting many out of business by the time they shift to an e-text basis.

From the Author's side, e-texts will end the pressure of "edition churn" [my term] that forces them to start the next edition before the current one is even on shelves. Instead, they can focus on developing supplementary material that goes beyond the core book. Delve deeper rather than keep going over the same shallow material again and again.

It is more likely that students will be able to outright buy an e-text than renting it for just a term [although there might be yearly fees to keep downloading the extensions, plus eventually they'd need to buy a new version that's compatible with their OS upgrades]. And given that students basically rent hardcopy texts by engaging in the endless cycle of buyback, owning the e-text may be better for them in the long run.

Don't expect instructors as a whole to ever like e-texts.

ED03: Self-Publishing – Elisha Huggins, Dartmouth College, Elisha.R.Huggins@dartmouth.edu

[Before I start with notes on this talk, I feel I should disclose my personal biases regarding this speaker. He's had a table in the display area at most AAPT meetings for the past several years, and I've talked to him on occasion. I always got the impression that he's the kind of person who's certain he's right simply because the mainstream doesn't accept him, the "they laughed at Galileo" sort of attitude. He may have valid points, but they tend to get lost in the zeal of his sales pitch and his "bitter outsider" demeanor. Well, cheerfully bitter, if that's possible. Still, regardless of whether he's right, he often comes across as the sort that gives self-publishing the taint of "vanity press".]

If you want to publish a nice, full color textbook with hardcover binding and look the way most of those on the shelf do, the printing costs come to about $100 a copy. That's before distribution costs, profit, etc, and not including payment to the content creators. However, a tape-bound black and white book can cost as little as $6 a copy to print using Print On Demand methods, and the cost to burn a CD is negligible. So, if you want to self-publish, no-frills books or books-on-CD are definitely the way to go.

In his case, he decided to write an introductory textbook that starts with Special Relativity and bases as much as possible on it. But, because it had never been done before, no publisher would touch it. It was a catch-22: until there was a text that started with special relativity, people wouldn't try teaching that way (in general); until people started teaching that way, there was no demand for a textbook and the publishers wouldn't make one.

Introductory physics courses are trapped in this catch-22, like Scrooge in need of some spirits to shock some sense into it. It focuses on the past (over a hundred years ago), while introductory chemistry focuses on the present and introductory biology focuses on the future. We may "duct tape" some modern physics onto the intro course, but that material rarely gets used since it's all at the end. Introductory physics is in danger of being moved into the History Department within the next 15 years.

Instead, we should integrate modern physics into the introductory course from day one. Some schools that have started using his course report evidence of success, although there have been failures that were due to misuse of the resources (i.e. teaching the standard class out of the new text, skipping around to avoid starting on modern physics).

Regardless of how you want to revamp things, though, these days it's pretty easy to assemble a professional-looking textbook and distribute it.

[I'll bring it up again in the panel discussion where it happens, but Stuart does counter the assertion of publishers being closed to niche texts. They will and do publish very small print run texts along the lines of the "Relativity First" text, even for just a tiny audience.]

ED04: What Can We Expect From The Textbook Of The Future? – Edward Redish, University of Maryland, redish@physics.umd.edu

[Redish admits that this talk was written at 7AM the day of the talk]

"If it's going to take the Titanic 10 years to sink, it might be useful to rearrange the deck chairs."

The textbook is optional in Redish's class, but recommended. About a tenth of the students buy it, while others check out copies from the library or outright ignore it.

Textbooks often ignore active learning research, lacking interactivity. However, they can **complement** active learning environments.

Sokoloff and Laws developed the "Physics Suite" that connects multiple resources into a package: textbook, teacher's guide, ILDs, workshops, tutorials, etc. At the core of the suite is a unified philosophy of learning guide. This cluster can be used in many ways and orders.

The philosophy at the root of the cluster has an epistemological orientation, explaining rather than being didactic. The chapter summaries are removed from the main text so that students don't have a cheat around the chapter...they need to come up with their own summaries.

Now, hardcopy texts do create constraints (as per ED02). One size fits none, corporate agglomeration tends to wipe out the niche publishers and edition churn burns out the authors.

Web delivery of content allows for many good things: hyperlinks (for depth and interconnectivity), multimedia (for active learning), nonlinear structure, and community development opportunities (a la Wikipedia). Students are already using Wikipedia for their classes, why not take advantage of that inclination? Especially since Wikipedia itself is a bad model for textbooks, as it focuses on details rather than the big picture.

Incoherent, generic development without a view of the overall scope tends to result in "learning in the small" and developing specific skills rather than getting the big picture. So development needs something to guide that big picture.

The textbook is one of the things that can help bring emphasis to the big picture, providing coherence and guidance among the resources. It can deliver a structure map, give guiding themes either implicitly or explicitly. Explicit themes could use icons to tell users where in the theme a given resource falls.

Development should be a closed-editor-list Wiki, with perhaps 25-50 allowed editors/content creators who are all on board with the big picture. By splitting the load, edition churn is lessened and no one burns out. "Let a thousand flowers bloom." [I hope Joe wasn't being ironic here, given the historical background of the Thousand Flowers incident in Maoist China...the flowers were encouraged to bloom and then rounded up as traitors.] Diversity can come back to textbooks, providing alternate structures for different users.

The analogy of turning textbook development into something more like Garage Band (a home music production program) and less like Amazon.com was presented.

http://www.physics.umd.edu/perg/papers/redish/RedishWither.pdf for more details. Wither or Whither, change or die. And the paper-only holdouts will eventually die, replaced by people who grew up with e-texts.

Panel Discussion: The moderator will be identified by M, Suzanne Brahmia by D, Stuart Johnson by C, Elisha Huggins by B and Edward Redish by A. Audience members whose names I didn't know will be identified by Q, otherwise they'll be identified by name or some other signifier. And if you're wondering why the lettering of panelists is "backwards" it's because I decided to go A-D from left to right before everyone had sat down, and they just happened to line up in reverse talk order.

Q: [Not a question, just some comments] High schools have to use the textbooks that the school board spent money buying, but the students have no desire to use 'em. Commenter promotes a "read the book after lecture if you feel like it, or use online materials" approach, with students encouraged to bring in their own resources. Despite spending patterns, actual usage patterns are shifting to the sort of model advocated by ED02 and ED04.

Q: How do we teach critical reading and thinking? They definitely need it in the real world, so how do we pull this off?

A: There is a large body of literature on how to teach these things.

D: But not specifically in physics, this needs to be worked out for our specific field.

A: "I challenge!" In any case, is it really valuable to teach them to learn to read a badly-written textbook?

D: Maybe not. But we can ask them to read other material, not just the textbook.

M: What about no-reading systems?

A: Weitzmann in Israel has students reading slightly edited current Biology journal articles.

Q: A textbook brings consistency of presentation. How will a Wiki-style text maintain this? Or, even more extreme, how can a "go find other resources" style maintain consistency? Won't inconsistency hurt the students?

Bruce Sherwood: Even with Open Source, you still have gatekeepers, who can enforce consistency and coherency.

A: Yyyyyeah...maybe consistency isn't really a great goal anyway. [Consistency is the hobgoblin of large textbooks.] Variety can be good too...we don't edit Shakespeare and Hemingway to present a consistent style, after all. There's a benefit to learning to deal with a variety of styles and voices, having different viewpoints is good. Maybe a little consistency, but keep it minimal.

Q: Start the teaching of critical reading by having students evaluate each other's reports and assignments, then move to outside texts for evaluation.

Q: On the topic of "intended order" and skipped sections, how can one read critically when big chunks of context are being shuffled or dropped?

D: Well, don't have them read textbooks that you have to hack up like that. Otherwise, explicate the cognitive processes for them as you read out of the book yourself, to set an example. Ask them for summaries and rephrasing, while pointing out the differences between student views and expert views. "Reading Out Loud" method.

B: Sometimes you just can't hold onto all the material in a hardcopy text, resulting in some material being demoted to "dwarf planet" status where it only exists in the e-text version. Does anyone have any better suggestions for cutting out material?

M: The "tyranny of the prerequisite" is a big problem with linear structure. You can't cut out certain things because they're assumed by the next course along the line, even if that means covering them badly.

A: Such linearity isn't really seen in real life. When you hit a gap, you go looking for the answer. So, in a classroom, have students Google or Wiki the topic in question when you hit a gap, and then hold a classroom discussion on it. Create a real critical environment out of the "gap problem", make a bug into a feature. Don't just tell them to look it up in the skipped chapters.

Q: [Comment] Different types of physics material require different reading skills, and if something's over your head the odds are you won't be able to pick it up just by reading. Texts should be organized for random access, like a wiki. Make them pull-oriented, students seek out what they need when they need it, rather than push-oriented where students get what you give them when you give it.

Q: Isn't there a debate between A and D's positions? D says to employ critical reading on the text, while A says to abandon the text entirely, ditching the five pound chunk of wood pulp. But...if we add critical reading skills, what content has to be dropped to make time?

D: Oh, we toss out the normal textbook too, no argument with Redish on that aspect. Use critical reading on the existing material, there's no need to drop anything to make room for a separate critical reading unit, since critical reading is integrated into everything else, just like we integrate teaching mathematical techniques. Model critical reading for them using the course content, just as we model the solution of an integral using physics content.

A: There's going to be an argument here anyway, but with Q. He'd love to teach critical reading separately if there were time, but given appropriate materials it can be taught in an integrated fashion.

Q: We don't have to work alone, perhaps collaborate on critical reading with the English Department. Anyway, it's a bad idea to trust the free market to make a Wiki work properly, is there any research around that might help guide the Wiki editors? The base state of our community seems too incoherent.

A: Use natural selection to build coherence, work is underway as we speak. Mind you, a lot of work is still needed by the community.

Q: [Comment] Sometimes the textbook does a better job than the instructor! Learning to learn the physics requires the ability to figure out ways around any cut material, we should use the texts to teach them how to extract information from bad texts. [I will add that I had a particularly bad lecturer in college, but the one thing he did that really helped me learn was require summaries of the textbook in advance of the topic being covered in class.]

Q: [Comment] Don't forget, students have a range of learning styles. Some can learn better from the text, but others can't. We need to cover as many styles as possible. Teach them to cope with off-style stuff in addition to catering to their preferred style.

Sherwood: A lot of scholarship goes into crafting textbooks, it's not easy for any one person to do this! It's hard to come up with structure in the first place, and faculty rely on the structure provided by the texts. Even strong autonomous learners can benefit from someone else's well thought-out structure.

M: The self-paced course fad has pretty much failed by this time, because students do want that structure.

Q: Will the transition to e-texts open publishers up to taking more risk on niche topics? [There was another part to this question that I didn't catch, possibly about online homework.]

C: Publishers already do niche texts, but the percentage of their output devoted to them probably won't change. The set-up, in terms of authors and layout, is still expensive regardless of how big the market is for the text. However, electronic homework represents a major format change for publishers.

M: Why aren't niche books catching on?

C: Conservative instructors who won't use them.

D: What about Sherwood & Chabay's text?

C: That is running counter to the trend, and catching on.

Sherwood: There are two cases he knows of where authors tried to make major changes, and it wasn't the publisher who nixed it, it was the outside review panel. "I won't use it if I have to change my lecture notes!" attitude.

Q: Sherwood's book is great, but students are only using it as a reference for problem solving examples. Do students really only need a problem solving manual?

M: Well, if you want to pander to their algorithmic approach and just have them churn out standard solutions, sure. [The students only want a problem solving manual, because they think that's the way to succeed, but we don't want that to be the actual way to succeed!]

- Session closes -

FC – PER Faculty Crackerbarrel

Jobs:

The Oregon State positions mentioned at the FacCrack in Seattle have been successfully filled.

A PER position is being added at NCState, under Beichner's aegis in an interdisciplinary center (such centers are springing up a lot lately).

Michigan State is trying to replace Walt Benenson at the Lyman Briggs College (tenure track Assistant Professor) and also adding a senior-level Science Education type for a cross-departmental position.

SIUE is replacing Kim Shaw and hiring a new PER position for the physics department, with strong provost support for STEM.

DePaul in Chicago is looking for a PER-or-experimental physics person next year.

CUNY Queensboro is still trying to fill a position, and there are GRAs available as research assistants.

CSU Chico has filled a position (Leslie Adkins?).

Kennesaw State is trying to create a PER group, and is looking to hire an extra PER person in 3 years.

Kansas State may be looking for a new postdoc for the PATHWAY program.

Colorado School of Mines is expecting to expand their PER program soon.

Toronto now has a 2-person PER group, and hopes to expand.

Tenure:

Tenures are going through pretty well at this point, very few denials or even deferments. Odds are, if you got hired for a tenure-track PER job, you're getting tenure.

Publication Issues:

Over the course of the recent editorial change at AJP, the entire backlog was cleared out (in the good way, not in the "throw 'em in the can" way). Karen Cummings is now acting as a first level editor, and works well with the top-level editor. However, there's currently a temporary acting editor, which may slow things down for a while.

PRST-PER is not experiencing backlog problems either, and has a turnaround of about 133 days between submission and publication, while AJP's turnaround is about twice that just for acceptance and then more for the print delay. Both have long tails, however, where a few unlucky papers just take forever to get through the process.

Once your paper is published, it's okay to also post a copy to your webpage, and use of arxiv.org is encouraged by both journals.

AJP is currently getting about 20 manuscripts a year, with 21% acceptance rate and 54% rejection (the remaining manuscripts are either withdrawn or are still in the pipeline). PRST-PER gets about 50 manuscripts a year, 33% acceptance and 47% rejection. The page rates at PRST-PER can be waived in hardship cases, and there's easy-to-apply-for "scholarships" to cover page costs (the fund is largely untouched at this point). TPT has about 33% acceptance rate, other details not known by those present.

The Journal of the Learning Sciences has been "infested" by PER at the editorial level, but Karen recommends sticking with AJP or PRST-PER, to keep it as easy as possible for people to find our work. Neither pipeline is jammed, so you don't really need to go to outside journals right now.

When tenure committees ask about journals, they really want to know acceptance rates and "impact factor". Not all journals have an impact factor calculated, but AJP and PRST-PER both have factors of about 3. By contrast, Science has an impact factor of 26. The PERC Proceedings don't have an impact factor, but may still count for tenure considerations (ask your committee).

Funding issues:

There just aren't enough funding opportunities. There are rumors that the NSF is trying to hide the fact they don't have enough funds to support all the tenure-seeking faculty that they've given startups to.

It's a good idea to seek out private funding, such as Hewlett Packard's laptop grants. Also, check out the Department of Education...NSF isn't the only public sector cashbox out there, after all.

A lot of existing funding contains requirements for outreach (i.e. museums and national labs) or "broader impact" (everyone else these days), you can get money by helping someone else spend their grants.

The Research Corporation may be getting a little more PER-friendly, mainly supporting young faculty and small schools.

Misc:

Karen Cummings needs a replacement editor for the spring edition of the APS Forum on Education.

Brian Foster will be running the FacCrack at Baltimore.

FG – Laboratories

FG01: Recurrent Studies: Physics Labs with Flavor – Mikhail Agrest, College of Charleston, agrestm@cofc.edu

[This was largely a collection of slides detailing every single lab used in a course, with some opening remarks to bind it together. Pedagogically...questionable.]

You need to excite students as well as electrons, to get them to be bright. Excited students are the most important thing. [The "excited" students in the sample pictures looked like they were being sarcastically excited, though. Stagy in general.]

Students use the results of one part of the lab to make predictions, then test the predictions. They do a "forward" study, then a "backwards" study, then compare. For instance, from the voltage of one element of a circuit, find the total voltage of the battery. Then, given a new battery voltage, find the voltage across one element. This is the recurrence of the title. If the forward part is done badly, then the backwards part can't be well-predicted. Students may, if they wish, be graded solely on how small their experimental error is.

FG02: Understanding Difficulties with Forces through Research-Based Labs – Sergio Flores, University of Ciudad Juarez, sergiflo@hotmail.com

Cancelled.

FG03: Rocketry Across the Curriculum at Mitchell Community College – Doug Knight, Mitchell Community College (North Carolina), dknight@mitchellcc.edu

Rocketry is used in most physics courses at MCC, as well as in clubs and outreach efforts. Most actual launches happen in Midland NC. In each course, there's a class project to launch something, as well as various labs that go into different aspects of the material.

In the calculus-based introductory course, the class project is to make and fly scratch-built (no kits) rockets using G motors (the higher the letter, the bigger the engine). The rockets are designed in CAD programs and submitted for review before actual construction. The performance is predicted, and they must launch an egg into the air and return it safely to the ground. Labs include measuring the impulse on Estes rocket cars, and doing motion-capture ballistics labs using pneumatic "stomp rockets". They compare in vacuo ideal graphs to the actual motion capture and figure out air resistance.

The algebra-based course uses somewhat smaller F motors in a somewhat simplified version of the class project, and slimmed down versions of the calculus-based lab projects. 80-90 percent success rate on the rockets.

Technical Physics, a class for majors in the "technical college" part of the school, focuses more on working with the hands. An E motor is used for a simplified launch project, but they make a test stand to measure the total impulse put out by a motor, testing a simple black powder motor of the kind you can buy in the model kit aisle at Wal-Mart. Lighter on the theory, heavier on the practice.

The conceptual physics class doesn't use rocket motors at all, instead has students use a water rocket (fill 2-liter bottle partly with water, pump it up with air, release). They still had to protect an egg in launch. They did a lab in which a balloon on a string (connected via a straw taped to the balloon) flies along to demonstrate Newton's 3rd Law. The male students tended to talk a bigger game, but the female students consistently win.

The MCC Projectile Society builds trebuchets and high powered rockets using K motors (about the size of a small baseball bat). This is kind of expensive, even with the motors being refillable. The rockets can reach Mach 1.1.

For outreach, MCC participates in the TARC rocketry challenge (rocketcontest.org) and a NASA funded student launch initiative.

Email the speaker for regulations regarding use of various chemical propellants.

FG04: Investigation of the Magnetic Field in Helmholtz Coils – Stephen Luzader, Frostburg State University, sluzader@frostburg.edu

FSU has three different types of Bainbridge-style setups for the e/m experiment, and while two work pretty well, one did not, as reported at a previous meeting. A number of experiments were carried out to try to figure out why it wasn't working, and eventually they discovered they'd misread the manual and were using the wrong value for the coil radius. Oops.

[The talk focused almost entirely on the nitty gritty of all the unnecessary experiments they did. No mention was made of how much, if any, student involvement there was in the diagnostic process. It strikes me that this would be a decent topic for a student project, but the approach in the talk was just equipment maintenance.]

FG05: Discrepancy in Trajectory By Anthropometry – Saami Shaibani, IMAAS, shaibami@imaas.org (in Virginia)

The human body is massively articulated, a complete description requires over a thousand data points for proper simulation. However, a lot can be done with a simplified model that uses 17 segments and 16 joints (head, neck, upper torso, center torso, pelvis, upper arms, lower arms, hands, thighs, shins, feet) like an action figure. Consider the biomedical parameters for each element, such as the mass, axes of motion, etc. This model ignores details like fingers, toes, jaw, etc.

In Injury Mechanisms Analysis, each segment is assigned a position vector, a rotation vector, velocity vector, angular velocity vector, accelerations, etc. Location, rotation, speed, orientation, etc. This is used in analysis of security camera footage for crime scene investigation.

"CSI in Physics 101" lab interviewed a medical examiner to get real data from a convenience store shooting, then set up a living person to replicate the actions. Anthropometry is VERY tied to specific dimensions, though, so you can't just grab a student who's kinda close. Then, they consider how motion affects anatomical landmarks, like the eyes or heart.

Use of anthropometry was able to prove, for instance, that a claimed "accidental" shooting of a clerk was no accident.

FG06: Going WILD While Staying Sane – Dave Van Domelen, Kansas State University, dvandom@phys.ksu.edu

"WILD" stands for Wandering Interactive Laboratory Demonstrations, and is a project funded by Hewlett Packard. KSU got a batch of tablet PCs, ancillary equipment and a small monetary grant to get sensors and software. The goal is to have students get away from the lab table occasionally, and even go outside to measure things found there.

This talk is mainly about practical considerations, as the pedagogy has not been significantly altered in the test case course yet, just the procedure. The course under revision is a conceptual physics lab that is a 1-credit standalone class.

The first part of the talk covered some pitfalls that were either anticipated ahead of time or discovered during the initial semester of implementation, such as "how do you do a 'wandering' lab when the weather is uncooperative?", security issues, and practical matters such as software and making sure the batteries don't go flat. KSU's solutions, which may not be broadly usable, as well as some suggestions for those seeking to run similar labs using something like the "$100 Laptop Project" were presented.

To deal with weather issues, any lab with a "wandering" component was written to have both indoor and outdoor activities, with instructors picking from a list of possibilities based on what conditions allowed. For software, OpenOffice (http://www.openoffice.org) was used in place of the Office suite, and while it presented some problems, most students were able to use it properly. DataStudio software was used along with PASCO sensors for data acquisition and analysis, but the audience was warned that PASCO has no current plans to issue a Linux-based version of that software (the $100 Laptop Program and similar commercial products like the Asus EEE PC save money by running Linux). TA training was emphasized as a solution to a number of issues, especially the battery problem (i.e. make sure the TAs check the laptops at the end of class and plug in any that the students failed to).

The second part of the talk presented the results of a short opinion survey performed at the end of Spring 2007. Student reactions were generally positive, and concerns regarding the potential unfamiliarity of the software turned out to be unwarranted.

FG07: Labs a la PER! – Taoufik Nadji, Interlochen, nadjit@interlochen.org (boarding school)

Students at Interlochen are largely interested in careers in art, and they dreaded the physics labs. The original cookbook labs worked poorly, and an attempt at homegrown reform was moderately successful but required a disproportionate amount of effort and resources to do what was essentially reinventing the wheel.

The instructor realized, "the students own the learning, not me," and was inspired to look into PER results. The new labs used Interactive Lecture Demonstrations, Real-Time Physics and UW Tutorials. The lecture portion added clicker tasks and the working out of theory with peer instruction. The course was linked into the history of physics for context. Vernier equipment was used, and LoggerPro software was acquired for motion analysis.

Since students don't really know how to write a good lab report, they were assigned stages instead, with feedback at each step. For instance, for a given experiment they might only have to report the objectives, or outline the procedures. A few "full" lab reports would let them sum up the process (every sixth lab had a full report).

A ten-element rubric was used for grading the full report, with the partial reports being aimed at one or two rubric elements at a time. Students would be informed in advance which elements they'd be responsible for on a given lab.

The results were pretty good.

The historical part of the course did things like have them go back over original works (like "Huygen's Theorem #13") and reprove old theories in modern notation.

Students seemed to like the new labs (and certainly weren't dreading them), and got decent FCI post-test scores. There was no pretest given, but assuming typical high school scores as a pre-test yielded gains of about 0.3.

FG08: Experimental Marshmallow Physics: Young's Modulus and the Speed of Sound – Kenneth Pestka, Dalton State College, kpestka@daltonstate.edu

This started with a simple question: could the Sta-Puft Marshmallow Man from Ghostbusters have supported his own weight? [Assuming a lack of mystical ectoplasm, of course...and I am amused to find that "ectoplasm" is in Word's autocorrect.]

Students looked at elasticity and Young's Modulus. A stress/strain graph is linear up until the yield strength, then turns over as the material breaks. A typical Young's Modulus experiment breaks wires, with very small displacements and sensitive measurements that are abstracted by their smallness. A more hands-on and intuitive experiment was desired for the topic.

Marshmallows to the rescue, with an experiment on compressile strength rather than tensile. Students plotted stress (F/A) versus strain (ΔL/L) while crushing a marshmallow in a special crushing apparatus. This had a good linear range and gave a result of Y = 29,000 Pa, give or take.

Marshmallows are, however, viscoelastic, with "creep" that causes Y to change over time. You need a long relaxation period between trials to reset the "memory" of the marshmallow, and conditioning exercises to get it into a short-term-friendly mode. Staleness of the marshmallows also matters a lot.

Still, even without a lot of care taken in conditioning, the experiment works pretty well. And it's robust even under some serious violation of assumptions (i.e. the actual volume of the marshmallow isn't changing much, since it bulges out at the sides, so it's less compression and more viscous flow).

Looking at the speed of sound (the square root of Y over density), and given a density of 320 kg/m^3, they got a speed of sound in marshmallows of about 10 m/s. There was no way to experimentally verify this, but since the speed of sound in rubber is about 50 m/s, it seems reasonable.

It's a cheap lab to do, feeds the students, and is fun. Future plans include doing something horrible to Peeps in the name of science.

PERC Bridging Session: Discipline-Based Education Research in other Disciplines (Invited)

[This was running concurrently with FG, and I kind of had to be at FG. Unfortunately, by the time FG let out, not only was the bridging session standing room only (and very little of that), the format had been rearranged so I missed most of the context anyway. As a result, these notes are actually borrowed from Dyan McBride and Fran Mateycik, graduate students at KSU.]

PS01: Making Science Learning Inviting: The Learning By Design Theory – Janet Kolodner, Georgia Tech, jlk@cc.gatech.edu

[title changed from the version in the program]

[Dyan's Notes]

In Case-Based Reasoning, people learn from experiences productively. Knowledge is seen as adding and/or reorganizing those experiences, dictated by both long-term and short-term goals. The experiences undergo interpretation in light of those goals, and learning happens in the process of achieving the short term goals and preparing long term goals.

An example was presented of middle school students presented with design challenges. The short term goal was to meet the challenge itself, but the long term goal involved practicing the tools and methods involved in performing the challenge.

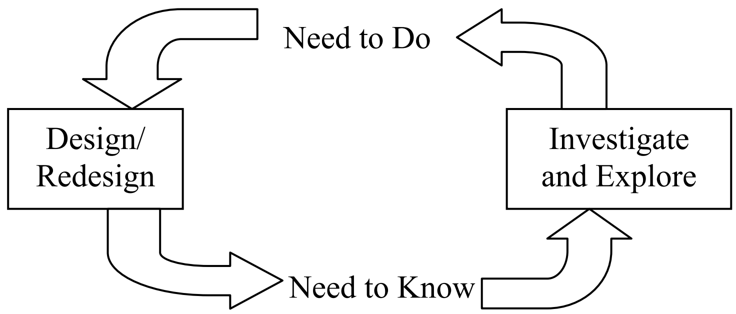
[Fran's Notes]

We want to promote self-explanation a la Chi (2000). How do we get and keep learners' attention? How do we get them to see importance in learning? How do we get them to articulate their thoughts and to brainstorm?

Case-Based Reasoning is a computational model of learning from experience. We extend our knowledge by adding new cases to the "database" in our brains, and also by re-explaining, re-encoding and re-indexing old cases, thereby extracting new generalizations.

CBR users go into a situation with a goal, and interpret experiences to shed light on how to achieve that goal. They achieve short term goals in order to get into a position to tackle a long term goal.

Problem-based learning matches case-based reasoning by giving students design challenges. Skills and practices are repeated, and students realize how valuable they are. The following cycle results:



The Design/Redesign block includes: understand the challenge, present and share, analyze and explain, construct and test, plan the design. The Investigate and Explore block includes: present and share, clarify, design investigation, conduct investigation, and analyze results.

An important part of this process is a "poster session" presentation and sharing with one another (note that "present and share" is on both sides). This is a critical classroom practice.

Learners have experiences worth learning from. We need to recognize the roles of individuals in the community, gradually introduce scientific practices, and provide cognitive and social affordances.

PS02: Problem Solving and Learning for Physics Education – Brian Ross, University of Illinois at Urbana-Champaign, bhross@uiuc.edu

[Dyan's notes]

How cognitive science might be useful for people interested in how students learn physics. A list of literature recommendations will be found in the PERC Proceedings.

Our brains have a number of specialized systems, including memory and reasoning. Memory can be broken up into Declarative Memory (knowing what) and Procedural Memory (knowing how), while reasoning can be split into Deliberative Reasoning (slow, such as the application of rules) and Heuristic Reasoning (associative, quick recognition).

These systems bring implications for learning. We need to transfer appropriate processing, not just Declarative but also Procedural Memory must be involved (Morris, Bransford, Franks 1977). Transfer depends on the overlap between Declarative and Procedural...understanding the solution versus just generating one. Problem solving is a procedural routine, not an end result. Conceptual understanding requires explicit reasoning.

When considering content, you need to realize that what you do depends on what you do it to, and that abstractions are difficult. [A]

Novices tend to focus on surface features (Chi 1981 and others), and giving examples of principles can have an influence on them. It can force unwanted comparisons and generalizations, making them think all situations with a given deep structure also must follow particular surface features. This is the "Authority of the Specific." [B]

Experts, on the other hand, tend to see the structural connections between things, but are still tuned into the surface features because of the concepts involved. The usual content can slow things down, but employing new contexts brings the danger of cueing inappropriate things.

So, how is prior experience used? Analogy and categorization are closely related, and one can examine cues, knowledge and output.

Novices cue on superficial things, with little effect from the deeper structure. They know some examples, formulae and categories, and when retrieving information tend to go for specific examples or formulae. Their problem solving techniques are formula-based and involve working backwards from the desired solution. They're also likely to draw analogies to previously solved problems.

Experts cue more on the structure, while their knowledge base is more likely to involve a few examples and a lot of general categories. They retrieve information based on these general categories and then make it fit the situation, problem-solving by means of category-based plans and procedures, only using analogy if the situation is unusual.

Additional cues can be provided by multiple representations, explanations, and conceptual analyses.[C]

Research methods in this area should include both qualitative and quantitative studies, often using qualitative studies to develop the protocols used in the larger studies. The independent variables should be "what matters", such as looking at superficial cues versus deep structure. The dependent variables will be the results of tests and finer analyses.

[Fran's notes mostly duplicate Dyan's, I'm just putting in a few things she wrote that Dyan didn't. The little superscripts up in Dyan's notes indicate where one of Fran's insertions goes.]

The first rule of style is to have something to say. The second rule of style is to say only one thing at a time when you have more than one thing to say.

A. Most things are very contextual. Young students don't learn that math is not context-based. While 5+2 fingers are 7 fingers, 5+2 markers might be 6 markers. What this means is that in the case of markers, a student might count two of the same color as one, so giving a red and green to someone with red, blue, yellow, orange and purple leaves them with six "effective" markers.

B. Or "Priority of the Specific". Students correlate the specifics of two situations rather than the abstract generalizations and become confused.

C. Newell and Simon, 1972.

PS03: Naïve Physics/Savvy Science: Casual Learning in Very Young Children...and the Rest of Us – Laura Schulz, Massachusetts Institute of Technology, lschulz@mit.edu

[Dyan's notes]

Children have shown that they can understand causal relationships without understanding the mechanical relationships involved. Most of them don't even realize how much they actually know. This also seems to extend to adults.

[Fran's notes]

Young children are resilient, and learn incredible amounts, but then at some point learning becomes incredibly difficult.

Causal discovery is great, but often comes with bad experimental design. So children may pick up the wrong conclusions from a badly controlled situation.

The processes that make us smart also hold us back, making it hard to revise what we know. We lock onto a piece of understanding quickly, but sometimes too firmly, especially when it's mistaken.

Children can understand causal but not mechanical relationships.

CP: PER Conference Poster Session

[Since I missed the bridging session, I tried to take some notes at the posters. I did not attempt to cover all of the posters, however, or even a significant fraction, nor did I necessarily take complete notes on those that I do cover. I'll list all authors, but only the email address of the lead author.]

CP20: Spending Time on Design: Does It Hurt Physics Learning? – Eugenia Etkina, Alan Van Heuvelen, Anna Karelina, Ruibal Villasenor, David Rosengrant, Cindy Hmelo-Silver, Rutgers University, Etkina@rci.rutgers.edu

Expanded on the information given in talks AG08-AG10. [Okay, my only notes on this one were, "This is the sort of thing I wanted to do for my PhD thesis work, but lacked the tools at the time." ☺]

CP31: Towards a Dynamic Model of Expert and Novice Categories in Physics – Edward Redish, Ayush Gupta, David Hammer, University of Maryland at College Park, redish@umd.edu

The categories of the title are set up as follows, the three being mutually exclusive:

1) Matter (a dog, an electron)
2) Process (heat, current)
3) Mental State (happiness)

Thus, if you accept that these categories are mutually exclusive, happiness is **not** a warm puppy.

CP34: Student Categorization of Problems – an Extension – Kathleen Harper, Zachary Hite, Richard Freuler, John Demel, the Ohio State University, harper.217@osu.edu

The students in Physics 131I (engineering honors program that takes physics, calculus and engineering graphics courses and makes sure they're all working in synch) were put through an ISLE curriculum and then given a Chi-style sorting task. This was intended merely to establish a baseline for some other work, but the students sorted in a surprisingly expert-like way after only a single quarter of intervention. Future work is planned to delve deeper into this and compare progress at the beginning of Physics 131I to that at the end.

CP57: Learning To Think Like Scientists with the PET Curriculum – Valerie Otero, Kara Gray, University of Colorado at Boulder, valerie.otero@colorado.edu

This project worked in the "Physical Science and Everyday Thinking" (PSET) project, with students "learning about learning" using the same types of activities as they used to learn about content. This improved learning about their own learning (reflection), about scientists' learning (the nature of science) and about the learning of others (i.e. how the kids they'd be teaching learn).

CLASS results showed gains in all categories, ranging from 4% to 17%.

CP71: Symbols: Weapons of Math Destruction – Eugene Torigoe, Gary Gladding, University of Illinois at Urbana-Champaign, torigoe@uiuc.edu

The top quartile in the study did equally well on numeric and symbolic problems, while the bottom quartile showed a large spread, with symbolic problems being significantly harder for them than numeric ones for many problems.

Symbolic problem solving increases cognitive load, and hides things such as units [not that students write those down much themselves when doing numeric problems]. The symbolic problem does not explicitly differentiate the states of each quantity (i.e. constant/variable/parameter).

[I am constantly harping on my students to solve problems symbolically before putting in the numerical values from the givens. They don't like to do this, and it's clearly harder for them, but this makes it a little more obvious why it's hard. For instance, it's easier for them to see what they need to solve for once it's the only thing left that's not a number.]

Thursday, August 2, 2007 - PER Conference

WS-B: Workshop: Physics Learning in the Context of Scaffolded Diagnostic Tests

Presenters: Edit Yerushalmi, Weizmann Institute of Science, ntedit@wisemail.weizmann.ac.il; Chandralekha Singh, University of Pittsburgh

Assessment is traditionally focused on the product, not the process, so it's hard to encourage problem-solving skills such as reflection. Alternative assessment is therefore necessary, such as the kind found in Inside the Black Box by Black and Wiliam (1998).

Feedback should be operational, explicit and targeted. Students need training in self-assessment to make this work.

A common feedback cycle involves returning graded work to students and letting them earn points for correcting their errors, explaining what was wrong, etc. However, once the class size gets large enough, the logistics of this become troublesome. Also, correcting the errors isn't the same as explaining why they were errors in the first place, but students fail to see the difference. They'd rather just spit out more product by correcting the errors, rather than think about the process that led them to the original error.

Recitation quizzes were used to teach reflective self-assessment. In the introductory recitation, students were given information on the desired product format and some guidance on process. Then, in a given recitation they would spend 20 minutes solving a context-rich problem. In the next meeting of recitation, they would spend 10-15 minutes self-diagnosing that quiz.

All students were given an un-annotated and ungraded copy of their original quiz, and had to first figure out if they got it right or not. They would also have to explain their errors, if any, and got the same credit for this that they would get for a quiz.

There were four groups in the study. Some were given no guidance in their reflection, others were given rubrics, and sample solutions of varying levels of detail were given to some of the groups. This exercise was primarily focused on promoting reflection on the process of problem solving. It should be considered a minor intervention in an otherwise traditional course.

Both the introductory material (which all got) and the diagnostic rubric (which only some students got) reflected an explicit problem solving strategy. Students had to work alone on this.

Workshop participants were asked to suggest avenues of study regarding this intervention. Contributions included: follow longitudinal student progress, compare this to "official" grading, look for affective effects and epistemological changes, examine the balance of errors in physics principles versus mathematical errors in the student reflections and see which type the students found more important, see if this strategy transferred to general practice, look at the effects of allowing group work, and see if it worked at all.

So, what do the students actually do? Is it helpful? How could it be improved? The variables involved that can be controlled by researchers include:

- The instructions on how to perform the problem solving and reflection

- The resources available (solution format, textbook, etc)
- The feedback provided on the self-diagnoses (this is very time-consuming)

The control group engaged in no self-diagnosis, and three experimental groups with differing instructions and feedback. The intrinsic motivation, "This will help you learn" was given explicitly to all three treatment groups. The quizzes were deliberately made hard enough that no one got them 100% correct on the first try, so there were always errors to discuss and reflect upon.

There were two quizzes in which the students had to perform the reflection exercises, plus a near transfer and a far transfer problem on the final exam related to the quiz problems. A separate rubric was created for instructors to use in rating the student work.

[Unfortunately, as an actual workshop, this worked incredibly poorly. The room was at about twice the recommended capacity, too loud for discussions, and it was just too crowded for people to want to move around freely like the organizers wanted us to. Also, in practice, we didn't seem to agree at all on how to use the instructor rubrics, and certainly didn't agree with the sample scoring.]

Aside: Not only do context-rich problems lack the sorts of cues students expect to be able to use for their pattern-matching strategies, they often have deliberately misleading cues.

[The group I was in found that the instructor scoring of self-assessment was remarkably lenient and inconsistent. For instance, a student who had made deep and widespread conceptual errors in their work said in self-assessment only that they'd made a sign error...and got scored 8/10 for reflection. This is hardly going to promote good self-assessment, is it?]

Even better students miss their errors when they don't have a full problem description or problem construction. And due to some mushy language use on the part of students (perhaps deliberately mushy), it's hard for graders to tell if the students recognize their own conceptual problems, or simply think it was an "Oops, I grabbed the wrong equation" sort of error requiring better memorization of the equation sheet.

Overall, it was determined during the research project that learning did occur during the self-diagnosis process. There was a good correlation between the quality of self-diagnosis and performance on the final exams' transfer problems. There was no correlation between quiz scores and self-diagnosis, but the quiz grading was iffy anyway.

Invited Talks

IT01: Cognitive Science: The Science of the (Nearly) Obvious – Bruce Sherin, Northwestern University, bsherin@northwestern.edu

Joked about paying the price for coming up with a too-clever title before writing the talk, and then finding he might not be able to cash the check that this title wrote.

When looking for a "grand theory" of physics cognition and learning, it's necessary to track changing knowledge structures from birth to death, or at least from introductory physics through the attainment of expertise. It's also necessary to have an appropriately reductive elemental description of what elements exist in responses, taking the elements from case-by-case to construct-from-general. A grand theory can constrain the "best" elemental account.

So far in PER, we've mainly focused in a tight range around introductory physics, so we haven't made a lot of progress towards what's needed for a grand theory. When experts are studied, they're often given trivial-to-them problems for comparison to how novices approach the same tasks, but we need to also look at how experts approach things that they find challenging. Other fields have been looking at the childhood angle, so we can wait and take advantage of their results later.

Even if we really only care about improving introductory physics, having a grand theory would help with it. Our actual knowledge about the physical world and mathematics progresses all the way from birth to death, and we need to be able to fit our piece of the instruction puzzle into place better by finding the context within the lifetime path. We can look to other fields for information outside of our focus, if need be.

For an example of "outside our focus" useful information, studies have shown that infants stare longer at "interesting" things, where interesting is often defined as "impossible as far as the baby is concerned". This indicates that even infants have probeable resources.

We should refocus our attention on what's **nearly obvious** to us. It's so basic that we ignore it, we need to explicate the implicit.

For instance, knowledge of the English language alphabet is chunked into the clumps ABCD – EFG – HIJK – LMNOP – QRSTUV – WXYZ. This is something that we don't teach explicitly (although it does map onto things like the alphabet song fairly well), but it seems to come up everywhere that this alphabet is used. These tacit chunks are one "grain size" down from the named structure "alphabet" but a size up from the named element "letter".

Another example would be clusters of equations and symbols. Students working on problems with whiteboards in a third semester introductory calculus-based physics course were studied in this case.

One group working on friction decided that the equation they were using was wrong, and that the coefficient of kinetic friction had to depend on the mass, such that bigger objects would be harder to slow down (i.e. a "more is harder to change" p-prim). Working with the symbolic form $\mu_k = \mu_{k0} + cm^n$, they settled on n=-1.

"Symbolic form" is a scheme that uses a conceptual schema merged with a symbolic template. In other words, an idea plus how to write it as an equation. The idea the student had was that the coefficient of kinetic friction really had two parts to it, one of which depended on mass. Idea one (two parts) had a conceptual schema that was "a whole is composed of two or more parts" while the template was [] + [] + ... The second idea (bigger mass should have less friction) had the conceptual schema "Prop-" (negative proportionality) and the template []/x.

There exist 21 symbolic forms in 6 clusters. For instance, the "competing terms" cluster has symbolic forms of "balancing" (a = b) and "competition" (a±b±c). The "terms are amounts" cluster has symbolic forms of "parts of whole" ([] + [] + []) and "base plus change" ([] + Δ).

diSessa's "sense-of-mechanism" is a subsystem of commonsense physics knowledge, made of p-prims (like "increased resistance = work harder").

Another example would be patterns found in arithmetic word problems, based on work in 1982-3 by several groups. The symbolic forms found here were "change" (A + Δ → A'), "combine" (A + B → C), "equalization" (A + ? → B) and "compare" (A – B → C).

Each symbolic form cluster can be tracked out to other things like p-prims, arithmetic patterns, etc. Symbolic forms may be the desired element for a grand theory of physics cognition and learning.

[Q&A]

"What is this useful for?" Well, while there's some hints here and there right now, it's mainly at the "of what use is a newborn babe?" stage. Redish pointed out that Elby's p-prim-based paired questions were an immediate benefit, however.

"Is there a type of expert best to study in order to get useful information for introductory physics insights?" Well, we want to look at ALL experts, ideally, those who have compacted knowledge and those with more loose-packed understanding. But we really want to see the experts at the point where they have to unpack their knowledge and extend their boundaries. [I suppose that means the best kind of expert is "confused".]

IT02: Facilitating Conceptual Learning Through Analogy and Explanation – Timothy Nokes, University of Pittsburgh, nokes@pitt.edu

How can we facilitate deep learning, as opposed to equation-manipulation skills? We should take cues from existing "expertise research" that looks at how people develop deep structural knowledge, use forward-working strategies and engage in transfer. Those with expertise understand relationships between their knowledge elements, not just the elements themselves.

Novices use near transfer from prior sample problems, and prefer this strategy even when given the relevant principles for a problem. They don't get how the principle links to the examples.

There are multiple pathways to a conceptual understanding of these links, including explanations and analogies or comparisons.

Explanation can facilitate learning and transfer. It helps to identify sub-goals and structure, and can lead to specializing (going from principle to example, deductive) and generalizing (going from examples to principle, inductive). Chi (2000) identified self-explanation as a tool that lets people generate inferences from the text to prior knowledge and repair mental models.

Specializing facilitates specific understanding, the construction of mental models, and performance in near transfer tasks. Generalizing facilitates abstraction, the construction of schema, and wide application (far transfer).

In an experiment, students were presented with either specializing or generalizing exercises on a particular topic (permutation via combinatorics), then tested for near and far transfer. The prediction was that, for both accuracy and speed, those given specialization exercises would perform better on near transfer, while those given generalization exercises would do better on mid or far transfer. The results were mixed, with generalizers being more accurate on far transfer and specializers being faster on near transfer, but no difference in the other two combinations (equal accuracy on near transfer, equal speed on far transfer).

Now for the "analogy" part of the talk.

Comparisons can facilitate acquisitions of schema. They do this better when there's a greater number of comparisons, a greater variety of comparisons, a clearer focus on commonalities, and a minimized cognitive load. But how about the type of comparison?

Two types of comparison categories were presented: Near Miss (NM) and Surface-Different (SD). NM uses the same principle in both cases, with some illustrative minor change that shows how the principle works. SD uses the same principle but has significantly different surface features in the two cases. For instance, let's say that situation A is a problem where N knights choose mounts from among M horses, N<M. A NM change would give situation B where the horses choose the knights instead but N<M sill holds, changing how the idea works. An SD situation C would have N children picking from among M puppies, N<M.

NM focuses on how variables are substantiated, where SD focuses on cross-context situations.

In this experiment, two groups were given the same problem as a worked example. Then half were given two NM exercises while the other half were given two SD exercises to work. This was repeated for a new worked example, swapping the NM and SD groups, so that everyone ended up doing two of each kind of analogous problem.

NM was predicted to facilitate principle use, while SD should have facilitated knowing which principle to use in the first place. In the experiment, students rated as "poor" in an unrelated learning test did perform according to this prediction. However, those rated as "good" on the test experienced a ceiling effect, doing well on all problems. Still, it seems that SD analogies are better at helping students learn which principle to pick for a given problem (learn to separate surface features from the underlying structure) while NM analogies improved their ability to apply the principles correctly (learn the ins and outs of the principle).

The "Explanations and Analogies" curriculum is currently being tested at the Pittsburgh Science of Learning Center.

[Q&A]

Analogies were made as direct and obvious as feasible, due to the novice nature of the students.

When asked how we know if students see SD and NM as instructors do, the reply was that we don't. That's part of what the study hopes to accomplish, though.

IT03: Learning Is Not The Same Thing As Problem Solving – Dan Schwartz, Stanford University, danls@stanford.edu (Stanford School of Education)

[Original title: Socializing Learning and Transfer. He ran into the same problem as IT01, but decided he didn't have to stick to the original title.]

At a previous PERC, it was pointed out how paleotransfer isn't found in experiments, even typical horizontal transfer. [He didn't use these terms, refer back to CB03 for definitions. Throughout this talk's summary, I will tend to use terms introduced in CB03.]

The existing Sequential Problem Solving test pretty much tests for the Routine Expert. Meanwhile, someone who engages in purely vertical transfer could also be considered the "Annoying Novice" who's always brainstorming and blueskying without being able to actually put any of these ideas into practice. What we really want is the Adaptive Expert, who balances horizontal and vertical transfer.

The Preparation for Future Learning (PFL) test looks to see if people are on the Optimal Adaptability Corridor (OAC).

Embedded worked examples don't really help purely horizontal learners, but they help those on the OAC a lot.

The contributions cognitive science can offer PER are many: multidisciplinary nature, verbal protocols, discourse analysis, careful descriptions of problem solving via examination of the verbal stream, and formal/executable models. However, cognitive science brings some risks along as well: motivation is rarely considered (i.e. people learn because learning is what brains do, not because of any motivation), problem solving models can be overused to explain learning (just because you can solve a problem doesn't mean you understand the material), and practitioners often confuse the map with the territory (models of thinking are not themselves thought).

Talking the talk may only enable part of walking the walk, in any case.

Two examples will now be considered from cognitive science: the symbolic distance effect and the benefits to learning of social interactions.

Symbolic Distance Effect (SDE)

Look at nonverbal outcomes of verbal rules, and ask, "How do we build new knowledge on the foundation of the old?" Can people start with rules and then move to experiential operation? [At this point, a lot of out-of-my-field terminology started getting used too quickly for me to process, so I missed some stuff.]

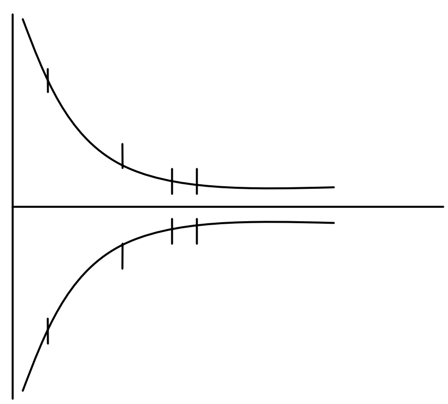
The Symbolic Distance Effect states that the larger the difference between two things, the easier it is to tell them apart, and percent difference matters. The SDE correlates with the activity of a part of the brain called the IPS [Intraparietal Sulcus, if I correctly interpret the results of a Google search on "IPS" and "brain"]. A smaller symbolic distance means more work done by the IPS. The IPS also activates for tasks involving perception, spatial reasoning and subtraction, but not addition.

If you consider the counting number line weighted by SDE, it looks like a logarithmic distribution, with numbers farther apart at low numbers and closer together at higher numbers. Perception in general seems to be logarithmic in nature.

When looking at just negative integers, the same effect was observed, with the same speed and accuracy results based on magnitude. Note, however, that in all cases the difference between fast responses and slow responses was on the order of 0.05s, so you really need mechanical analysis to pick up the effect. Adding a negative symbol consistently required an additional 0.2s for flipping the signs, but the difference between slow and fast remained the same, and bigger magnitude was still slower.

However, when the comparison was made between a positive and a negative integer, the SDE flipped. Now, numbers close together (like +1 and -2) were seen as far, while those far apart (like +8 and -6) got slower near-like effects.

It turns out that placing a boundary between classes creates a different sort of SDE. Distance becomes "distance from the boundary", where in this case the boundary is zero. Rather than just mirroring the logarithmic number line, it becomes necessary to go to two dimensions to find the distance between numbers. [Note: the diagram I copied below seems to contradict some of the near-far stuff, and I think I may have missed a flipover somewhere, because this paragraph and the previous one also seem somewhat contradictory.]



So, the hash marks in the badly reproduced graph above would be 1, 2, 3 and 4 from left to right on top, and -1, -2, -3 and -4 from left to right on bottom. +2 and -2 are closer to each other than +1 and -1.

While staying in-class, 7th graders perform like adults, and younger children are just slower and less accurate, but otherwise follow the same patterns. However, 7th graders don't have the cross-class SDE inversion that adults do. They have different rules perceptualized, and may not see classes the same as adults do.

There is not yet a model describing how children acquire the mental model that as adults causes the SDE inversion.

Social Effects:

Non-verbal processes can regulate verbal learning. But is there a non-verbal, non-problem-solving learning benefit to social interaction? In other words, working with someone can certainly help learning if they provide aid in problem solving or explanations. But does the social aspect alone provide a benefit?

An experiment was set up to tweak the student belief about social interaction, in something akin to a Turing test. Using a virtual reality helmet, students interacted with a computer animated character that was actually generated by a canned "Eliza" style routine (a very primitive fake artificial intelligence) that mostly followed a script. One group was told that this was an "avatar" that represented another student in the study, while the other was told it was an "agent" run by computer. Student skin conductance was measured as a way to determine affect.

In both cases, the students spent some time interacting with the character in order to get comfortable with them, doing things like playing with objects in the virtual environment and engaging in unrelated chit-chat. There were written problems in the environment that both student and agent would read and discuss, and the agent communicated verbally in the initial experiments (a "silent avatar" option was added later, where the "avatar" read along with the student but did not talk). Regardless of whether the student thought the character as avatar or agent, it gave the same advice and made the same statements, controlling for the other potential learning benefits of the interaction.

Students thinking they were dealing with an avatar got better posttest scores, both on problems identical to those solved in the simulation and on transfer problems. The "avatar" group also had higher emotional arousal as measured by skin conductance, and it's a previously known effect that arousal correlates positively with performance [nervous laughter from audience at the double entendre at this point].

The mere belief that social interaction is happening improves learning and raises emotional affect.

Then they asked if the social aspect had to involve interaction, or if mere presence was enough to generate the affective effect. The main rise in skin conductance would happen before the agent even responded, so they added the "silent avatar" who was present and moved around the same as before, but did not speak.

In the silent avatar case, just having someone around who you thought was a person was almost as good as interacting with someone. Silent avatar subjects did as well on factual questions as talkative avatar subjects, but not as well on inference and application (problem solving) questions. Thus, at least for the transmission of factual knowledge, lecture is actually better than reading from the book, if only because there's the social dimension of being in the room with other people. We seem to absorb information better as a group than on our own. However, getting socially involved helps on all levels.

Thus, active engagement brings more social interaction to classes, which leads to higher arousal which leads to better learning.

Implied social interaction, or expected social interaction, was also shown to be effective. So, a student taking notes for someone else not in attendance might actually learn better than a student taking notes for themselves. [In other words, I didn't just type in 60+ pages here for nothing...I got a social-interaction benefit out of it.]

[Q&A]

Moderate arousal is best. Too high is almost as bad as too low.

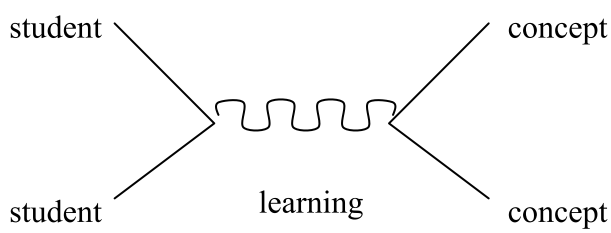
[The scheduled panel discussion was abandoned in favor of giving more time for individual Q&A after each talk.]

Lunch Speakers (Finkelstein and Price)

The "Rising Above the Gathering Storm" report on the education situation (http://www.nap.edu/catalog.php?record_id=11463) has come in and made its recommendations, and Congress has acted, advocating 10,000 new teachers and 10 million minds, recruiting the best and the brightest to avoid a crisis in science education. The time to act is now, and PERC should be involved.

No, not the Physics Education Research Conference...the Physics Education Research **Corporation**.

Tired of being rated behind Latvia on the TIMSS? Come to the PERC! We'll provide all the research you need, and the results to boot! Never worry about slipping below standards again with our new teacher evaluation forms, where students are free to rate you as a great teacher or the greatest teacher. We'll provide you with a full range of impressive representations to include in your report, such as Many Worlds diagrams, or the always useful Feynman Virtual Learning diagram:



But wait, there's more! Having trouble with your grant proposal? We'll provide you our time-tested Grant Proposal Mad Libs™, and you'll never have to sweat over an NSF grant again!

So, come see the PERC for all your government-approved educational needs!

[In case anyone needs to be told this, the entire talk was a spoof. I only add this because Murphy's Law means that if I don't spell it out, someone will think Noah was being serious and send him an angry letter.]

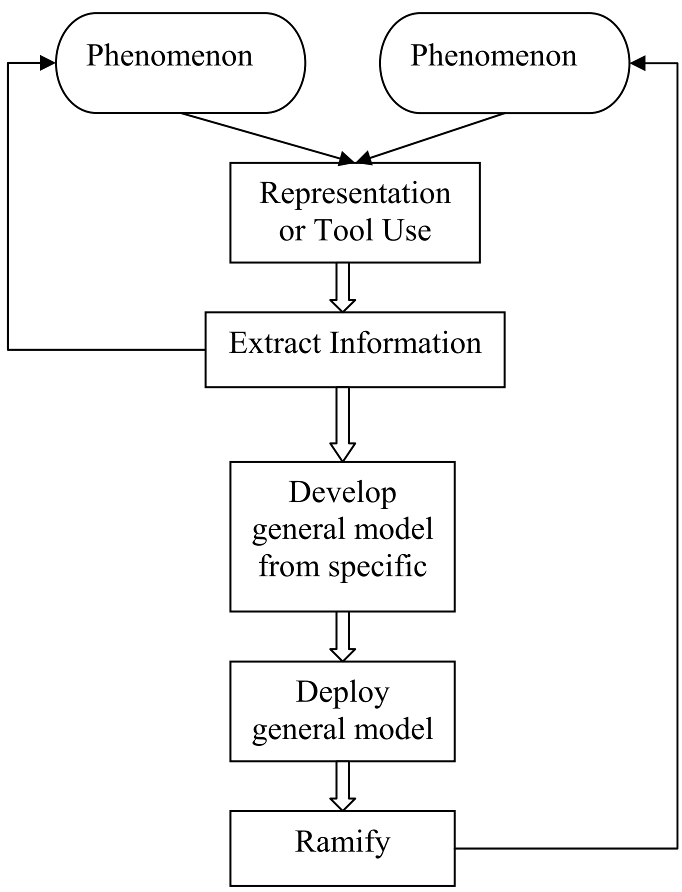
TP-B: A Conversation About Models, Modeling, Representations and Cognitive Science (targeted poster session)

TPB01: Modeling, the Genesis of Models and the Relationship with Cognitive Science – Eric Brewe, Florida International University, eric.brewe@gmail.com

Modeling started as a theory of science, which progresses by creating models that fit the world around us. Models exist in our minds, but must be made communicable to be useful. They're developed using representational tools.

Internal models are ad hoc, and experientially driven. Scientific models are intentionally developed to be broadly used, and are broken up into specified models and general models.

We create specified models first, then build up our general models from assemblies of specified models. Once a general model exists, it can be ramified to spawn new specified models to fit the situation.



Organizing the class around models has had a big impact on the curriculum. Various cases are split up and given to different groups in the class to saturate the field.

[Aside, about this time I was starting to space out a bit, it had been a long week, so my notes on TPB02 and TPB03 may be a bit iffy. I took notes on TPB04 first.]

TPB02: Physical Modeling – Ruth Chabay, North Carolina State University, rwchabay@unity.ncsu.edu

Talking about idealizing or simplifying messy situations [I have a parenthetical note here, "often (something) for students" and I cannot for the life of me figure out what that missing word is...it looks like "duse"] and working with existing models: add to them, change them or refute them. Not always building models from scratch.

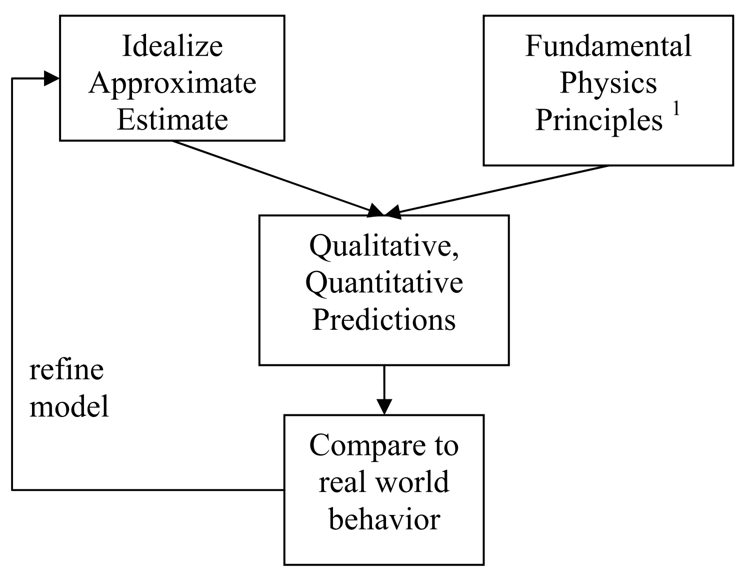
Examples were taken from the Chabay & Sherwood (2007) textbook (pages 56 and 494).

Gordon Aubrecht suggested "Adaptive Fermi questions" [although I can't recall what for].

Students propose models and evaluate them. For instance, models to explain how sparks happen. Models like the mean free path or simple electric field ionization are examined and discarded, eventually being guided to a working model. Framing can help get the students to propose models you want to have brought up.

Once a model is plausible, does it make testable predictions?

The following flowchart was presented for construction of physical models:



[1] Conservation of energy would be a Fundamental Physical Principle, while Newtonian Gravitation would be a model.

TPB03: The Centrality of Models in Making Sense of the Physical World – Dewey Dykstra, Boise State University, ddykstra@boisestate.edu

[Dewey had to leave at lunch to make his flight, due to the ongoing problems at the hosting site that left him having to book the flight before they fixed the "no housing on Thursday night" situation.]

Realist view: reality exists, models reflect reality.

"Radical" constructivist view: models are all that there is, there's no objective reality to compare to, and a model is only valid for the model maker. [Dewey calls this radical constructivism, but most in attendance agreed that true radical constructivism went a LOT farther out on the branch than Dewey's poster. I suggested simply calling this "strong" constructivism.]

Amdahl, in There Are No Electrons (1991) says, "Never confuse your watermelon for the universe." In a strong constructivist view, it's not useful to present or communicate our models to others, everyone has to make their own model. We may end up recapitulating the essence of someone else's model, but that'll be more due to our shared or common experiences than due to any inherent "reality" of the model.

If we all have valid experiences, then we should keep disequilibriating students until they settle on models that are "good". We can't control what goes on inside their heads, but we can try to provide the best set of common experiences to shape its development.

Dewey uses completely unguided inquiry in his class, presenting the students with no models at all and making them construct their own models in an environment that favors models that are consistent with experimental results. His testing results seem to support this method.

During discussion, Gordon Aubrecht and Yuhfen Lin brought up the point that a shared culture makes it easier to share models than Dewey would have us believe. We're part of a shared environment, and it's more ubiquitous than we may realize. In other words, those common experiences may be so overwhelming that model transmission may be practical anyway.

TPB04: Exploring Modeling as a Process of Simplifying Real Objects, Interactions, Systems and Processes – Eugenia Etkina, Rutgers University, etkina@rci.rutgers.edu (with Anna Karelina and Arron Warren of Rutgers)

After a workshop on modeling, the question was asked, "What's a model?" This provoked some thought, hmmmm.

Physicists don't actually study reality, we study our models **of** reality. They may be linear, exponential or sinusoidal, and deal with points or continua, but they're all simplified from messy reality. They also must be predictive, and may be descriptive or explanatory. The predictive power of a model is always limited by the simplifications required in its creation.

Models are used to simplify the following things: objects, processes, systems of objects and their interactions, and the interactions between objects.

Reality is simplified into a conceptual model which is then quantified into a mathematical model.

In class, we rarely present students with situations that can be correctly modeled in different ways, and give them little or not instruction on how to pick the best modeling for a situation. For instance, when modeling the behavior of a circuit, different setups work best for predicting behavior when the resistance of the circuit approaches that of the ammeter, or that of the voltmeter.

So, when teaching models, we should try to make sure it's not a single-solution sort of thing. Make the students think about their assumptions, the domains of validity, and so forth.

Gordon Aubrecht brought up an HO scale train set as an example of the literal limitations of models. Much about it is right (surface details, general behavior) while much is incorrect (the power source, construction materials) or even missing.

AG08-10 results were cited as an example of how modeling helps.

Discussion:

We focused on how the four posters were similar in their approaches and how they were different.

Similarities:

- All spoke of the limitations of models in some way.
- All had some level of constructivism.
- We have to make assumptions about reality.
- Model making is an iterative process, at least in the validation portion.
- Each study offered a choice of models to test.
- All dealt with idealization and simplification.
- There is no such thing as a "true model". All models are wrong in some way.

Differences:

- How much guidance is needed? Some advocate none, others quite a bit.
- Some stated models were purely mental (01, 03, 04) while one treated them as analogous to physical things (02).
- 03 was more of a proselytizing poster, the rest simply presenting results.
- Two used real experiments (01, 04) and two used thought experiments (02, 03).
- The grain size of models was seen differently in the various posters.
- 03 felt models could not be generalized, 01 felt models must be generalized.
- Each poster had different theoretical underpinnings.
- Fundamental principles may (01) or may not (02) be models.

[I did not attend the second afternoon session, since doing so might have made us miss our flight.]

And so ends another installment of my AAPT meeting notes. I don't plan on attending Baltimore due to scheduling issues, but hope to be allowed out of the country to attend the Edmonton meeting in summer 2008.

Dave Van Domelen
dvandom@phys.ksu.edu