Educational Engineering Theory

John Stewart and Gay Stewart

June 6, 2004

1 Introduction and Motivation

Using concept inventories, the physics education research (PER) community has identified a fundamental flaw in traditional introductory physics instruction. Students do not learn to answer conceptual (non-calculation based) problems. The PER community also feels it has found the solution to this problem in the use of Interactive Engagement Methods (IE). These two results are quite attached in papers and talks within the PER community. We argue that the two results should be viewed as having greatly different levels of experimental support. The result that students do not learn conceptual reasoning in traditional instruction is now very soundly established with test results from a wide range of institutions showing very poor student conceptual gains. The problems on the Force Concept Inventory (FCI) and the Concept Inventory of Electricity and Magnetism (CSEM) are by and large simple applications of concepts most introductory instructors thought they conveyed. It was the assumption that focusing on the quantitative application of physical law would have the side effect that students would be able to apply the laws qualitatively. This assumption has categorically been proved wrong.

To address the failing of traditional instruction, a wide variety of new instructional methods have been created including peer instruction, micro-computer based laboratories, and group problem solving using white-boarding techniques. The new instructional methods are grouped under the title interactive engagement (IE) methods. Research by Richard Hake at a variety of institutions show that institutions that do not use IE methods have an average conceptual gain of 20%, institutions without a PER specialty that use IE methods show a 40% gain, and institutions with a PER specialty using IE methods show a $60-70\%$ gain. Conceptual or Hake gain is defined in terms of a normalized difference between pre-test and post-test scores

$$\text{Conceptual Gain} = \frac{\text{posttest} - \text{pretest}}{100 - \text{pretest}} \cdot 100\%$$

The results in themselves are very troubling. The highest performing PER institutions have students whose conceptual performance must be better than their quantitative performance, which strongly suggests at least some teaching to the test is occurring. Given the wide differences among the IE methods, it is curious that all methods and all combinations of methods yield approximately the same results. This suggests that there is an underlying factor that the methods share which is actually producing the gains. We propose the common element is simply the class seriously dealing with conceptual problems, that is having conceptual reasoning featured in a serious segment of the class where students understand it is something over which they will be tested. Traditional education simply does not cover conceptual reasoning at any level. We would expect that the first step to reform to be actually covering the conceptual material where the students read conceptual examples, saw conceptual examples worked in lecture, worked conceptual problems in homework, and could expect conceptual problems on tests. Just by acclimating students to this type of problem we should expect some gain. If we assume the student's pretest score was completely random, then the difference between the 20% gain of traditional instruction and the 40% gain of IE is only 5 problems on the concept inventory. The problems on the concept inventory are so fundamental that as one designs interactive activities and exam problems that test the concepts it is virtually impossible to not include activities and problems that are very close to the concept inventory problems.

We can accidentally provide some experimental support to the view that the success of interactive engagement is not in the method but in having conceptual reasoning as a real part of a science class. We reworked a science class at the University of Arkansas, University Physics II (UPII, more information on UPII is available elsewhere at this web site). The rebuilt course included readings that featured conceptual problems, conceptual lecture examples, conceptual homework problems, and conceptual exam problems; a full integration of conceptual reasoning into all elements of the course. Student score consistently a conceptual gain of 42%. Based on current PER research, one would have to assume this gain was due to the substantial interactive character of our laboratories. IE methods often mix purely interactive methods that could not be delivered in a traditional course with elements which could be modified for presentation in a traditional course. The peer instruction questions could be offered as homework, questions at the end of activities and some questions within activities could also be offered for homework. Because of our full integration, the majority of the conceptual material that could be traditionally presented was offered as homework and not included in the lab manual. The labs were converted for use by the algebra-based version of the course. This course, CPII, is taught out of a traditional textbook in a traditional lecture style. The only non-traditional element is the interactive labs. Students in the first offering of CPII scored only 16% on the concept exam, consistent with traditional instruction. From this we must conclude that feature of UPII that produces the 40% gain is the conceptual problems integrated into the homework, not the interactive character of the labs.

I note that this conclusion is drawn from an electricity and magnetism class where our student do not enter with strongly held misconceptions. Further research is needed to understand the effect of strongly held misconceptions in the mechanics class.

2 General Theory

To discus physical systems with fellow researchers, scientists have developed a precise theoretical language for describing the systems. This language is sufficiently detailed to capture the important features of the system, but not so detailed that irrelevant features prevent understanding. In this document, I seek to develop such a theoretical structure for educational systems.

The core of any theoretical description of any system, is a basis that captures the important features of the system in enough detail to understand the system. After much trial and error, we propose the independent steps found in the expert response to the educational system as the appropriate elements for our educational basis. These would take the form of independent intellectual productions found in the solution to problems, the working of activities, or the taking of detailed notes during reading and lecture. We will discus in more detail how these fundamental objects can be identified with reproducible precision later. For now, let the fundamental intellectual step in addressing a science class be called a *process* and denoted by P_i. The state of knowledge of a student can be represented by the set of probabilities, ρ_i, the student responds correctly to a problem requiring correct application of the process for its solution or a sequence of processes $\rho_{i,j,k}$ for its solution. We will write the state of knowledge as

$$K = \{\rho_1, \rho_2, ...\rho_n, \rho_{1,1}, \rho_{1,2}, ...\rho_{n,n}, ...\}$$

This vector clearly does not fully represent a student's mental state. It is a gross approximation to the state of a student's knowledge, but any useful theoretical representation of knowledge will be a gross approximation. The student knowledge vector is different for each student, i, and changes with time t as the class progresses. A single students progress through the class is written $K_i(t)$ and the full effect of the class is given by the set of knowledge vectors for all students $\{K_i(t)\}$.

Student knowledge changes with time even if the student does nothing. After the student leaves the class, most if not all of the probabilities in the knowledge vector will decay to zero over time as the student forgets. Some of the probabilities may actually increase either by the student seeing the full picture or by losing probabilities that were low and allowing the student to use only the pieces they truly understood. The details of how the decay progresses or how and if the increase progresses if at all would be exceptionally interesting research topics and are naturally formed in terms of the student knowledge vector.

Our concern, however, is with the functioning of the class and how it affects the knowledge vectors of the population of students. A class is built out of pieces that are meant to convey certain skills and build student knowledge; sections of lecture, homework problems, reading subsections and examples, and parts of laboratory activities. The division is not hard and fast but the designing instructor could subdivide the class and hopefully say this part of the class was supposed to convey this knowledge. We will call a non-subdividable part of the class that attempts to affect student knowledge a *fundamental interaction* and designate it as

$$I_i(\{P_j\}, \{D_j\}, \{S_j\}, t, \Delta t)$$

where i denotes the student, t is the time at which student i experiences the interaction, Δt is how long it took, and P_j, D_j, and S_j are processes the interaction should impact. The processes are divided into three educationally interesting subsets; $\{P_j\}$ is the set of primary processes, those found in an experts response to the interaction such as steps in a homework solution, steps in a lecture example, responses to a group discussion, or writing in a lab manual. $\{D_j\}$ are the direct processes, processes that the interaction is directly designed to impact but which do not appear in an expert response. For example if a lecture presents a formula but not an example the designer could still assert the student should be able to apply formula. The educational interaction is faulty if it does not impact the primary and direct processes. $\{S_j\}$ are the secondary processes, processes the designer hopes are affected. For example a discussion of gravitational potential energy would hopefully allow a student to reason about electric potential energy.

Fundamental interactions are presented in specific sequences to produce an effect greater than the sum of the interactions. Let a sequence of interactions be called a *composite interaction* and be denoted

$$I_i(\{I_j\}, \{S_j\}, t, \Delta t)$$

where $\{I_j\}$ is the sequence of interactions. Since all primary and direct processes are already captured in the fundamental interactions, the composite interactions contain only secondary processes. These secondary processes may include processes that are among the principal and direct processes of the contained interactions. It should be noted that the lists of processes may contain a processes more than once.

At multiple times during the class the state of student knowledge will be evaluated by tests or quizzes. The evaluation will contain problems and parts of problems that can be solved with some sequence of processes. The students score on the evaluation, the fraction of the problem they got correct, is related to the probability they could answer other questions involving the same processes. If a problem part requires the process sequence $P_1 P_3 P_2$ for an expert solution, we will record the fraction of the points received for the probability $\rho_{1,3,2}$. Therefore an evaluation, E, is a subset of the student knowledge vector at the time of the evaluation

$$E_i(\rho_4, \rho_{20}, \rho_{3,2,3}, ..., t)$$

3 Framing Educational Questions

We can use the formalism of the previous chapter to frame educational questions quantitatively.

3.1 Today's Big Question

The primary concern of PER researchers today is the role of misconception in student conceptual performance and the effectiveness of IE methods at overcoming those misconceptions. This is currently thought of as a comparison between the performance of two different classes, C_1 and C_2. A class is considered a set of processes, interactions, and evaluations. Let us divide the union of the processes sets for the two classes into two sets, conceptual processes Ω_c and quantitative processes Ω_q. Define an average for a specified evaluation E over all students over the probabilities of getting an evaluation question correct if the question can be answered by a processes in the set Ω_i as $< E >_i$. The basic tenant of PER research today can be stated

$$< E_1 >_c \gg < E_2 >_c$$

if and only if C_1 contains interactive engagement interactions and C_2 does not. To me, as stated this proposition is fairly un-testable since it doesn't specify how many interactive engagement interactions are required, what fraction of the total interactions are required for an interactive engagement class, or how the processes found in the evaluation are split between the primary, direct and secondary processes in the interactive engagement interactions. Further, when comparing two different education systems the variables of student population, instructor, and class design introduce so much variation in the class structures that it is unlikely anything quantitative can be said.

A supportable quantitative result is more likely if this question can be framed for a single class where the student population and instructor are the same. Divide the interactions into two sets, those which have IE character, Γ_{ie} and those that do not Γ_t where Γ is used for a set of interactions. Let Ω_{ie} be the processes impacted by the interactive engagement activities and Γ_t be the processes impacted by traditional instruction. The claim would then become

$$< E >_{ie} \gg < E >_t$$

and

$$< E >_t \approx 0$$

This statement avoids the problems that traditional instruction contains few conceptual processes and therefore scores poorly on composite concept evaluations. This proposition can be refined in various ways. I would find it very interesting to compare the performance of the interactive engagement activities based on the process divisions, primary, direct, and secondary defined earlier.

3.2 My Big Question

To me, the proposition in the previous section is far to exotic to be the first question of interest in physics education. I feel that previous educational reform effects have failed primarily because the researchers attempted to improve the system before they actually understood the system. Our research therefore is primarily concerned with how educational systems function rather than how a specific technique can be used to improve them. I feel the primary problems with education is that both the educator and the student greatly overestimate the amount learned in a class. Students and teachers think in terms of topics mastered, but my experience is that students learn very locally and have a very hard time generalizing between two very related skills. The question I would most like answered is to what extent students learn anything outside of the processes found in the primary process set of the interactions. I strongly suspect

$$< E >^p \gg < E >^d \gg < E >^s$$

where the superscripts stand for primary, direct, and secondary process sets. If this proposition is true it converts our perception of student knowledge from a few big things to a lot of little things.

3.3 The Role of Repetition

Processes will generally occur more than once in the interaction set of a class. It has been my experience as an instructor that students do not master a skill simply by seeing it once. By counting the number of occurrences of a process in the interaction set of a class, an occurrence vector, O, can be constructed

$$O = (n_1, n_2, ...n_{N_p}, n_{1,2}...)$$

Comparing the occurrence vector with the student knowledge vector will all the determination of how much if any repetition is necessary for student mastery.

3.4 Real Coverage

The occurrence vector will allow the determination of the real coverage of the class. Processes that are only found in the course reading, but which find no other expression in the class interaction set, may be viewed as not actually covered by the class. The course reading defines the theoretical topical coverage of the class, the coverage vector the real coverage. Their relation would be very interesting.

3.5 Relative Efficacy

The current belief is that traditional lecture is an ineffective way to educate students and that interactive laboratories are the only effective conceptual education method. More in general, we can compute the relative efficacy of various types of interactions.

3.6 Some Fundamentals

First however, I would like to know some of the basics. How many processes are there in a science class? How are the processes distributed through the interactions, that is if I give an hour lecture how many processes have I expose? Likewise, how many processes are required on average to complete a homework problem? How many of the processes are conceptual?

4 Identifying Processes

To make any of this work, a method is needed for identify processes in a way that is not simply the opinion of one observer. We initially tried simply examining examples and asserting a set of processes were required for the solution of the example. This generated an unconvincing model. From this experience, we developed a more supportable technique for building processes models. The key is to have tangible constraints on the creation and association of processes. The constraints being used for the current modelling effort are

- Each use of a process must have textual support, that is one must be able to identify a block of text that represents the use of the process.
- The process titles must be usable as subsection titles for the modelled example.
- A generic example built of process titles and general models for how the process works must provide an intelligible generic framework for the solution of the problem modelled.

I would expect many different process identification schemes to be developed by different researches eventually culminating in an agreed upon process set maintained by the APS.