

# Difficulty Based Evaluation of Teaching Methods and Materials

Dan Kalman

July 15, 1998

## Overview

How can you tell if one instructional method is better than another? An obvious response is to measure what the students learn as a result of each approach. Presumably, if one method is better than the other, the students taught by the better method will learn more, or learn better. Therefore, you can infer the relative merits of the two methods from assessments of student performance. This approach may be criticized for neglecting the effects of student effort. It is conceivable that two approaches of instruction might yield the same levels of student achievement, but only by virtue of much greater effort by students subjected to one of the methods. Although the educational outcomes may be the same, most teachers and students would prefer a teaching method that produces those outcomes with less effort. In that sense, one of the two methods would be better than the other, in spite of the inability of student performance to distinguish between the two.

The situation described above, in which students expend measurably disparate levels of effort to end with the same levels of performance under two different teaching methods, may seem unlikely to occur. Students taught by two different methods might very well expend different levels of effort, but it would be surprising if such a difference were so closely correlated with the effectiveness of the teaching method as to completely mask a significant difference between two methods.

In this essay it is argued that such an outcome, far from being considered unlikely, should actually be the expected result of student performance based comparisons of teaching methods. This point will be discussed in the following section, **Effort Self-Regulation**. A succeeding section will propose an alternative method of comparing teaching methods, **Difficulty Based Evaluation**, and will describe several directions for research based on this method.

## Effort Self-Regulation

For illustrative purposes, consider the following simplistic model of mathematics student behavior. Each student has a target grade that is desired for each course. The  $A$  students are shooting for an  $A$ , the  $B$  students seek a  $B$ , and so on. Over a course of instruction, students receive periodic feedback from the teacher indicating the level of performance. Each student then calibrates his or her level of effort based on that feedback. Thus, a  $C$  student who somehow manages to make a  $B$  on a test will relax a little bit. Conversely, an  $A$  student who scores a  $B$  will work harder.

Now imagine that students are taught in two different classes by two different methods, but that the performance measures used to determine grades are roughly equivalent. Further suppose that the method of presentation in one of the courses is confusing and poorly organized, while in the other course, the material is presented in a clear and coherent fashion. According to our simple model of student behavior, the classes would result in roughly equivalent learning for the two groups of students. In one class the students might have to work much harder than in the other, but the  $A$  students in each class would make the necessary effort to earn an  $A$ , likewise the  $B$  students,  $C$  students, etc.. In this way, by regulating their levels of effort, the students would more or less exactly compensate for the differences in the instructional delivery. Measuring what these students learned, we should find no reason to prefer one of the methods of instruction over the other.

While the model of student behavior is ridiculously simplistic, it does reflect an important aspect of learning, and one that is too often ignored in empirical research on teaching effectiveness. Our students have many demands on their time outside of the ones imposed in our classes. We should not be surprised if they develop strategies for balancing these demands based on the outcomes they wish to obtain. Anyone who has taught college mathematics certainly knows how common it is for students to keep track of grades in progress, and set specific goals for final grades. Who has not been asked, “What will I have to score on the final to get a  $B$  in the course?” Certainly this behavior, which might be termed *student effort self-regulation*, varies among students. The accuracy of the simplistic model will be much greater for some students than for others. But it seems clear that it describes a consistent and pervasive influence on student behavior, and one that has the potential to seriously confound any achievement based comparison of teaching methods.

Achievement based comparison can also be challenged on philosophical grounds. I have never felt that my job or goal as a teacher is to put learning into the heads of my students. Ultimately, that is something that each student must do. No, my goal is to facilitate the process. If I can feel that I have made the student’s process of learning a little easier, then I have done my job. In this context, what is the natural way for me to choose between two competing approaches to teaching a subject? Clearly, I want the one that students learn from most easily.

In curriculum development and improvement projects, it is rare to observe significant effects on student

performance. Some of these projects are rejected by critics who ask, “Why should I change to a new method if you cannot show empirical evidence that it will lead to gains in student performance?” It is my view that this question is fundamentally misdirected. If some new approach allows students to achieve equivalent results with less effort, or with greater enjoyment, that new approach should obviously be preferred.

Of course, the consideration of student attitudes is a familiar aspect of teaching method evaluation, and it is not intended to consider it in this discussion. Rather, attention will be focussed on difficulty, as apprehended by the students.

## **Difficulty Based Evaluation**

The idea that is being proposed here is simple: compare two methods of instruction by measuring how difficult it is for students to learn from each. How might difficulty be measured? Here are few suggestions.

### **Survey Methods**

To find out how difficult a learning task is for students, ask them. One possible experimental design along these lines might compare two presentations of a topic in a course. A control group of students would be taught using the standard methods; the experimental group would be taught one topic in a new way, and all other topics in the same manner as for the control group. Students would be asked to rate the difficulty of mastering a variety of procedures, concepts, notations, etc. encountered in the course. These would be related to several different topics, selected to represent a range of difficulty as judged by someone who has expert mastery of the material. The topic presented by the experimental teaching method would be just one of these topics. In analyzing the results of this survey, the student responses regarding topics taught by the same method for both groups would be used to control for the variations in student perception of difficulty. The goal would be to find measurable differences in the perceived difficulty for the experimental topic between the two groups of students.

### **Time on Task**

Get students to record accurately the time spent studying for a particular topic, prior to some assessment. Correlate the time on task and the level of achievement for the two competing instructional methods. This approach presents implementation difficulties, and care would be required in assessing the accuracy of the time recording process.

## **Instant Learning Assessment**

Give a lesson and then instantly assess what the students learn. Compare the results for different methods to teach the same thing. The idea here is to remove the confounding effects of students working to achieve a subjectively chosen level of mastery by testing them before they have the opportunity to study.

Here is a possible experimental design for this kind of study. Work with student subjects in individual sessions. Give each subject a set of topics to study during the session. One topic is considered the experimental treatment and is presented to each subject in one of two different ways. The remaining topics are controls and are presented in the same way to each subject. Tell the students that they will have a fixed time period to study several units. Each unit will be studied, and then the student will be asked to answer questions about the material. If these questions are answered successfully, the student will earn a reward. The incentive will be for the student to amass as large a reward as possible in the available time. Now the students will want to work fast so they can try as many topics as possible. They will not want to try to answer questions too soon, though, because they won't win the prizes. So the students will really be motivated to learn the material as well as possible in the minimum amount of time. Now we can use the performance of the students on the control topics to profile student abilities. We randomize the order of the topics presented to the students to control for that effect. If repeated with a large enough number of students, it should be possible to determine if one or the other of the two approaches to the test topic that are being compared is 'easier' in the sense of producing better results in time limited study. In this setting, we seek to eliminate the effects of student effort self-regulation by allowing a fixed amount of time for the session, and providing a uniformly high incentive for all participants to do well on all topics.