Using Field Experiments to Change the Template of How We Teach Economics

John A. List
University of Chicago and NBER

Mark Twain once said that “All you need in this life is ignorance and confidence; then success is sure.” I suspect that he is right for describing many walks of life: whether I visit companies such as United Airlines, non-profits such as Smile Train, or someplace closer to home—the classroom—I find that field experiments can improve what we do and how we do it. Consider the current state of economics teaching. The landmark economics textbook for my generation was created by Paul Samuelson more than 60 years ago, and his template has been followed ever since by most modern economics textbooks. This is despite an economics field that is quite different from when Samuelson first scribed his masterpiece.

The central theme of this article is that with developments in economic thinking and methodology, that template is no longer consistent with current best practices in economics. The current template does not convey to students how modern economists approach economic problems, missing an important opportunity to showcase the empirical developments of the last few decades. With this as a background, I argue that our current teaching template is in need of reform, and in this short article I highlight three areas of my own research that can easily be included in the classroom. Using my own research as the vehicle for change is for convenience only, and is meant to follow my plenary discussion at the AEA’s Third Annual Conference on

---

1 This teaching note is adapted from my plenary talk at the AEA’s Third Annual Conference on Teaching Undergraduate and Graduate Economics that took place in Chicago, IL, May, 2013.
Teaching Undergraduate and Graduate Economics. Of course, one can find several other excellent recent empirical examples of how to improve our pedagogy with an evidenced-based approach. The problem, it seems, is not the supply of excellent empirical examples rather the demand to make use of them in our textbooks.

The Need for Change

To understand the underpinnings for the standard template, consider Samuelson and Nordhaus’ (1985) discussion of how economists differ from other scientists when doing empirical work. They write

“The economic world is extremely complicated. There are millions of people and firms, thousands of prices and industries. One possible way of figuring out economic laws in such a setting is by controlled experiments……like those done by chemists, physicists, and biologists……..Economists have no such luxury when testing economic laws. They cannot perform the controlled experiments of chemists or biologists because they cannot easily control other important factors. Like astronomers or meteorologists, they generally must be content largely to observe.”

Following the spirit of this passage, the traditional method of empirical research is a semi-automatic approach of writing down a model, downloading mounds and mounds of secondary data, and invoking various econometric assumptions to make causal statements based on the empirical results. Technically, the problem that this approach confronts is creating a convincing counterfactual so the analyst can make causal inference—the gold standard in the social sciences.
Recent years have witnessed economists using many approaches to construct valid counterfactuals. To show this point, I find it instructive to consider empirical methods more broadly (as I have done in elsewhere, most recently in Al-Ubaydli and List, 2012). The Easternmost portion of Figure 1 highlights some of the popular empirical models economists have used to analyze naturally-occurring data. Each of these approaches uses a different set of assumptions to make a cause-effect statement.

<table>
<thead>
<tr>
<th>Controlled Data</th>
<th>Naturally-Occurring Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab</td>
<td>NFE, NE, PSM, IV, STR</td>
</tr>
<tr>
<td>AFE</td>
<td></td>
</tr>
<tr>
<td>FFE</td>
<td></td>
</tr>
<tr>
<td>NFE</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1: A field experiment bridge**

Experiments, on the other hand, provide a convincing method of creating the counterfactual since they directly construct a control group via randomization. Randomization acts as an instrumental variable by creating variation in treatment among participants. Proper randomization allows you to make causal statements, importantly going beyond correlation.

The oldest type of experiment is in the Westernmost portion of Figure 1, the laboratory experiment. Laboratory experiments typically make use of randomization to identify a treatment effect of interest among student subjects. The laboratory approach was pioneered by Vernon Smith, who won the Nobel Prize for his contributions in 2002. By now thousands of laboratory studies have been published in economics.
Field experiments represent a movement to take the data generation process beyond the walls of the laboratory. Two decades ago, the primary data generators were lab experimentalists. The past 15 years has witnessed an explosion of creative ways to generate data in the field. Harrison and List (2004) propose a classification scheme that helps to organize one’s thoughts about the factors that might be important when moving from the lab to the field.

The first departure from the typical laboratory experiment is the “artefactual” field experiment (AFE), which mimics a lab experiment except that it uses “non-standard” subjects. Such subjects are non-standard in the sense that they are not students, but participants drawn from the market of interest. Moving closer to how naturally-occurring data are generated, Harrison and List (2004) denote a framed field experiment (FFE) as the same as an AFE but with field context in the commodity, task, stakes, or information set that the subjects can use. This type of experiment is important in the sense that a myriad of factors might influence behavior and by progressing slowly toward the environment of ultimate interest one can learn about whether, and to what extent, such factors influence behavior one by one.

Finally, a natural field experiment (NFE) is the same as a FFE in that it occurs in the environment where the subjects naturally undertake these tasks, but where the subjects *do not know* that they are participants in an experiment.\(^2\) Such an exercise is important in that it represents an approach that combines the most attractive elements of the experimental method and naturally-occurring data: randomization and realism. In addition, it importantly tackles a selection problem that is not often discussed concerning the other types of experiments (see List, 2012).

\(^2\) This raises the issue of informed consent. For a discussion on this, and related, issues see Levitt and List (2009) and List (2008, 2011b).
At this point the reader might appreciate the problem with the Samuelson and Nordhaus (1985) description above: economists today are not simply passive observers. The reality is that economists have made enormous strides in developing empirical methods that shed light on important economic issues. Economists no longer develop a deductive model “assuming rationality” and then run a regression to “test” that deductive model in a way that satisfies almost no one. These most recent experimental advances have led recent textbooks to at least mention the experimental approach (Samuelson and Nordhaus’ (1989)):

"The economic world is enormously complicated, with thousands of prices and millions of households. In an exciting new development, economists are turning to laboratory and other controlled experiments to understand complex economic processes..."

While modern texts are now, at least, mentioning that experimentation is done in economics, at issue is how economists think about empiricism in economics and whether they make use of the new approaches in their lessons. By and large, modern textbooks do not.

**Bringing the Classroom Alive with Field Experiments**

My research focuses on using field experiments to deepen our understanding of the economic science. In the remainder of this paper, I will show how some of these field experiments can be leveraged in the classroom. Specifically I give examples from 3 major areas that make use of AFEs, FFEs, and NFEs. Although in the classroom I have found that NFEs usually provide the most interesting conversation for the students, discussing examples of each shows how empirical economists do field experiments and provide students with insights into the economics of everyday life. The three examples highlight teaching points in three areas:
1. Factor markets: Why do women get paid less than men in labor markets?

2. Introducing Behavioral Economics: How can we use behavioral economics to motivate teachers?

3. Introducing Discrimination: What seven words can end third-degree price discrimination?

If the reader desires a more patient (as well as accessible) discussion of these three issues and many more related field experiments, I direct them to Gneezy and List (2013).

1. Why are men paid more than women in labor markets?

   When it is time to explain factor markets students face a core challenge: at that moment in the course, students have focused completely on goods such as cell phones, pizzas, beer, or widgets. Changing gears and trying to understand inputs used to produce those outputs many times confuses even the best students. They have yet to think hard about the determinants of how much labor, how many machines, or land needed to produce the final product. And, while they have heard about reasons why people might earn different wages in the labor market, they have not thought critically about a general understanding of the determinants of wages.

   All texts discuss pay differentials, but bringing those to life with field experimental data to help students truly understand “The Whys” behind factor price differences can be quite rewarding. How I tackle this problem is by starting with the standard Census data showing a pay gap exists between men and women. I then introduce important reasons why this pay gap might exist: human capital differences, labor market attachment, discrimination, hours worked, etc. I then introduce a relatively new explanation: on average men tend to be more competitively inclined than women. I then ask—is that true? And, if so what is the explanation for it?

   A rich literature using lab experiments has developed that suggests men choose to compete more than women. This is fascinating laboratory work done by Uri Gneezy, Muriel
Niederle, Aldo Rustichini, Lise Vesterlund, and by now many others. For example, in an experiment about making basketball free throws (where men are better than women), men are more likely than women to choose to have their pay tied to their success. But, this is also the case in activities where women are better than men. For example, in solving anagrams women are better than men, yet more men choose to have their pay tied to their success when solving anagrams.

One can then ask—does this behavior manifest itself in the field? To gain more insight into this issue we developed the following natural field experiment (see Flory et al., 2012). We advertised otherwise identical office jobs across 16 cities in the US, in one treatment arm, we varied the payment scheme: fixed wage ($15/hr.) versus competitive scheme ($12/hr. + $6/hr. if you outperform a fellow worker). We did it with nearly 7,000 observations. Here is what we found in terms of measuring the probability of completing an application for the job:

Figure 2: Application Probability by Treatment and Gender
What it tells us is that in the fixed wage treatment, men complete applications at a rate slightly greater than 50 percent whereas women complete applications at a rate slightly higher than 40 percent. The right-hand panel tells us that both men and women shy away from the competitive environment, but women shy away from the competitive pay scheme more than men.

So it seems that there is some evidence that males slightly prefer competitive pay schemes more than women, at least on average and in this natural field experiment. Combining these data with the laboratory data provides evidence consistent with the notion that men are more competitively inclined than women. But we decided to look more closely at the issue and to test if the differences in competitiveness are solely due to “sex” differences, or whether they are at least partly socially determined.

To do that we visited distinct societies with an artefactual field experiment. We found two distinct societies. The first was one matrilineal, a small Khasi society in India where the wealth runs through the youngest female in the family and men “are sick of playing the roles of breeding bulls and baby-sitters” (Ahmed, 1994). The second was patriarchal, the Masai society in Tanzania where women are treated as property, and women feel that “Men treat us like donkeys” (Hodgson, 2001). We then developed a simple game that paralleled the laboratory games that had been developed in the Western World: we had people throw balls into a bucket. Here are the results we obtained:

Figure 2: Percent Choosing to Compete by Society
Figure 3 provides three panels of data. The left-most panel provides an example of a result that has been found in U.S. studies: men choose to compete much more than women. Data from the Masai tribe are similar to the U.S. data in that men compete much more often than women. Yet, a very different results is observed amongst the Khasi. In the Khasi sample women actually compete much more often than men in these simple games.

This results in and of itself fills up a lecture with interesting discussion! But, then one moves to interpretation, which can consume another hour if you do not hold people back. This is because one speculative interpretation of this result is that Khasi society may remove social barriers that prevent naturally competitive women from expressing their true personalities. Another interpretation of the data is that Khasi society may allow competitive women to earn greater rewards for their effort and to pass on wealth to their daughters, both of which increase the fecundity of their competitive genes. In either case, everyone agrees that these results need to be replicated and that further treatments need to be carried out to detail the underlying structure at work, but the discussion of them in the classroom generates excitement and shows students
how modern economists go about answering questions of input prices and reasons for varying prices.³

2. Introducing Behavioral Economics

Let me next turn to another field experiment that is useful to introduce behavioral economics while serving to discuss an important social issue: how to increase student achievement. The underlying notion is whether we can leverage the concept of loss aversion to improve primary education. Loss aversion is a concept developed by Daniel Kahneman and Amos Tversky, and suggests that people are averse to losses: people value losses much more than comparable gains. The phenomenon has been found in both lab and field experiments, and was a key reason why Danny Kahneman won the Nobel Prize in 2002. Can we put this concept to work to improve education?

Here is what we did. We created the following framed field experiment in which teachers in Chicago Heights were given a chance to increase their salary through a bonus. Specifically, if their students improved their test performance the teachers could earn as much as $8,000 over a one-year period. The expected value of the reward was $4,000—an 8% increase in their annual salary, so it was substantial.

Teachers were randomized into three groups: one group was the control group—teachers in this group received no reward. A second group received the reward as a gain, or what one might consider a ‘traditional’ bonus scheme: these teachers are informed of the reward and the rules to gaining the reward at the beginning of the school year (September). They are paid the bonus at the end of the year (June) if their students achieved. The third group was the loss

³ You can read more about this artefactual field in Gneezy et al. (2009).
group—teachers in this group received $4,000 in September but they had to write a check to us for the difference in June if they earned less than $4,000. Here are the results:

Figure 4: Standardized Thinklink Math Score by Treatment

![Graph showing standardized thinklink math scores by treatment.]

Notes: Student’s score is normalized to have mean zero and standard deviation one within each grade. Results pool the individual and team treatments.

We find that while the incentives worked to motivate teachers across both the gain and loss treatments, the students who were in the classrooms of teachers who were randomly placed in the loss group had much greater test score growth (roughly 0.22 standard deviations) than students placed in the classrooms of teachers in the other two groups.4

3. Introducing Price Discrimination

My third example concerns the teaching of price discrimination. Economists usually state two major reasons why people receive different price quotes for the same good: i) because discriminators have a general distaste for a specific type of person—by say age, gender, or race

4 You can find more information about these experiments in Fryer et al. (2012).
(see Becker, 1971) and they are willing to give up profits (or utility) to cater their prejudice; or ii) a discrimination whereby the entrepreneur is using observable characteristics to infer how to price in a way that maximizes his own profits. Pigou denoted this type of discrimination as third-degree price discrimination. In our natural experiments we aimed to not only measure if discrimination was occurring, but to detail the underlying nature of that experiment.

We began by sending economic agents—both disabled (those confined to a wheel-chair) and abled agents—to automobile repair shops around Chicago to have identical cars repaired. We found that the disabled received price quotes that were about 30 percent higher than the abled received. This finding is squared under the theory of taste-based discrimination because mechanics might charge the disabled agents more because they are inherently biased against them—they have a distaste for people confined to wheel chairs. Alternatively, this result is consistent with the model of 3rd degree price discrimination if the body shop mechanics charge disabled more because they believe that there are search differences (people in wheel chairs search less than those not confined to wheel chairs) so they can charge more to that group, earning greater profits. In surveys we did we found that indeed not only do mechanics believe that there are search differences, but it is also true—disabled search less, which is consistent with 3rd degree price discrimination.

To delve more deeply into the question we performed a second natural field experiment. We began with a replication treatment, including 12 disabled and abled agents seeking to fix their vehicles. The results were as expected: disabled testers received roughly 20% higher offers, and this is shown in the left-hand panel of Figure 5. We then did the same experiment, but we had both groups explicitly say “I’m getting a few price quotes today.” These results are in the right-hand panel of Figure 5:
As you can see, this simple statement eliminated the differences in quotes, providing data consistent with third-degree price discrimination, not the simple taste based discrimination model.\(^5\) This is because if simple taste based discrimination was occurring, then even in the case where both the disabled and abled are “getting a few price quotes,” the disabled should still be discriminated against: if they were disliked in the earlier field experiment they should be disliked now. Likewise, the third-degree price discrimination model predicts that if discrimination is driven by perceived search differences, once those differences are negated there should be no difference in treatment. This is what happened.

**Concluding Thoughts**

This was a very brief presentation of how field experiments can be used in teaching. Each example provides a different view to students of economics, and gives them a good sense

\(^5\) You can find more information about this experiment in Gneezy et al. (2012).
of how modern economists approach economic thinking, and thereby offer alternative ways to structure the empirical dialogue that we hold with our students.

So where does that lead us concerning the original Samuelson-Nordhaus (1985) quotation? Here is how I would rewrite it:

In the past, economists have believed that a complicated world necessarily neutered Homo Experimentalis, rendering empiricism as a passive enterprise based on pure observation. The world is indeed very complicated, with many simultaneously moving parts. This fact represents a keen advantage, not disadvantage, of experimentation. This is because randomization allows the researcher to balance the unobservables across treatment and control groups, leveraging the data generating process to deduce causal relationships. In many complicated settings, more traditional empirical methods have difficulty going beyond correlations.

References


Sadoff, Sally, Roland Fryer, Steven Levitt, and John List. “Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment”, NBER Working Paper No. 18237


**Further reading**


