PLS 506
Program Evaluation

Take Home Final Exam

Directions: You should answer all of the questions on this exam as clearly and briefly as possible. Each question is equally weighted and partial credit for incorrect answers will be awarded so answer all questions and their sub-parts. Be sure to properly label and answer all parts of each question. Your exam is due by **5:00 PM on Tuesday July 30, 2013**. You may not talk with others in the class about the questions or your answers. You should sign, date, and attach the honor pledge to your exam. I will be available to answer any questions you have about the exam during regularly scheduled office hours. You can also arrange an appointment to discuss the exam.

1. Suppose you were interested in the measuring the concept of "poverty" among citizens.
   a. Describe two different indicators that can be used to measure this concept.

   b. Provide an operational definition for both indicators and be sure that the list of categories for each variable is exhaustive and mutually exclusive but be sure to keep your lists relatively short (e.g., no more than 6 categories).

   c. Why do social scientists prefer using multiple indicators of a concept rather than a single indicator?

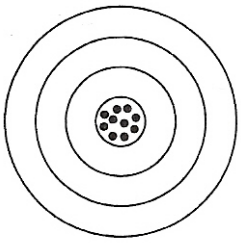
   d. Suppose you had the choice of gathering data for your indicators from the following sources: archival records; surveys (telephone, web, or mailed); focus groups; interviews; direct observation; or trained observer ratings. Which data source would be the most reliable measure for each of your indicators of "poverty"? Which data source would be the least reliable measure of your two indicators? Be sure to explain your reasoning.

   e. What are some possible sources of measurement error that could occur when trying to collect data using the preferred sources noted in the previous question? What steps would you take to minimize the measurement error?

   f. How would you assess the reliability of each measure of your indicator? How would you assess the validity of each measure? What is the relationship between the measures' reliability and validity? Be sure to explain your answers using examples related to measuring the concept of "poverty"

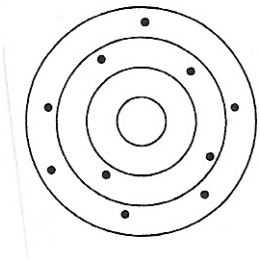
   g. What are some potential construct validity problems for each of your measures? Be sure to provide examples.

h. Using the analogy of a shooting target with the accuracy on the center of the target indicating the precision of the measure, a social science researcher applies a measure that produces the following distribution:

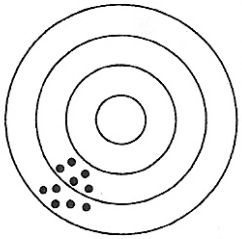


Target A

Why would you prefer the distribution in Target A to the distribution of Target B or C below?



Target B



Target C

i. Suppose your only options were measures of an indicator that produced the types of error illustrated by Target B or C. Which would you prefer and why? In your answer be sure to comment on the reliability, validity, and types of measurement error.

2. In 1980, the Chicago Public School (CPS) system adopted a school choice program that allowed all incoming freshmen to apply to virtually any high school in the school district, the third largest in the U.S. with more than 60 high schools. Roughly half of the CPS students opt out of their neighborhood school. Since the schools with good test scores and high graduation rates would have more students than space, the CPS resorted to a lottery. Imagine two students, otherwise statistically identical, each of whom wants to attend a new school. Due to the random selection process, one student gets to go to a new school while another is left behind. The operation of this program therefore provides a "natural experiment" to test whether selecting a good school improves educational performance because there are grades both before and after the selection process. Note: This example and findings are from Steven D. Levitt and Stephen J. Dubner, 2006. *Freakonomics [Revised and Expanded]: A Rogue Economist Explores the Hidden Side of Everything* (Harper Collins: New York, NY). Pages 143 - 146.

   a. Why is a situation like this referred to as a "natural experiment"?

b. What type of research design does this data provide? Use the conventional notation of X for a treatment, O for observation, C for control, R for random, and NR for non-random to diagram out the research design.

c. Describe how this research design controls for (or fails to control for) the following threats to internal validity: maturation; history; instrumentation; testing; and selection.

d. In this case, the data indicate that students who entered the school choice lottery were more likely to graduate than students who didn't. Students who choose to enter the lottery also tended to be smarter and more academically motivated. However, students who won the lottery and went to a "better" school performed no better than statistically equivalent students who lost the lottery and were left behind in a "bad" school. Similarly the students who opted out of the lottery tended to test at about the same levels, albeit at lower levels, both before and after the lottery but their performance was no worse by remaining in their current school. Formulate a hypothesis that explains these findings?

e. What are the main threats to external validity in this "natural experiment"?

f. How might construct validity be a problem in this case?

g. While the results referred to above were "statistically" significant, why might the evaluator be more concerned about the magnitude of the effect (i.e., substantive significance) rather than its statistical significance?

3. Consider the following two research designs diagramed using the conventional notation of X for a treatment, O for observation, C for control, R for random assignment, and NR for non-random:

R X O
R O

Option 1

R O X O
R O O

Option 2

a. Explain how randomized assignment in options 1 and 2 can or cannot be used to rule out or control for the following threats to internal validity: maturation; history; instrumentation; testing; and selection. Be sure that you provide examples to support your arguments.

b. If the research designs depicted in Options 1 and 2 were modified so that there was no method of random assignment to treatment and control groups, what new threats to internal validity would emerge?

c. What is the principal threat to external validity in both research design options?

d. When might you prefer option 1 to option 2?

e. When might you prefer option 2 to option 1?

f. Using language that can be understood by someone without training in statistics and research methods explain the advantages of relying on randomized experiments and the value of using random assignment to treatment and control groups.

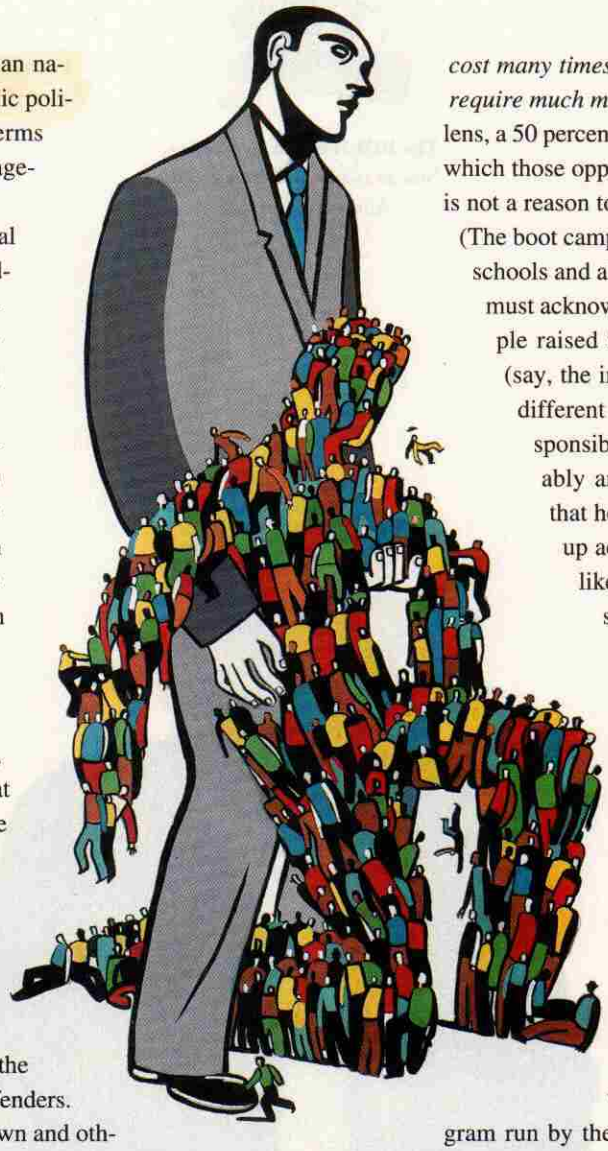
4. During this semester, we spent a great deal of time discussing a variety of issues related to conducting program evaluations. We discussed the value of the information generated from the careful application of social science research methodology to the study of program processes and impacts. Conversely, we discussed many of the challenges associated with conducting program evaluations. Read the attached essay by Amitai Etzioni that was published in the *Atlantic Monthly* in July 1994. Based on this essay, do you think Etzioni would agree or disagree with the following statement:

   - *Program evaluations provide critically important information because policy makers and the public care a great deal about whether government programs have a measurable impact on public policy problems*.

   In your answer, be sure to describe Etzioni's basic arguments? Do you agree or disagree with his arguments and the premises behind them? What do his arguments suggest about the value of program evaluations to policy makers and the public? How important does Etzioni thing program evaluations are? What type of information does he think the public and policy makers want?

Incorrigible

Bringing social hope and political rhetoric into instructive contact with what it means to be human



JUST how incorrigible is human nature, and what lessons on public policy follow once we come to terms with the sobering answer to this age-old question?

On one level, both from personal experience and from numerous studies, we know that it is extraordinarily difficult to change habits, personality traits, culture, and social institutions. Take weight control. Ninety-five percent of the many millions of Americans who diet each year in pursuit of a more attractive and healthier figure regain within one to three years almost all of the pounds they previously shed. Even when faced with certain death our bodies often seem unable to follow our minds. Many men who have had heart attacks continue to smoke, though they have been warned that they are significantly increasing the likelihood of a second (and most likely fatal) attack.

Studies demonstrate that numerous drug-rehabilitation, crime-prevention, and job-training programs for welfare clients yield results no better than the new boot camps yield for young offenders. But just as dieters disregard their own and others' experiences, the very strong propensity of all these programs to fail has not stopped Congress from making them a cornerstone of its anti-crime policy.

We all know how difficult changing human behavior is, but this knowledge has not changed our basic optimistic predisposition. Once we truly accept that human nature is surprisingly resistant to improvement, however, some rather positive, constructive lessons follow.

by Amitai Etzioni

Lower your expectations. When seeking change, expect it to cost many times as much as you first imagined and require much more time. When viewed through this lens, a 50 percent graduation rate from boot camps—which those opposed to the camps see as a failure—is not a reason to close them but a stunning success. (The boot camps are doing better than many public schools and almost as well as some colleges.) We must acknowledge that hoping to assimilate people raised for twenty years in one subculture (say, the inner city, as a gang member) into a different subculture (of work and social responsibility) in only a few months is laughably ambitious. And we must not expect that helping disadvantaged children catch up academically by means of a program like Head Start will keep them on the straight and narrow when they graduate from the program. This expectation is especially absurd if we return them to a world—rife with prejudice and economic and social disadvantages—that agitates against hard work and progress by legitimate means. Growing evidence shows that kids may need long-term and costly "maintenance" programs. A Loyola University study, for instance, found that children who participated in a special program run by the Chicago public schools from age three until the end of third grade tested at or above national averages on completion but that a decade later many of these gains had melted away. The researcher, J. S. Fuerst, now believes that disadvantaged youngsters may need up to *nine years* of supplemented education. He may well be underestimating what they actually require.

A firm grasp of our stubborn, improvement-resistant nature will keep our future policymakers—one hopes against hope—from being swayed by social-policy entrepreneurs. Take the reports on America Works, a for-profit

Illustrations by Alison Seiffer

corporation that trains welfare clients for *one week* and then finds them steady jobs. The company has recently become the darling of New York Mayor Rudolph Giuliani's administration, because it fits the new mayor's ideology (greater reliance on the private sector) and because the city hopes it can place many of its 1.1 million welfare clients in jobs this way and slash welfare expenditures. In New York, America Works has so far placed only about 1,100 people; the initial reports of its incredible successes were based on 290 cases. You can bet your next year's salary (I herewith offer mine) that most welfare clients will not be put to work after one week of training. Only blurry-eyed optimists will believe that such a program can work on a large scale.

Creaming is okay. America Works does provide a valuable lesson: because it is paid much of its fee only after a welfare client has been on the job for four months, it prefers to work with clients who are easy to place—those most disposed to work and most able to do so—and drops from its program those who show signs of presenting a difficulty. Anyone who shows up five minutes late for a training session is out, for example. Critics call such an approach "creaming," and consider it a serious defect.

This criticism presupposes that there are other methods of converting an entire barrel of milk into cheese—or, to put it less metaphorically, that an organization can help all or most welfare clients to find jobs without making an immense per-client effort.

But if we start with a more realistic notion of human transformation—which acknowledges that it is typically very costly and challenging, and that the deeper we reach into the barrel, the greater the difficulties we'll have—creaming suddenly stops being a mark of defective policymaking and becomes a wise policy to try to implement. Taking on the relatively easy cases first gives us much more bang for whatever bucks we can spend on a given endeavor than taking a random sample of all potential clients. The resources saved this way can then be applied to some of the more difficult cases. Policymakers should, though, recognize the fact that the going will get tougher and tougher.

Don't expect to scrape the bottom of the barrel. More difficult to come to terms with is that even if we dedicate considerable resources and persevere for years, we will still have at least some nontrivial "residue": people who cannot be reached. Just as there are some patients too ill to cure, there are some welfare clients who cannot be weaned and some criminals who cannot be rehabilitated. How to deal with them is a major civil-rights issue. For example, the State of Washington has passed a law that requires prisons to release violent sex offenders not immediately after they have completed their sentence but only after a judge or a jury has determined that they no longer pose a danger. Such indeterminate sentences don't become defensible until we accept that not only are most people rather difficult to change but some people are incorrigible—at least given the current state of medical and social-science knowledge.

Don't allow the best to defeat the good. Both the social-science literature and the daily press are full of reports about this or that government program that turned out to be a disaster. The common practice is to compare the program as it has turned out with what it was originally envisioned as. Instead we should take it for granted that, given our nature, very few if any programs will work out swimmingly. The best basis for comparison is not the dreams of the founder, or even the blueprint, but something else that works. It will not do merely to establish that phonics does not work when it comes to teaching kids to read—tell me what works better. It will not do simply to criticize training programs for welfare mothers—show me something else that helps them find jobs. As long as the social goal at hand must be served, we must settle for the comparative best (which is often not so hot), rather than chase elusive perfection.

Be multi-faceted but not holistic. An official at the U.S. Department of Education has reported that the department is considering running an experimental program based on the hypothesis that whatever role modeling, encouragement, and assistance young students obtain in school is negated by life in drug-infested and poverty-stricken areas, by broken homes, by an environment of unemployment and asocial behavior. The goal is to discover whether granting these youngsters "everything"—providing their parents with jobs and housing, trying to change the neighborhood's culture, arranging for after-school tutoring, and so forth—will help them catch up with students from privileged areas and stay caught up.

On the one hand, the logic of such an approach is inescapable. The social world is one messy ball of wax; we can hardly hope to intervene in one strand of a person's life (say, schoolwork) and ignore all the others. This logic has led James Q. Wilson and other scholars to suggest that inner-city children should be taken from their troubled parents and put into kibbutzlike boarding schools.

On the other hand, the costs of such a holistic approach are so overwhelming that even if it works, it is impractical for the large number of people who need help. We must therefore search for approaches that acknowledge the ball-of-wax nature of many of our problems but are less exacting than the holistic approaches. This is the reason that prenatal care and vaccinations are so attractive as programs. These services can be provided if some attention is paid to other factors (treatment may need to be brought to mothers rather than requiring mothers to come to clinics), but it's not necessary to provide housing, full employment, and years of supplemented education just so that the programs will work (although these may be desirable in their own right).

It's no use pretending that poverty or welfare will be abolished, AIDS or cancer cured in this century, drug abuse or teen pregnancy sharply reduced. Let's instead dedicate our efforts to effective but clearly delineated projects in each of these areas. This humbler approach is likely to have a very attractive side effect: it may enhance public willingness to pay for such projects and may also restore public trust in our leaders and institutions.