Problem Set 3

11.126/11.249/14.48
Due April 24, 2007 in class

1 I. Class Size: Hanushek vs. Krueger (20 points)

Hanushek's literature review on the effects of class size suggests that there is no strong association between class sizes and test scores in the United States. Table 6 shows that across many (non-experimental) studies, small classes are associated with lower test scores as often as higher test scores. Table 7 shows that the same result holds for the relationship between class size and a student's *improvement*in test score over the course of the year. On the other hand, Krueger's paper on the Tennessee STAR experiment found large gains from small class sizes.

1. (10 points) Suppose that small class sizes really do have a substantial effect on student achievement, i.e., Krueger is right. How might you explain the patterns summarized by Hanushek?

   Answer: A few different answers are possible.

- Hanushek's data are not from an experiment and so they suffer from omitted variables bias. In particular, it is possible that weaker students tend to be assigned to smaller classes (in order to help them catch up). If these students are not only weaker to begin with but have less potential for improvement, then even the value-added model may give biased results.

- A permanent change in class size might have large effects on test scores because teachers can change their teaching style, even if a transient change in class size would not. In the STAR experiment, teachers were guaranteed small classes for several years, and this more closely mimics a program to reduce class sizes across an entire school district.

- Krueger found that small classes had a much more important effect in the first year than in subsequent years; this one-time kick might get lost in a value added model because most small-class students would have been in small classes in earlier years too. (This consideration can cut both ways, though–see below.)

2. (10 points) Now suppose that small class sizes don't have a very strong effect on achievement (on average), i.e., Hanushek is right. How might you explain the findings of the STAR experiment if the average effect of class size is negligible?

   Answer: Again, a few answers are possible.

- The STAR experiment might have suffered from Hawthorne effects (i.e., teachers in the treatment group were unusually motivated because they knew they were part of an experiment and wanted the researchers to find that small classes are a good thing). Krueger tries to test for this, but his test is at best indirect.
- The STAR experiment suffered from high attrition rates. This isn't a problem if the attrition is random–or unrelated to class size–but we might worry that parents were more likely to enroll their child in a different school if their child was assigned to a large class, and that more motivated, education-minded parents were particularly likely to take this step. If so, then the STAR data would be biased toward finding that small classes raise achievement (since high-scoring students would disproportionately leave the control group). Krueger tries to test for this kind of bias, but the test is imperfect.
- Krueger finds a large effect only for the first year. This finding is consistent with the conclusion that the average effect of class size is quite small (even if it is large for Kindergardeners who have never been in a small class before).

2 II. Vouchers and School Choice (25 points)

In the readings on school choice and in our discussions in class, the focus has always been on the effect of allowing a small number of children to enroll in the school of their choice (or their parents' choice). That is, we are looking at the effect of school choice on an individual student, holding the school system as a whole approximately constant. Imagine implementing a voucher program nation wide. Write a short essay discussing the additional effects such a program might have, effects that are not captured by small scale experimental studies. Do these effects make you more or less inclined to support a national voucher program?

Answer: Overall, people did really well on this question. There are various angles you could take, but here are some points that you could make:

- One broad channel through which voucher systems are supposed to improve outcomes is to reallocate students from poorly-performing schools to better schools. Studies of individual voucher recipients show (perhaps) that students who enter voucher schools do better, but it is not clear that this improvement can be scaled up. For example, many private schools have a cost advantage because their teachers are willing to work for lower pay–either because the students are easier to work with or because their work is partly an act of charity. However, a large-scale voucher program would increase the proportion of private school students who are "difficult", and the supply of teachers willing to work cheaply for non-economic reasons might be limited. Thus the cost advantage of private schools could

dry up once vouchers were implemented on a large scale. This consideration should make vouchers less attractive.

- The other main channel through which vouchers are supposed to affect outcomes is by giving each school incentives to improve. Studies of individual students who switch schools miss this effect entirely. By examining the difference in learning a student receives in the destination school (i.e., where students want to go) versus the source school, these studies miss any effect on the average quality of all schools. This consideration should make vouchers more attractive.

- Vouchers are also likely to affect the sorting of students across schools, although exactly how sorting will be affected depends on the details of the program. A plausible result, however, is that the "cream" of the poor schools will leave for better schools (either because private schools are more likely to accept these students or because students from these families are likely to be more motivated to take advantage of school choice). If students benefit from being around more intelligent and motivated peers, then this will be good for the students who move, bad for the students left behind in the poor schools, and possibly bad for the students at the destination schools (if the cream of the poor schools are still below average in the good schools). How this affects your opinion of vouchers is likely to turn largely on a value judgement: do you want to help bright and motivated students escape bad surroundings, potentially at the expense of the students with the lowest economic prospects?

- The characteristics of teaching as an occupation are likely to be affected strongly. Job security will decline and unions will be weakened, but there will be more variety in the types of teaching job available. The net effects on salaries is ambiguous. In the long run, these effects will change the number and mix of people who enter the teaching profession. (It is hard to be definite about the net direction of these effects, but the decline in job security is likely to reduce the number of people becoming teachers but raise their average quality–since good teachers have less need for tenure.)

- Geographic mobility might increase now that people no longer need to live in the same district where their children go to school. We should see less "Tiebout sorting" (where people with similar preferences for spending on a public good, like education, live together) and more sorting on other dimensions (preferences for parks, perhaps, or proximity to a job). While this effect is probably small, it makes a large voucher program more attractive.

3 III. Incentive and Accountability Systems (25 points)

1. (10 points) In 1996, Chicago Public Schools instituted an accountability reform. An important element of the reform was to put poorly-performing schools on probation. Probation means that the school loses some of its autonomy, and if it fails to improve staff can be dismissed or reassigned to other institutions; it is something that teachers and administrators would strongly like to avoid. The rule is that a school goes on probation if fewer than 15% of its students perform at or above national norms (the national median) on a reading test, with students in Special Education programs omitted from the calculation. Discuss some of the "perverse" incentives of this program, i.e., incentives that were probably not intended by the program designers. (I see four perverse effects.)

   Answer: I gave full credit for any four of the following (or a few other ideas that a small number of people mentioned):

- There is an incentive to prevent low-performing students from taking the test or having their test count. One way is to put them into Special Education.
- There is an incentive to "teach to the test"; here, that includes focusing on reading at the expense of other subjects.
- There is an incentive to focus resources on students who are "on the bubble", i.e., near the national norm. The school doesn't gain anything if it improves the scores of students who will easily pass or who are doomed to fail.
- Only schools where approximately 15% of students are performing above the norm have strong incentives. Schools where more than 25% of students perform above the norm don't have much to fear, even though they are still mediocre by national standards. Some schools could be in such rough shape that getting 15% of their students over the norm is an unattainable goal, in which case they have weak incentives too.
- Finally, there is an incentive to cheat outright, either by revealing the contents of the test ahead of time or by changing students' answers after the fact.

2. (15 points) Suppose you are designing an incentive system for schools based on their test scores. You can base your incentive scheme on one of three measures: a) the raw average of test scores in a school (normed to the population at large); b) a school's change in average test score compared to the same school's score last year; or c) a school's average test score minus the school's "predicted" score, where the predicted score comes from a regression of test scores on student Socio-Economic Status, race,

and other similar measures. Discuss the advantages and disadvantages of each option.

Answer:

- Option (a). The advantage of this approach is primarily that it is simple and transparent. The disadvantage is that average test scores are affected by many things other than school quality, like the home environments of the students. This is bad in two ways: 1) it is unfair, because some schools are punished for having a student body that is difficult to work with, and 2) it weakens incentive effects, because a "bad" school may not be able to beat a "good" school no matter how much effort it puts into improving its teaching quality.

- Option (b). This is still a fairly simple and transparent option, and it probably does a good job of correcting for other factors that affect test scores. (Because many of these factors could be hard to measure but are probably fairly constant from one year to the next, it may be even better than option (c) in this regard.) This option also gives the strongest incentives to schools that are currently underperforming, which could be seen as either an advantage or a disadvantage. Its primary disadvantage is that improvements in test scores this year raise the standard for next year; schools will trade off incentive benefits this year against lower benefits next year, and this will kill much of the incentive effect.

- Option (c). This should give a more accurate measure of each school's teaching quality than option (a), and it avoids the perverse dynamic incentives of option (b). However, it has three significant disadvantages:

  - It will be impossible to include all of the (non-school) factors affecting test scores in the regression model, so changes in incentives might be less related to changes in teaching quality than under option (b).
  - This option is by far the most complicated and least transparent. Staff and parents may fail to understand what makes one school get rewards when another school (perhaps with high raw test scores) does not. The formula used could also be open to charges of corruption (if a low-scoring school where the superintendant used to work gets substantial benefits, people will be suspicious–perhaps rightfully so!).
  - The bar is inherently set lower for schools with many students from traditionally disadvantaged backgrounds. You might be concerned that setting lower expectations for these students will be self-fulfilling, even if it is statistically correct that their test scores will be lower on average.

4 IV. Policy Brief (30 points)

You are an education aid to Janet Napolitano, Governor of Arizona. Napolitano is a moderate Democrat who respects both the government and markets. She believes strongly that Arizona's future depends on improving the quality of the state's K-12 education. She is considering proposing either a demonstration voucher program (which could include private schools) or a demonstration public school open enrollment program. She has not yet decided on the size of the program and she is still considering whether it should be run in one location or multiple locations. Before she makes these decisions, she wants to better understand what such a program might accomplish.

Earlier today, the governor called you into her office and asked you to write a short policy brief (approximately 600 words), based upon your reading, which summarizes what existing studies say about the effectiveness of a voucher program or an open enrollment program in raising student achievement and in improving school quality. If existing studies contain important lessons about how to design a demonstration, you should briefly include those lessons as well.

Answer: There isn't a "right" answer and some people took an unusual but reasonable tack on this question. However, typically I was looking for the following:

- A summary of the evidence on voucher effects from Milwaukee, both from the Greene, Peterson & Du paper and from Rouse's paper (you could also mention the original paper by Witte, Sterr & Thorn, but that wasn't necessary). You should mention each paper's findings and say something about why they come to differing conclusions with the same data. Ideally you would mention Rouse's observations on the P-5 schools and the possibility that the benefits of vouchers operate through reduced class size.

- A summary of the evidence from Chicago Public Schools' open enrollment program. This evidence is cleaner than the evidence from Milwaukee, in the sense that the authors could directly observe lottery participation, tracking lottery losers was less of an issue, and participation was much more widespread. The authors of the study found some evidence that lottery winners had gains on "non-academic" dimensions, but there was little evidence of test score gains.

- Some discussion of how to design a demonstration program. Because this is a demonstration program (which may or may not be brought to scale), it is important to design it with an eye to evaluating its effects. Some points you could make:

  - Evaluation is easier if you hold lotteries to ration spots in oversubscribed schools. Arguably this is also a fairer mechanism than allowing receiving schools to cherry pick the best applicants. (If we are introducing a voucher program rather than an open enrollment program, you could also randomly hand out vouchers to a subset of

disadvantaged children–but you would get a lot of complaints about unfairness.)

- As best you can, it is important to track lottery losers wherever they go. You would like to test lottery losers even if they enroll in a non-choice private school (or an out-of-state school if that is feasible). The fact that outcome data was missing for a non-random subset of the lottery losers was probably the biggest flaw of the Greene, Peterson & Du paper.
- It helps to have data on the same children before and after they enter their new school. This means that you want to be collecting data before the demonstration program is implemented and that you want to track individual students over time.
- You might want to make an effort to capture some of the large-scale effects from problem II. This is hard; the right way to do it would be to randomly assign a subset of "bad" schools to have their students participate and randomly assign a subset of "good" schools to be eligible for receiving students, but that's probably not politically feasible. Still, you might want to allow a large proportion of the students in some selected area to participate; then you could at least gather case study evidence on how the "bad" schools in this area responded. Of course, these schools will resist allowing a substantial proportion of their students to leave.

Other things you could talk about: theoretical effects you might expect a school choice program to have, how to design the program to be as fair and effective as possible, or the relative merits of vouchers and open enrollment programs.