Dissertation Review
"Essays on Incentives and Information in Schools" by Dagmara Celik Katreniak
Reviewer: Karna Basu

Introduction

Dagmara's dissertation is well-written, innovative, and pertinent. The fieldwork undertaken is impressive. And the empirical results add to our understanding of outstanding questions in behavioral economics, development economics, and the economics of human capital. Below, I discuss several aspects of her dissertation, with my observations and some suggestions for improvement.

Taken together, the two chapters of her dissertation aim to examine student performance in Ugandan schools. Within, there are a number of questions of both academic and policy relevance. How can student performance be improved? What is the effect of performance incentives on student happiness? Do students correctly assess their own abilities? How can the accuracy of students' self-assessment be improved?

The dissertation provides some answers to the above questions, and more. In many cases the empirical results are thought-provoking and suggest some natural directions for future research.

Main results

Chapter 1 focuses on the main randomized experiment, which examines determinants of student performance and happiness. Treatments varied along two dimensions. First, students received differing levels of feedback: no feedback, group-level feedback (relative to other groups within a class), and class-level feedback (relative to other classes). Second, students received differing levels of rewards for performance: no rewards, financial rewards, and reputational rewards. The main outcomes to be measured were performance on exams (math and English) and happiness.

This amounts to a very rich experimental setup as it allows us to gain several insights about what *combinations* of factors affect student outcomes. The first set of treatments is purely about information provision. The second set involves what economists typically consider standard incentives. Since these treatments were layered on top of one another, we learn about the interaction of information and incentives.

Dagmara has carried out a thorough empirical analysis which looks at standard treatment effects as well as heterogeneous treatment effects by slicing the population along many lines of interest. Treatment effects generally have the expected signs. But there are a number of nuanced and intriguing results. First, treatments were most potent in combination—in

particular, feedback combined with financial rewards had strong positive effects on performance in both subjects. This, I think, is quite a major finding as it helps unpack why, in past studies, the impacts of student incentives have sometimes been ambiguous.

Second, math performance seemed more responsive to treatments than did English performance. Third, some treatments lowered student happiness, which might be related to changes in stress levels. And fourth, there were a number of interesting heterogeneous treatment effects, especially across gender and initial student quality. I will comment further on these later in the document.

Chapter 2 is concerned with students' self-assessment. This topic links nicely to Chapter 1—a treatment that offers feedback is effective only if students lack precise information about their own skills. The fact that the feedback treatments had an effect suggest that students may not have an accurate picture of their academic standing. Chapter 2 confirms that.

First, Dagmara finds that students are generally overconfident, as seen in past studies. Second, feedback improves students' self-assessment, but students remain overconfident. Third, there are again heterogeneous treatment effects.

Comments

My comments are divided into three categories. First, I discuss where the papers fit into the academic literature. Second, I list specific questions that emerged during my reading of the dissertation. Third, I make some suggestions for a unifying conceptual framework. None of the suggestions below is critical, but I leave them for the author to consider.

On the first point: I expect the papers in this dissertation will soon be essential reading in the literature on student performance and human capital building. The experiments were beautifully constructed, the data analysis is clean, and the results nicely complement and add to past studies. In fact, for this reason I would encourage Dagmara to place her papers in a larger context. Her work focuses on interventions *on* the students themselves. There is also a large literature on other kinds of interventions aimed at improving student performance. While Dagmara has included quite acceptable literature reviews, here are some suggestions for additional papers that emphasize *teacher* interventions as ways to improve student performance: Fryer, Levitt, List, & Sadoff (2012); Muralidharan (several papers); Duflo, Hanna, & Ryan (AER, 2012). These papers all present very interesting angles on ways to induce teachers to both show up more and to teach better. It would be worth thinking open-endedly about how student-side and teacher-side interventions might interact with each other. Finally, Abhijit Banerjee has an interesting chapter on these issues in his edited book, "Making Aid Work".

In Chapter 2, I would be interested to learn more about how overconfidence is linked to outcomes in other settings. I'm not as familiar with this topic, but one reasonable place to start would be with the works of Daniel Kahneman.

On my second point, here are some questions that directly relate to the dissertation:

- How lasting should we expect treatment effects to be? In particular, if feedback fails to close the overconfidence gap, do we expect repeated feedback to continue improving performance?
- Should education policymakers care about student happiness? Is stress perhaps a sensible motivator, and couldn't we reasonably conjecture that there is a tradeoff between short-term and long-term happiness? How might this affect Dagmara's policy conclusions?
- Since randomization occurred at the school level, some more discussion on sample size and power would help.
- In the across-group feedback treatment, students were given enough information to construct their group's average score. But in the across-class treatment, students were told their class' rank compared to other classes, but not the average score for their class (this is my understanding based on my reading). Why was this the case? In general, it would be worth including some greater explanations for the *specifics* of the treatments.
- The gender differences in treatment effects might be nicely related to Hoff & Pandey (AER, 2006).
- The English vs Math discussion on page 44 is not clear.
- The group composition results are very interesting and have implications for how workers etc should be organized in production (positive assortative vs negative assortative matching).
- Result 12 in Chapter 2 is very intriguing and deserves a more detailed discussion.

Finally, my third point: I would strongly urge Dagmara to bring in a simple theoretical framework, at least before submitting the papers for publication. There are several mechanisms that the experiments explore, and the papers would be much stronger if the reader were given a conceptual guide to thinking through the mechanisms.

Here's what I have in mind. Let students have some outcome, Y, that they wish to maximize. Y is a noisy function of two variables—relative performance and absolute performance. Performance depends on effort, which is costly. Let there be some distribution of student types, where each student has, say, a different marginal cost of performance. This benchmark model will be very easy to solve, yielding some optimal level of effort for each student.

Now it will be possible to rigorously think about how different hypotheses and interventions relate to the model. Under the above model, feedback should have no effect since it is assumed that students know their own cost functions. So the only intervention that should matter involves rewards, which is equivalent to making Y more sensitive to performance.

So, since we know that feedback matters, we can reject the benchmark model. Now, consider a modified model in which students assess their performance-cost function wrong. Overconfidence implies that students think their cost is lower than it really is. Now, think about

how feedback changes this assessment, and how this in turn changes equilibrium levels of performance.

Apart from helping clarify the existing empirical results, I think a model like this will also immediately suggest some ideas for additional heterogeneous treatment analysis. This would make the papers strong contenders for publication in the top journals.

Conclusion

I enjoyed reading this dissertation. In my opinion it satisfies the requirements for a PhD in Economics, and I recommend proceeding with the defense. My best wishes to Dagmara as she continues her research.