

Response to "Does Teacher Preparation Matter?"

Michael Podgursky

*Middlebush Professor of Economics and Department Chair
University of Missouri - Columbia*

Professor Linda Darling-Hammond and collaborators attempt to estimate the effect of Teach For America (TFA) teachers versus non-TFA teachers with various types of certification on several measures of grade 4 and 5 student achievement in the Houston Independent School District. They use student-level longitudinal achievement data that links students to the relevant math or reading teacher. For the most part they find that TFA teachers are less effective than teachers with traditional certification.

Darling-Hammond and authors focus their criticism on a study done by CREDO using similar data several years ago. However, they devote less than two sentences (pp. 2-3) to the major study in this area conducted by Mathematica.¹ The researchers do not seem to appreciate, or at least do not acknowledge, that their findings, which are seriously at variance with those in the Mathematica study, are also based on inferior research methodology.

The Mathematica study was a carefully designed experiment in which students were randomly assigned to teachers within schools. The Mathematica researchers studied TFA and a control group of non-TFA teachers at six sites around the country (including Houston) and found positive effects of TFA teachers in math, even when compared with experienced certified teachers.² They also found positive, but generally insignificant effects, on reading.

Why are the Mathematica results more reliable? Since teachers are not randomly assigned to schools, and students are not randomly assigned to teachers within schools, it is important to carefully account for both of these effects when studying the effect of any teacher characteristic on student achievement. That's why a well-designed random assignment study like Mathematica's always trumps a non-experimental study like that of Darling-Hammond et al.

Suppose that TFA and other novice teachers are equally effective, but that on average the TFA teachers are assigned to the most difficult schools and are given the most challenging students within these schools. Then a non-experimental study like Darling-Hammond's is likely to reach the incorrect conclusion that TFA teachers are less effective teachers

¹ See the report at: http://www.teachforamerica.org/documents/mathematica_results_6.9.04.pdf

² Darling-Hammond, et. al write: "... in both of these studies [CREDO and Mathematica] the comparison group teachers were disproportionately untrained and unlicensed teachers. Neither of these studies explicitly compared TFA teachers to teachers with standard training and certification, controlling for other student, teacher, and school variables." (pp. 2-3). This is incorrect, see table VI.2 p.33 of the Mathematica report with the URL in the text.

Why would this be the case? Don't the "controls" in the Darling-Hammond study like prior student achievement and student free and reduce-price eligibility take account of non-random assignment of students to teachers? Unfortunately, there is no reason to believe that they do a sufficiently good job to prevent bias in estimating teacher effects. Consider student socioeconomic status (SES). Many studies have documented the powerful effect SES has on student achievement. Yet in this study, the only control for student SES is free and reduced lunch eligibility, a simple binary indicator (yes/no), measured with a great deal of error. A student with an intact and supportive but low income two-parent family will get a "1" along with a student whose mother is a crack addict. Clearly, this crude variable has not controlled for the large difference in human capital investments accruing to children from these two very different households. Moreover, Darling-Hammond et.al. use only a single year of prior test score data (more on this below). Since test scores have a great deal of measurement error, they have only very crudely measured the achievement status of the students prior to entering the classroom. Thus, SES is not fully captured by a lagged test score either.

For these reasons and many more, from a scientific standard it is much better to have a study design with random assignment to "treatment" and "control" groups than to rely on data in which students and teachers are matched in non-random ways that are not known and only crudely controlled for by the researcher.

Thus, one well-designed random assignment study such as the Mathematica's, trumps 50 or 250 non-experimental studies if there are systematic (but unmeasured) patterns in the way TFA teachers and students are matched.

(If true experiments are not available, in policy analyses economists try to look for "natural experiments" in which some exogenous phenomenon yields an approximation to an experiment. For example, economists have examined the student achievement effect of teacher professional development and class size with these methods.)

However, even by the standards of the non-experimental longitudinal research literature the Darling-Hammond study falls short. First, it is possible using their data to compare TFA and non-TFA teachers within the same school—but they do not. (In technical terms, they fail to include a "school effect.") Clearly, this would be a much better control for socioeconomic characteristics of the community than the covariates they use. In addition, their statistical methodology fails to make use of the fact that they have repeated test scores for each student. They potentially have three test scores (grades 3-5) for each student for each type of test (e.g., math, English), as well as several different tests per student, yet they make no effort to use this information in estimating the TFA effect. The simplest and most commonly used technique would be to include a "student effect" for each student in their sample. Mixture models such as those employed by William Sanders utilize the full set of information from all the tests in estimating teacher effects. At a minimum, using a more complete profile of tests would give us a better measure of the baseline academic achievement of students before they are exposed to TFA versus non-TFA teachers.