Available online at www.sciencedirect.com
ScienceDirect
Economics of Education Review 26 (2007) 673–682
Economics of Education Review
www.elsevier.com/locate/econedurev

Teacher credentials and student achievement: Longitudinal analysis with student fixed effects

Charles T. Clotfelter, Helen F. Ladd*, Jacob L. Vigdor

Sanford Institute of Public Policy, Duke University, Durham, NC 27708, USA

Abstract

We use a rich administrative dataset from North Carolina to explore questions related to the relationship between teacher characteristics and credentials on the one hand and student achievement on the other. Though the basic questions underlying this research are not new—and, indeed, have been explored in many papers over the years within the rubric of the "education production function"—the availability of data on all teachers and students in North Carolina over a 10-year period allows us to explore them in more detail than has been possible in previous studies. We conclude that a teacher's experience, test scores and regular licensure all have positive effects on student achievement, with larger effects for math than for reading. Taken together the various teacher credentials exhibit quite large effects on math achievement, whether compared to the effects of changes in class size or to the socio-economic characteristics of students.
Published by Elsevier Ltd.

Keywords: Education production function; Teacher quality; Teacher credentials

1. Introduction

Education researchers and policy makers agree that teachers differ in terms of quality, and that quality matters for student achievement. Despite extensive research, however, debate still rages about whether measurable teacher credentials can reliably predict either teacher quality or student achievement. We shed new light on this issue by using rich administrative data from North Carolina to explore a range of questions related to the relationship between teacher characteristics and credentials on the one hand and student achievement on the other. The teacher credentials in which we are most interested are those that can be affected in one way or another by policy.

This paper builds on our previous cross-sectional research on teacher credentials and characteristics (Clotfelter, Ladd, & Vigdor, 2006), but differs in its use of longitudinal data. These data include all North Carolina students in grades 3, 4 and 5 in years 1995–2004 for whom we can identify their teachers of math or reading. The longitudinal aspect of the data allows us to include in our models student fixed effects, which provide powerful protection against the left-out variable bias that typically plagues research of this type. Such data also permit us to explore in some detail the mechanisms through which teacher credentials exert their impacts.[1]

*Corresponding author.
E-mail address: helen.ladd@duke.edu (H.F. Ladd).

[1]Given the space constraints for this article, we are not able to describe all the details of the analysis. For those details and a broader set of results, the reader should refer to the longer version of this paper (Clotfelter, Ladd, & Vigdor, 2007).

0272-7757/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.econedurev.2007.10.002

2. Empirical framework

Although we focus here only on the findings related to teacher credentials, all the findings emerge from fully specified models of student achievement estimating using student level data. In recognition of education as a cumulative process, the standard starting point in the literature is a "value-added" model in which the learning that a student (denoted by i) brings to the classroom in year t is incorporated in the form of her achievement in the relevant subject in the previous year.[2] Specifically, we can write:

$$A_{it} = \alpha A_{it-1} + \beta_1 \mathrm{SCF}_i + \beta_2 \mathrm{SCV}_{it} + \sum_j \left[\beta_3 \mathrm{TCF}^{(j)} + \beta_4 \mathrm{TCV}_t^{(j)} + \beta_5 C_t^{(j)}\right] D_{it}^{(j)}, \quad (1)$$

where the relevant variables or vectors of variables are defined as follows:

A_{it} is achievement of student i in year t as measured by a normalized test score in reading or math.

A_{it-1} is achievement of the ith student in the prior year.[3]

TCF is a vector of teacher characteristics, such as the teacher's race and gender, that are fixed over time for any specific teacher.

TCV_t is a vector of teacher characteristics that vary over time, including, for example, years of teaching experience, attainment of higher degrees, or attainment of a particular type of license.

C_t is a vector of classroom characteristics that vary depending on the student's classroom each year. These include class size and characteristics of peers.

SCF_i is a vector of measurable student characteristics that are fixed over time, such as a student's race, gender, and age in grade 3.

SCV_{it} is a vector of student characteristics that vary over time. These include indicator variables for thing such as grade repetition or movement to a new school.

u_{it} is an error term.

In addition, $D_{it}^{(j)}$ is an indicator variable for whether the student had the jth teacher in year t. The coefficients β_1–β_5 are vectors rather than individual parameters, and α represents the extent to which knowledge persists from one year to the next.

As a model for estimating the effects of teacher credentials on student achievement (denoted by the coefficient vectors β_3 and β_4), Eq. (1) is flawed in at least two ways. First and most important, the relevant coefficients will be biased because, as we have documented in our prior research, teachers with stronger credentials tend to be matched at both the school and the classroom level with students who are more educationally advantaged (Clotfelter et al., 2006). That positive matching means that too much of the achievement of high achieving students would be attributed to the credentials of the teachers rather than to the unobserved characteristics of the students. Second, the inclusion of the lagged achievement variable on the right hand side of the equation is a problem, both because the variable is likely to be measured with error and because any correlation of achievement over time would make the variable endogenous.[4]

We address these two biases in two ways. First, we make use of our longitudinal data by replacing all the student-specific variables that do not vary over time with student fixed effects. In other words, we allow for the intercepts of the achievement equation to vary by student. Provided the effects are linear, the inclusion of these student fixed effects eliminates any bias associated with the nonrandom matching of teachers and students. This conclusion follows because their presence means that the only variation used to estimate the coefficients of interest is variation within, not across, individual students.

Second, to address the statistical problems that arise when the lagged achievement variable is included as a control variable on the right hand side of the equation, we either delete that variable or we move it to the left hand side by respecifying the dependent variable as a student's gain in achievement (i.e. as $A_{it} - A_{it-1}$). Neither solution is perfect.

[2] Hanushek (1997) and Hedges, Laine, and Greenwald (1994) provide contrasting summaries of the research literature on how teacher credentials and other inputs affect student achievement. Boardman and Murnane (1979) and Todd and Wolpin (2003) provide thoughtful analyses of model specification.

[3] For simplicity, the lagged achievement term refers here to the same subject as the dependent variable. In other analysis not reported here, we have included lagged achievement terms for both math and reading as well as squared terms for each of them, on the ground that prior achievement in both math and reading could affect current year achievement in either subject.

[4] Though the dependent variable is subject to measurement error as well, that error simply shows up in the error term and is not a source of bias.