Journal of Public Economics 95 (2011) 798–812

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Teacher training, teacher quality and student achievement

Douglas N. Harris [a], Tim R. Sass [b,*]

[a] Department of Educational Policy Studies and La Follette School of Public Affairs, University of Wisconsin, Madison, 213 Education Building, Madison, WI 53706, United States
[b] Department of Economics, Florida State University, 288 Bellamy Building, Tallahassee, FL 32306, United States

ARTICLE INFO

Article history:
Received 6 February 2007
Received in revised form 10 November 2010
Accepted 16 November 2010
Available online 25 November 2010

Keywords:
Student achievement
Teacher quality
Teacher training

ABSTRACT

We study the effects of various types of education and training on the productivity of teachers in promoting student achievement. Previous studies on the subject have been hampered by inadequate measures of teacher training and difficulties in addressing the non-random selection of teachers to students and of teachers to training. We address these issues by estimating models that include detailed measures of pre-service and in-service training, a rich set of time-varying covariates, and student, teacher, and school fixed effects. We find that elementary and middle school teacher productivity increases with experience (informal on-the-job training). The largest gains from experience occur in the first few years, but we find continuing gains beyond the first five years of a teacher's career. In contrast, we do not find a consistent relationship between formal professional development training and teacher productivity. However, this may be partly driven by estimation issues as we find more significant positive effects of formal training in the subject-grade combination where estimates should be most precise (middle school math). There is no evidence that teachers' pre-service (undergraduate) training or college entrance exam scores are related to productivity.
© 2010 Elsevier B.V. All rights reserved.

1. Introduction

It is generally acknowledged that promoting teacher quality is a key element in improving primary and secondary education in the United States. Indeed, one of the main goals of recent presidential administrations has been to have a "highly qualified teacher" in every classroom. While recent research has documented the central role of teacher quality in promoting student achievement, there is no consensus on what factors enhance, or even signal, teacher quality. This has fueled debate over how best to prepare new teachers and how to improve the quality of the existing teacher labor force. To better understand the determinants of teacher quality, we consider the relationship between teacher productivity and teacher training, including formal pre-service university education, in-service professional development, and informal training acquired through on-the-job experience.

Uncertainty regarding the effects of teacher training is due in large measure to three methodological challenges in estimating the effects of training on teacher quality. First, it is difficult to isolate productivity, especially in teaching where a student's own ability, the influences of a student's peers, and other characteristics of schools also affect measured outcomes. The problem is exacerbated by the fact that assignment of students and teachers to classrooms is usually not random, leading to possible correlations between observed teacher attributes and unobserved student characteristics. Second, as in other occupations, there is an inherent selection problem in evaluating the effects of education and training on teacher productivity. Unobserved teacher characteristics, such as motivation or intelligence, may affect the amount and types of education and training they choose to obtain as well as subsequent performance of teachers in the classroom. Third, it is difficult to obtain data that provide much detail about the various types of training teachers receive and even more difficult to link the training of teachers to the achievement of the students they teach. Addressing all of these issues in a single study presents significant data and estimation challenges.

In this paper we address these challenges and present new evidence on the effects of human capital obtained before entering the teaching profession (pre-service education) and after entry. The latter includes in-service professional development, advanced degrees, and informal training acquired through experience. We utilize a rich statewide administrative database from Florida that allows us to tie student performance to the identity of their classroom teacher at all grade levels and in turn link teachers to their in-service training, their college coursework and majors, and their pre-college entrance exam scores. These data provide an opportunity to analyze the effects of both pre-service and in-service training on teacher productivity while addressing the twin selection problems associated with teacher acquisition of training and assignment of students to teachers.

Our analysis proceeds in two steps. First, we estimate student achievement models that include a substantial set of covariates that measure the time-varying observable characteristics of individual students, their classroom peers, and their school's principal. We also incorporate multiple levels of fixed effects that control for unmeasured

* Corresponding author. Tel.: +1 850 644 7087.
E-mail addresses: dnharris3@wisc.edu (D.N. Harris), tsass@fsu.edu (T.R. Sass).

0047-2727/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpubeco.2010.11.009

time-invariant student, teacher and school characteristics. This first-stage model includes detailed information on the quantity and characteristics of education and training teachers receive after they have entered the classroom, including both formal graduate education and professional development courses sponsored by schools and school districts. We also include measures of teacher experience, which represent informal on-the-job training. This first step yields estimates of the fixed effect for each teacher, which represents the teacher's contribution to student achievement or "value added" that does not vary over her career. In the second step we take the estimated teacher fixed effect and regress it on characteristics of teachers' (time-invariant) undergraduate coursework, controlling for teacher pre-college cognitive/verbal ability with college entrance exam scores.

We begin in Section 2 by describing past literature on teacher training. Our data and methodology are discussed in Sections 3 and 4, respectively. Our results, presented in Section 5, suggest that only two of the forms of teacher training influence productivity; content-focused teacher professional development is positively associated with productivity in middle and high school math and on-the-job training acquired through experience is correlated with enhanced effectiveness in teaching both math and reading in elementary and middle schools. The implications of our findings are discussed in Section 6.

2. Previous literature on the effects of teacher training

In early work on teacher productivity, researchers estimated education production functions by regressing aggregate student achievement levels on measures of teacher training and various other controls using cross-sectional data (see review by Hanushek (1986)). A subsequent generation of studies used student-level two-year test-score gains and richer sets of teacher training variables to evaluate the impact of teacher training on student achievement. The state of the literature through the year 2000 has been extensively reviewed by Wayne and Youngs (2003) as well as by Rice (2003), Wilson and Floden (2003), and Wilson, et al. (2001). Rather than duplicate previous surveys we highlight new research findings over the last decade.

One of the main shortcomings of earlier generations of research is their reliance on observed student characteristics or "student covariates" to account for student heterogeneity. There is evidence that better trained and more experienced teachers tend to be assigned to students of greater ability and with fewer discipline problems (e.g., Clotfelter et al. (2006), Feng (2009)). Given this positive matching between student quality and teacher training, the student-covariate studies' lack of controls for unobserved student characteristics would tend to upwardly bias estimates of teacher value-added associated with education and training.

Since 2000, eight studies of teacher productivity in the U.S. have addressed this selection-bias problem using student fixed effects to account for unobserved heterogeneity. Five other recent studies exploit either experiments with random assignment of students and teachers or situations where there is "apparent random assignment" (based on observable student characteristics) or "natural" experiments where assignment is based on exogenous factors. Table 1 provides a summary of the recent research on teacher training and productivity, broken down by the methodology employed. No matter what the empirical strategy, nearly all of the recent studies of teacher productivity include some measure of teacher experience, which serves as a proxy for on-the-job training. Results for elementary math are about evenly split between positive and insignificant effects of teacher experience on student achievement. In contrast, all but one of the eight recent studies that separately analyze elementary reading find that student achievement is positively correlated with teacher experience. At the middle school level the subject-specific findings are essentially reversed. Studies that include middle school consistently find positive effects of teacher experience on math achievement whereas the findings for the effects of experience on middle school reading achievement are evenly split between positive and insignificant correlations. The three studies of high school teachers yield conflicting results.

As discussed by Rockoff (2004) and Kane et al. (2006), the estimated effects of experience may be biased if sample attrition is not taken into account. For example, less effective teachers might be more likely to leave the profession and this can give the appearance that experience raises teacher value-added when, in reality, less effective teachers are simply exiting the sample. Alternatively, selection could work in the opposite direction; more able teachers with higher opportunity costs may be more likely to leave the profession, leading to a spurious negative correlation between teacher experience and student achievement. One method of addressing the attrition issue is to include a teacher-specific effect, to control for unmeasured teacher ability, along with the experience measures. The teacher-specific effect should purge the influence of teacher time-invariant ability on experience, yielding unbiased estimates of the marginal product of experience.[1] While the recent student-covariate studies all include a teacher-specific effect, only two of the student-fixed-effects studies, Hanushek et al. (2005) and Rockoff (2004), employ teacher fixed effects in addition to student fixed effects. Both of these studies analyze only a single school district. In our work we are able to include both student and teacher fixed effects using data for the entire state of Florida.

In addition to experience, the other commonly measured aspect of teacher training is the attainment of graduate degrees. Nearly all of the recent student-fixed-effects and random-assignment studies include a measure of post-baccalaureate degree attainment, typically whether a teacher holds a master's degree. Except for positive correlations between possession of a masters degree and elementary math achievement found by Betts et al. (2003), Dee (2004) and Nye et al. (2004), recent research indicates either insignificant or in some cases even negative associations between possession of graduate degrees by a teacher and their students' achievement in either math or reading.

In contrast to experience and possession of advanced degrees, the pre-service undergraduate training of teachers has received much less attention in the recent literature. Of the 18 studies we review, only seven investigate the relationship between pre-service education and subsequent teacher productivity. Two studies consider the college major of teachers and neither finds a positive and significant relationship between major choice and teacher productivity (Aaronson et al. (2007); Betts et al. (2003)). Similarly, most studies find no significant correlation between college selectivity and later effectiveness as a teacher (Kane et al. (2006); Clotfelter et al. (2006, 2007)). One study, Clotfelter et al. (2010), does find a positive and significant relationship between the prestige of the undergraduate institution and productivity of high school teachers. Among education majors, one study finds a significant relationship between productivity and some specific elements of teacher preparation programs (Boyd et al., 2008).

There are at least two shortcomings of recent estimates of the impact of undergraduate education on teacher productivity. First, most recent work has relied on relatively gross measures, like college major, which may obscure significant variation in college coursework. Even the Boyd, et al. work measures preparation program features

[1] While the inclusion of teacher effects greatly reduces the potential bias associated with teacher attrition, it does not necessarily eliminate it for two reasons. First, since multiple observations are required to compute teacher effects, elementary school teachers who leave after one year are necessarily excluded. This is not a significant problem for middle and high-school teachers, however, since they teach multiple classes within a single period (though it remains a problem for estimating the effects of experience, which can still only be done for teachers with two or more years in the classroom). Second, if there is an unobserved time-varying component of teacher productivity that is correlated with the likelihood of attrition, then this will not be fully captured by the teacher effect. For example, as noted by Murnane and Phillips (1981) and others, the presence of young children in the teacher's home may lower teacher productivity and also increase the likelihood of attrition. We test whether teacher-specific effects eliminate attrition bias in our empirical work below.

Table 1
Results of recent studies of the effects of teacher training on student achievement in the united states, by teacher training type.

Type of training				
Method/studies	Undergraduate studies	Graduate degrees	In-service training	Experience
Gain score with student covariates				
Aaronson et al. (2007)	Major (MH 0)			(MH 0)
Hill et al. (2005)				(ME 0)
Kane et al. (2006)	GPA (ME/MM 0, RE/RM 0) College Selectivity (ME/MM 0, RE/RM 0)			(ME/MM ++, RE/RM ++)
Boyd et al. (2008)	Preparation Program Characteristics (ME/MM Mix, RE/RM Mix)			
Panel data with student fixed effects				
Betts et al. (2003)	Major (All Mix)	MA (ME +, RE 0, MM 0, RM 0, MH 0, RH ++)		(ME 0, RE 0, MM −, RM 0, MH 0, RH 0)
Boyd et al. (2006)				(ME/MM ++, RE/RM ++)
Clotfelter et al. (2007)	Univ. Prestige (ME +, RE 0)	MA (ME 0, RE− −)		(ME ++, RE ++)
Clotfelter et al. (2010)	Univ. Prestige (CH ++)	MA (CH 0)		(CH ++)
Hanushek et al. (2005)		MA (ME 0, MM 0)		(ME +, MM +)
Jepsen (2005)		>BA (ME 0, RE 0)		(ME +, RE +)
Rivkin et al. (2005)		MA (MC 0, RC 0)		(MC +, RC 0)
Rockoff (2004)		MA (ME 0, RE −)		(ME 0, RE ++)
Random assignment and "natural experiments"				
Clotfelter et al. (2006)	Univ. Prestige (ME 0, RE 0)	MA (ME − −, RE − −)		(ME ++, RE ++)
Dee (2004)		MA (ME +, RE 0)		(ME 0, RE ++)
Ding and Lehrer (2005)		MA (ME 0, RE 0)		(ME 0, RE +)
Garet et al. (2008)			(RE 0)	
Garet et al. (2010)			(MM 0)	
Jacob and Lefgren (2004)			(ME 0, RE 0)	
Nye et al. (2004)		MA (ME +, RE 0)		(ME +, RE +)

Each cell starts by listing the specific variable under consideration, except for the last two columns where in-service training and experience are defined the same way across studies. Effects on student achievement are given in parentheses. The first letter indicates the subject area: M = math, R = reading, C = combined across subjects. The second letter indicates the grade level: E = elementary, M = middle school, H = high school, and C = combined across elementary and middle. This is followed by information regarding the effects of the specified variable on student achievement scores in the previously specified subject and grade in the preferred specifications: ++ = positive and significant in nearly all preferred specifications; + = often positive and significant; 0 = insignificant; − = often negative and significant; − − = negative and significant in nearly all preferred specifications; and Mix = mix of positive/significant and negative/significant.

rather than actual courses taken.[2] Second, none of the recent studies that include measures of undergraduate training control for the pre-college ability of future teachers. In our work we consider the specific courses taken by teachers and control for pre-college ability with college entrance exam scores.

The only recent studies of the impact of in-service professional development (PD) on teacher productivity in the United States are Jacob and Lefgren (2004), Garet et al. (2008) and Garet et al. (2010).[3] Jacob and Lefgren exploit a "natural experiment" that occurred in the Chicago public schools where low-performing elementary schools (based on students' failing to meet national norms on a reading exam) were placed on probation and given resources to purchase additional staff development services. Employing a regression discontinuity design, they find that the additional professional development resulting from probation had no effect on student achievement in either math or reading. Garet et al. (2008) utilize an experimental design to study the impact of an intensive PD program for early reading teachers and subsequent follow-up with in-school coaches. While the PD intervention increased teacher knowledge and changed instruction, neither the PD program alone nor the PD program with subsequent coaching yielded improvements in student reading scores. Similar to Garet et al. (2008), Garet et al. (2010) conduct a randomized controlled trial of the efficacy of a professional development program for middle school math teachers designed to improve teaching of rational numbers. The program was associated with increased teacher knowledge of rational numbers, though the effect was significant at only an 85% confidence level. Like the earlier reading study, the math PD intervention produced significant changes in teacher instructional practice, but did not yield any statistically significant improvement in student achievement. While methodologically strong, the Jacob and Lefgren (2004), Garet et al. (2008) and Garet et al. (2010) studies consider PD in relatively narrow contexts and thus provide limited insight on the efficacy of PD in more general settings. In the present study we expand on their work by considering PD received by teachers in all schools at all grade levels and distinguish between training that focuses on content and that which emphasizes pedagogy.

3. Data

To study the effects of teacher training we make use of an extensive panel data set of school administrative records from Florida.[4] The data cover all public school students throughout the state and include student-level achievement test data for both math and reading in each of grades 3–10 for the years 1999–2000 through 2004–2005.[5] Summary statistics are available in Table 2.

[2] At least two of the older gain-score studies, Eberts and Stone (1984) and Monk (1994) do include detailed measures of courses taken.

[3] Angrist and Lavy (2001) analyze the effects of teacher professional development on teacher productivity in Israel.

[4] A more detailed description of the data is provided in Sass (2006).

[5] Until recently, the state of Florida administered two sets of reading and math tests to all 3rd through 10th graders in Florida. The "Sunshine State Standards" Florida Comprehensive Achievement Test (FCAT-SSS) is a criterion-based exam designed to test for the skills that students are expected to master at each grade level. The second test is the FCAT Norm-Referenced Test (FCAT-NRT), a version of the Stanford Achievement Test used throughout the country. Version 9 of the Stanford test (the Stanford-9) was used in Florida through the 2003/2004 school year. Version 10 of the Stanford test (the Stanford-10) has been used since the 2004/05 school year. To equate the two versions of the exams we convert Stanford-10 scores into Stanford-9 equivalent scores based on the conversion tables in Harcourt (2002). The scores on the Stanford-9 are scaled so that a one-point increase in the score at one place on the scale is equivalent to a one-point increase anywhere else on the scale. The Stanford-9 is a vertically scaled exam, thus scale scores typically increase with the grade level. We use FCAT-NRT scale scores in all of the analysis. The use of vertically scaled scores to evaluate student achievement is important since a one-unit change has the same meaning for low- and high-achieving students.

Table 2
Summary statistics for Florida public school students and teachers, 1999/2000–2004/2005.

	Math			Reading		
	Elementary (Grades 3–5)	Middle (Grades 6–8)	High School (Grades 9–10)	Elementary (Grades 3–5)	Middle (Grades 6–8)	High School (Grades9–10)
In-service (student-level) variables						
Achievement gain	18.783	13.635	9.812	16.663	16.500	−1.096
Std. dev. of achiev. gain	25.718	23.212	26.510	26.601	25.545	26.039
Achievement level	649.597	676.771	722.819	655.110	698.568	709.156
Std. dev. of achiev. level	37.632	37.615	39.042	38.728	38.996	34.040
Number of schools attended	1.031	1.032	1.017	1.032	1.026	1.017
“Structural” mover	0.009	0.248	0.297	0.009	0.174	0.352
“Non-structural” mover	0.102	0.133	0.138	0.104	0.118	0.148
Fraction female peers	0.499	0.493	0.515	0.499	0.513	0.523
Fraction Black peers	0.212	0.236	0.178	0.222	0.197	0.161
Fraction mover peers	0.142	0.409	0.419	0.143	0.325	0.515
Fraction “strc.-mover” peers	0.001	0.250	0.265	0.009	0.177	0.340
Average age of peers (mo.)	121.392	150.925	178.634	121.444	153.170	181.401
Average class size	25.318	27.097	27.330	25.386	25.973	27.364
Teacher experience	11.053	9.344	11.552	11.100	9.980	10.679
Total in-service hours	53.579	44.866	36.808	51.543	53.652	43.510
Content in-service hours	19.819	12.941	14.401	20.143	21.548	17.224
Other in-service hours	33.760	31.945	22.408	31.400	32.105	26.296
Advanced degree	0.322	0.317	0.386	0.303	0.338	0.373
Principal experience	11.281	11.753	12.293	11.581	11.934	12.858
New principal at school	0.137	0.150	0.164	0.134	0.145	0.160
New school	0.001	0.001	0.001	0.001	0.004	0.001
Pre-service (teacher/school spell-level) variables						
Education major	0.961	0.771	0.605	0.966	0.610	0.469
Math ed. Major	0.000	0.119	0.355			
English Ed. Major				0.002	0.306	0.319
Math major	0.000	0.023	0.123			
English major				0.004	0.202	0.390
SAT total score	965.703	988.325	1052.284	970.469	1019.337	1036.485
No. of obs. (in-service)	439,766	682,705	501,284	487,465	435,908	325,630
No. of obs. (pre-service)	1,117	606	324	1,300	415	367

Unlike other statewide databases, we can precisely match students and their teachers to specific classrooms at all grade levels.[6] We can determine the specific classroom assignments of middle-school and high-school students, who typically rotate through classrooms during the day for different subjects. This enables us to better separate the effects of teachers from students and their peers. Having data from all K-12 grades also allows us to estimate separate models for elementary, middle and high school which affords us the opportunity to see how the impacts of teacher education and training vary across the three school types.

Not only do our data directly link students and teachers to specific classrooms, they also provide information on the proportion of each student's time spent in each class. This is potentially important for correctly matching teachers and their students at the elementary school level. While primary school students typically receive all of their academic instruction from a single teacher in a single “self-contained” classroom, this is far from universal. In Florida, 5% of elementary school students enrolled in self-contained classrooms are also enrolled in a separate math course, 4% in a separate reading course and 4% in a separate language arts course. In addition, nearly 13% of elementary students enrolled in self-contained elementary classes are also enrolled in some type of exceptional student education course apart from their regular classroom, either special-education or gifted courses.[7]

We restrict our analysis of student achievement to students who receive instruction in the relevant subject area in only one classroom. In our elementary school analysis, only students in “self-contained” classrooms are included. Elementary students spending less than one hour per day in the class are not considered as a member of the classroom peer group. At the middle and high-school levels, students who are enrolled in more than one course in the relevant subject area (mathematics and reading/language arts) are dropped, though all students enrolled in a course are included in the measurement of peer-group characteristics. To avoid atypical classroom settings and jointly taught classes we consider only courses in which 10–40 students are enrolled and there is only one “primary instructor” of record for the class. Finally, we eliminate charter schools from the analysis since they may have differing curricular emphases and student–peer and student–teacher interactions may differ in fundamental ways from traditional public schools.

Despite limiting the sample to students who have only one class per subject, our ability to match the content taught in each classroom with the content on the state test varies by grade and subject. In elementary schools, the matching of subjects is relatively easy because students typically have only one teacher and, as indicated above, we have dropped the small percentage of students who have more than one teacher. In middle and high school, however, more students have multiple classes for the same subject, especially in reading, so that more students are dropped; and some who are not dropped may have their reading scores influenced by classes such as social studies, that involve reading but where developing reading is not the primary purpose.[8] Also, even if we were able to match the test content to a specific classroom, our ability to isolate the effect of

[6] Currently, the Texas data do not provide a way to link teachers and students to specific classrooms. For North Carolina, one can only (imperfectly) match specific teachers and students to classrooms at the elementary school level. Matching is done by identifying the person who administers each student the annual standardized test, which at the elementary school level is typically the classroom teacher.

[7] Since previous studies lack data on students' complete course enrollments, they either ignore the fact that students may receive instruction outside their primary classroom or deal with the issue in an ad-hoc fashion.

[8] Koedel (2009) provides some evidence that social studies teachers influence reading test scores at the high school level.

classroom factors may be further constrained in the case of reading achievement because some students may do a considerable amount of reading in their leisure time.[9] Few students do math during leisure time so this is less of a problem in that subject. For all of these reasons, we have greater confidence in our results for mathematics than for reading.

The ability to link teachers to their university coursework is another important feature of the Florida data. For relatively young teachers (those who attended a Florida public university or community college since 1995) our data include complete college transcript information, including entrance exam scores, courses taken and degrees received. Because Florida has a uniform course numbering system, we are able to create variables that describe each course according to its focus on teacher content knowledge, pedagogical knowledge, and classroom observation/practice in teaching.[10] We then aggregate these measures for each teacher to capture the relevant characteristics of each teacher's entire undergraduate training. We also know the major associated with each college degree and can thus distinguish future teachers who graduated with an education major from those who earned a degree in various non-education disciplines like mathematics and English literature.

4. Econometric model and estimation strategies

4.1. Measuring teacher productivity and within-career education and training

While the issue of measuring a teacher's output is controversial, particularly outside the economics literature, we shall simply define the relevant product as student achievement measured by standardized tests. Consequently, we view a teacher's productivity as their contribution to student achievement, holding other inputs constant. To empirically measure the impact of education and training on teacher productivity it is therefore necessary to first develop a model of student achievement.

Boardman and Murnane (1979) and Todd and Wolpin (2003) model student achievement as a cumulative process in which the current achievement level depends on the individual's initial endowment (e.g. innate ability) and their entire history of individual, family and schooling inputs. They then derive sets of restrictions that yield empirically tractable models of current achievement. Following their approach, we assume that the cumulative achievement function does not vary by grade, is additively separable, and linear.[11] Second, family inputs are assumed constant over time, and the impact of parental inputs on achievement, along with the impact of the initial individual endowment on achievement, induce a (student-specific) constant increment in achievement in each period.[12] Third, the marginal impacts of all prior school inputs decay geometrically with the time between the application of the input and the measurement of achievement at the same rate. Thus lagged achievement serves as a sufficient statistic for all prior schooling inputs. We further assume that lagged achievement follows a Markov process and is uncorrelated with the error.[13] Given these assumptions, current achievement becomes a linear function of prior achievement, contemporaneous student/family and schooling inputs, and an individual specific constant.[14]

We represent contemporaneous achievement inputs by a vector of time-varying student/family attributes, $\mathbf{X}_{it}$ (where the subscripts denote individuals (i) and time (t)), classroom peer characteristics, $\mathbf{P}_{-ijmt}$ (where the subscript $-i$ students other than individual i in classroom j at school m), time-varying teacher characteristics (e.g. experience and in-service training), $\mathbf{T}_{kt}$ (where k indexes teachers), time-invariant teacher characteristics (e.g. innate ability and pre-service education), δ_k, and non-teacher classroom-level inputs (such as books, computers, etc.), $\mathbf{Z}_j$. If we assume that, except for teacher quality, there is no variation in education inputs across classrooms within a school, the effect of $\mathbf{Z}_j$ becomes part of the school-level input vector, $\mathbf{S}_m$. School-level inputs can be decomposed into those that vary over time and those that are invariant over the time period of analysis. The time-varying components we measure are the administrative experience of the principal, the principal's administrative experience squared, whether the principal is new to the school and whether the school is in its first year of operation. Time-invariant school inputs are captured by a school fixed component, ϕ_m. The achievement function can then be expressed as:

$$A_{it} = \lambda A_{it-1} + \beta_1\mathbf{X}_{it} + \beta_2\mathbf{P}_{-ijmt} + \beta_3\mathbf{T}_{kt} + \beta_4\mathbf{S}_{mt} + \gamma_i + \delta_k + \phi_m + \nu_{it} \quad (1)$$

where γ_i represents the student-specific fixed component and ν_{it} is a mean zero error.

We include three measures of teacher education and training in the vector of time-varying teacher characteristics, $\mathbf{T}_{kt}$. Experience, representing on-the-job training, is captured by a set of indicator variables representing various levels of experience; the omitted category is teachers with zero experience. This specification allows for non-linear effects of teacher experience on student achievement. In-service training is measured by a vector of variables representing the number of hours

[9] If some students consistently read more outside of class and the effect of leisure-time reading on student achievement is constant across grades then student fixed effects should account for these differences.

[10] Courses were coded using the course descriptions in the State of Florida State Course Numbering System. The following categories are used: *Education theory/foundations* include courses that cover general education theory or general issues in education. *Pedagogical-instructional* includes general instructional methods and theories to instruction. *Pedagogical-management* includes classroom management issues in general or for different groups of students. *Pedagogical-content* includes combinations of subject and pedagogy. *Other development* includes issues such as ethics, professionalism or administration. *Classroom observation* includes observation in the classroom. *Classroom practice* includes courses that require field experience. *Subject content* includes subject content (e.g. math). Each course was assigned a total value of one which was, in some cases, distributed over several types of training. An example may help to illustrate: SCE 4361 Introduction to Middle School Science Teaching was coded as *pedagogical-content* (0.3) and *classroom observation* (0.7). This is the based on the course description: "Introduction to the roles and responsibilities of science teachers with an emphasis on middle school students. Extensive fieldwork required."

[11] We partially relax the grade-invariance assumption in our empirical work below by estimating separate elementary, middle and high school equations, thereby allowing the coefficients on schooling inputs to vary across grade groups.

[12] An observationally equivalent assumption is that the impact of home inputs is time invariant and the amount of home inputs change by a constant rate over time. If family inputs change over time (at a non-constant rate) and are correlated with variables in the achievement model, this will lead to biased estimates of the model parameters. For example, if parents compensate for an inexperienced teacher by spending more time with their child in learning activities at home, this would impart a downward bias on the estimated effect of teacher experience on student learning. There is little consistent evidence whether or not home inputs systematically vary over time with the quality of schooling inputs, however. Bonesronning (2004) finds that class size has a negative effect on parental effort in Norway, suggesting that school and home inputs are complements. In contrast, Houtenville and Conway (2008) find that parental effort is negatively correlated with school-level per pupil expenditures on instructional personnel, implying that school resources and parental effort are substitutes.

[13] Without this assumption, lagged achievement would be endogenous and OLS estimation would potentially be biased. Alternatively, one could estimate instrumental variable models by assuming there is a higher order degree of serial correlation in the residuals and employing twice and greater lags of achievement as instruments for prior-year achievement (see, for example, Sass (2006) and Koedel and Betts (2009)). Given that our achievement data begins in third grade and our model contains a student fixed effect, such an approach is infeasible at the elementary school level. For middle and high school, we explored the use of instrumental variables (IV) estimation, but the iterative procedure we employ would not converge when used with IV regression.

[14] A thorough discussion of these assumptions and the derivation of the linear education production function model can be found in Todd and Wolpin (2003), Sass (2006) and Harris and Sass (2006).

spent in various types of professional development courses. Both current-year hours of training as well as the amount of training in each of the five prior years are measured separately to allow for delayed implementation of new teaching strategies, human capital depreciation and possible negative impacts of contemporaneous training on student achievement associated with absences from the classroom. Finally, attainment of post-baccalaureate degrees is included to capture the effects of additional formal education obtained after entering the teaching profession. The vector of coefficients on these time-varying teacher characteristics, β_3, thus represents the impact of within-career education and training on teacher productivity. In some cases, we allow β_3 to vary over time, but have omitted the time subscript for simplicity.

In our baseline estimates we assume λ is equal to one and thus the dependent variable is $A_{it}-A_{it-1}$ or the student achievement gain:

$$A_{it}-A_{it-1} = \Delta A_{it} = \beta_1 \mathbf{X}_{it} + \beta_2 \mathbf{P}_{-ijmt} + \beta_3 \mathbf{T}_{kt} + \beta_4 \mathbf{S}_{mt} + \gamma_i + \delta_k + \phi_m + \nu_{it} \quad (2)$$

This implies that the decay rate on prior inputs is zero; school inputs applied at any point in time have an immediate and permanent impact on cumulative achievement.[15] This is of course a strong assumption; we therefore examine the robustness of our results to changes in the assumed value of λ.

4.2. Teacher–student sorting

The three-way-fixed-effects approach we employ will alleviate bias associated with sorting of students and teachers to schools and classrooms based on their time-invariant characteristics. However, as Rothsein (2010) argues, if students are dynamically assigned to teachers on the basis of prior unobserved shocks to student achievement and these shocks are serially correlated, then even models with student, teacher and school fixed effects will yield biased estimates of teacher quality.[16] Using data from a cohort of students in North Carolina, Rothstein conducts falsification tests that support the existence of dynamic sorting. However, Koedel and Betts (2009) find evidence that dynamic sorting of student and teachers to classrooms is transitory and that observing teachers over multiple time periods mitigates the dynamic sorting bias.

For all of our analyses, we impose two sample restrictions in order to mitigate any potential bias from dynamic assignment of students to teachers. First, following Koedel and Betts, we limit our sample to teachers who taught at least two classes of students during the sample period. This restriction has the additional advantage of eliminating teachers whose estimated effects would be very imprecise due to few student observations per teacher.[17] Second, we conduct the strict exogeneity falsification tests proposed by Rothstein by testing for apparent "effects" of both the identity of future teachers and the characteristics of future teachers on current achievement gains.[18] We then limit our analysis sample to those districts in Florida in which the Rothstein test fails to reject the null of strict exogeneity for either teacher effects or teacher characteristics at a 95% confidence level.[19]

4.3. Selection into training

Not only can the matching of students to teachers be non-random, the acquisition of training by teachers may be non-random as well. If the amount of PD a teacher obtains is correlated with other unmeasured teacher characteristics that influence teacher productivity, our estimates of the impact of PD on student achievement could be biased.

If some teachers consistently engage in more professional development than others, due to their tastes, ability, or other time-invariant factors, teacher fixed effects will account for these differences and self-selection in PD acquisition will not lead to bias. Similarly, if there are systematic differences in PD acquisition across schools, due to principal preferences or other school-level factors that are relatively time invariant, these differences will be taken into account by the inclusion of school fixed effects in the achievement model. If the amount of PD is driven by observable factors in the achievement model, like experience, then the estimated impact of PD on student achievement will not suffer from selection bias.[20] However, if teachers select into PD in ways that are correlated with time-varying performance, a problem analogous to the dynamic assignment of students to teachers would exist. For example, if teachers who temporarily exhibit low productivity either choose to obtain more PD or are systematically assigned to acquire more PD by administrators, but would have seen their productivity rise in the absence of any intervention, then the estimated effects of PD will be biased upward.[21]

We present evidence in the Appendix A showing that PD acquisition is unrelated to prior teacher performance, suggesting that Rothstein's dynamic assignment problem does not extend to teacher selection into PD. Instead, we can predict PD hours based on school-level factors, observable teacher characteristics, and re-certification requirements, all of which are either exogenous or accounted for as covariates in the regression models.[22] The standard

[15] Thus, for example, the quality of a child's kindergarten must have the same impact on his cumulative achievement as of the end of the kindergarten year as it does on his achievement at age 18.

[16] For example, if students who have an unusually good test score in one year are systematically assigned to particular teachers the following year and then fall back to their typical achievement gain, estimates of the teachers' effects on student learning will be biased.

[17] Since elementary school teachers typically teach one class per year while teachers in middle and high school teach multiple classes per year, eliminating teachers who only taught a single course during the estimation period reduces the sample by about 15% in elementary school, but only decreases the middle and high school samples by about 1%. Because our sample restriction eliminates elementary school teachers who start teaching in the final year of the sample or who depart after their first year of teaching, our findings should be viewed with caution when evaluating training for first-year elementary school teachers.

[18] To implement Rothstein's falsification test of future teacher effects we employ the econometric specification used by Koedel and Betts (2009), pp. 15–17. The test of future teacher characteristics is discussed in the appendix to Rothsein (2010).

[19] Results of the exogeneity tests for each county in Florida are presented in Tables A1 and A2 of an Appendix available from the authors. We also eliminate some small districts where there were too few teachers to conduct the Rothstein test. Results using the full sample, including districts in which strict exogeneity is rejected, are qualitatively similar to those in the text, which employ the restricted sample. See Appendix Tables A3 and A4.

[20] Correlation between PD and other explanatory variables, like experience, will of course produce multicollinearity, which can inflate the standard errors of the regression coefficients and bias the test statistics toward statistical insignificance. In the Appendix to this paper we show that PD acquisition does tend to fall with experience (Figures A1–A3). When experience is removed from the achievement equation, the PD estimates become more precise, as one would expect (see Appendix Table A26). However, the results are qualitatively similar to those with both experience and PD in the model.

[21] In 2000, the Florida legislature passed legislation establishing new professional development procedures. This included a provision requiring "each school principal to establish and maintain an individual professional development plan for each instructional employee assigned to the school. The individual professional development plan must . . . [be] related to specific performance data for the students to whom the teacher is assigned." (2000 Florida Statutes, 231.600(4)(b)5). However, monitoring of district and school compliance by the Florida Department of Education did not begin until a "Professional Development System Protocol" was initiated in Spring 2003. As of 2006, Bergquist (2006) found district-level compliance with state mandated PD procedures was relatively strong, but implementation of the teacher-level requirement was in the "unacceptable" range, indicating little or no evidence the standard was being implemented. Given our sample ends in the 2004/05 school year, it seems the legislative mandate to link PD to prior performance did little to contribute to the potential endogeneity of PD acquisition in our analysis. For further details see Florida House of Representatives (2008).

[22] See Appendix Figures A1–A3 and Appendix Table A5 and A6. There is also strong evidence from recent reports by Florida schools districts that acquisition of in-service training is driven largely by certification requirements. For example, in 2006/07, the stated purpose for over 75% of in-service hours was certification renewal and another 15% of in-service hours were for the purpose of acquiring additional certifications. Only 8% was for "professional skill building." See Florida Department of Education (2007).

"professional certificate" in Florida is good for a period of five years. To renew her certification, a teacher must complete 120 h of PD or earn six semester hours of college credits. The renewal application must be submitted during the last year the current certificate is valid.[23] Thus certification imposes an exogenous minimum quantity of PD on all teachers. Further, if teachers minimize the present value of the cost of PD by back-loading PD acquisition toward the end of their certification period, then variation in the certification period among teachers would lead to exogenous variation in the timing of PD across teachers.

4.4. Computational issues

Estimation of Eq. (2) is computationally challenging since it includes three levels of fixed effects: individual students (γ_i), teachers (δ_k) and schools (ϕ_m). Standard fixed effects methods eliminate one effect by demeaning the data with respect to the variable of interest (e.g. deviations from student means). Additional effects must then be explicitly modeled through the inclusion of indicator-variable regressors. Given our data includes tens of thousands of teachers and thousands of schools, such standard methods are infeasible.

We combine two different approaches to solve the computational problem associated with estimating a three-level fixed effects model. First, we utilize the "spell fixed effects" method proposed by Andrews et al. (2006) and combine the teacher and school fixed effects into a single effect, $\eta_{km} = \delta_k + \phi_m$. This combined effect represents each unique teacher/school combination or "spell." The achievement equation thus becomes:

$$\Delta A_{it} = \boldsymbol{\beta}_1\mathbf{X}_{it} + \boldsymbol{\beta}_2\mathbf{P}_{-ijmt} + \boldsymbol{\beta}_3\mathbf{T}_{kt} + \boldsymbol{\beta}_4\mathbf{S}_{mt} + \gamma_i + \eta_{km} + \nu_{it} \tag{3}$$

The second approach is an extension of the iterative fixed effects estimator proposed by Arcidiacono et al. (2005).[24] The essence of the Arcidiacono et al. method is to estimate the fixed effect for each individual by calculating each individual's error in each time period (i.e. actual outcome minus the individual's predicted outcome) and then compute the mean of these errors for each individual over time. With each estimate the individual fixed effects are recomputed and the process is iterated until the coefficient estimates converge.

Taking deviations from the teacher–school spell means and subtracting the de-meaned student effect from both sides, the achievement equation becomes:

$$\begin{aligned}\left(\Delta A_{it} - \overline{\Delta A}_{km}\right) - (\gamma_i - \overline{\gamma}_{km}) &= \boldsymbol{\beta}_1(\mathbf{X}_{it} - \overline{\mathbf{X}}_{km}) + \boldsymbol{\beta}_2\left(\mathbf{P}_{-ijmt} - \overline{\mathbf{P}}_{km}\right) \\ &\quad + \boldsymbol{\beta}_3(\mathbf{T}_{kt} - \overline{\mathbf{T}}_{km}) + \boldsymbol{\beta}_4\left(\mathbf{S}_{mt} - \overline{\mathbf{S}}_{km}\right) + \nu_{it}\end{aligned} \tag{4}$$

where the overbar and km subscript denote the mean of the relevant variable over all students and all time periods covered by teacher k at school m. Eq. (4) is estimated by ordinary least squares (OLS), using initial guesses for the individual effects. This produces estimates of $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$ which are then used to calculate predicted outcomes for each individual and in turn update the estimated individual effects. The process is iterated until the coefficient estimates converge. Standard errors are obtained by bootstrapping.[25]

4.5. Measuring the effects of pre-service education on teacher productivity

In order to gauge the effects of teacher ability and college preparation on future productivity we follow a two-step estimation procedure first proposed by Dickens and Ross (1984).[26] In the first step we calculate the estimated teacher-school effects from the estimation of Eq. (4). The predicted teacher–school spell fixed effect can be expressed as the difference between the average achievement gain for all students in group km minus the product of the estimated coefficients and the group averages of the explanatory variables:

$$\hat{\eta}_{km} = \overline{\Delta A}_{km} - \hat{\overline{\gamma}}_{km} - \hat{\boldsymbol{\beta}}_1\overline{\mathbf{X}}_{km} - \hat{\boldsymbol{\beta}}_2\overline{\mathbf{P}}_{km} - \hat{\boldsymbol{\beta}}_3\overline{\mathbf{T}}_{km} - \hat{\boldsymbol{\beta}}_4\overline{\mathbf{S}}_{km} \tag{5}$$

These teacher–school effects can be decomposed into three time-variant components: the part of the teacher effect due to the education they receive as undergraduates, the portion of the teacher effect due to pre-college ability, and the school effect. In the second step we gauge the impact of pre-service education on later teacher productivity by first de-meaning the teacher–school spell effect by school to remove the school effect component and then regressing the resulting within-school teacher effects on a vector of pre-service education variables for teacher k, $\mathbf{U}_k$, their entrance exam scores, $\mathbf{E}_k$, and a random error[27]:

$$\hat{\eta}_{km} - \overline{\hat{\eta}}_m = \hat{\delta}_k = \boldsymbol{\omega}_1\mathbf{U}_k + \boldsymbol{\omega}_2\mathbf{E}_k + \xi_{km} \tag{6}$$

The estimates of the coefficient vector $\boldsymbol{\omega}_1$ indicate the partial correlation between the characteristics of a teacher's pre-service education and their future contribution to student achievement, holding constant their measured pre-college ability. Thus they can be interpreted as signals of future productivity. Following Dickens and Katz (1986), Eq. (8) is estimated by weighted least squares, with the square root of the numbers of students per teacher/school spell as weights.

5. Results

5.1. Effects of experience and professional development training

Initial estimates of the student achievement model, including student, teacher and school fixed effects, are presented in Table 3.[28]

[23] See http://www.fldoe.org/edcert/renew.asp.

[24] Arcidiacono et al derive their estimator in the context of a model with only fixed effects and no other covariates. However, it is straightforward to extend their approach to models with covariates. Details of the derivation are available upon request. Arcidiacono et al. (2005, p.9) demonstrate in simulations that their estimator is unbiased, regardless of the number of observations per student. Their approach does introduce a small bias in a model with so-called "peer fixed effects," but that is not relevant here. A refined procedure which yields unbiased estimates in the presence of peer fixed effects is discussed in Arcidiacono et al. (2007). We demonstrate in the Appendix (see Table A7) that the Arcidiacono et al. (2005) iterative estimator produces estimates that are extremely close to the standard OLS estimates.

[25] The standard errors from the bootstrap procedure do not account for clustering of students within a classroom or classrooms within a school. This is partly compensated for by the fact that we include classroom peer measures and teacher fixed effects (which correspond to a common average error for all students a teacher ever teaches). While recent research has produced bootstrap procedures for clustering at a single level (Cameron et al. (2008b)) and OLS methods for clustering at multiple levels (Cameron et al. (2008a)), neither method is applicable to our situation.

[26] Alternatively, one could directly estimate the effects of teacher pre-service characteristics by entering them as explanatory variables in Eq. (2) and removing the teacher fixed effect, δ_k. However, given our college transcript data begin in 1995, we only have pre-service information for a fraction of teachers. Consequently, in many cases there are only one or two teachers with pre-service information in a school and the student and school effects are not well identified.

[27] We center the teacher-school spell effects on the school mean so that the resulting teacher effects are relative to the mean teacher effect for all teachers at the school, whether or not their pre-service characteristics are known. If we did not de-mean the teacher-school effects ex-ante and instead included school indicators in Eq. (7), the resulting teacher effects would measure teacher productivity relative to other teachers at the same school with pre-service information.

[28] The non-training control variables in the model generally perform as expected. The primary exceptions are the classroom peer variables. While the estimated magnitudes of the peer effects are in line with those from previous studies, such as Hoxby (2000), their direction runs counter to that found in previous work. In the Appendix (see Tables A15-A20) we estimate models that exclude classroom peer variables and find their omission has no significant impact on the estimated effects of training on teacher productivity.

Table 3
Determinants of student math and reading achievement in all grade levels (models with student, teacher and school fixed effects).

	Math			Reading		
	Elementary (Grades 4−5)	Middle (Grades 6−8)	High School (Grades 9−10)	Elementary (Grades 4−5)	Middle (Grades 6−8)	High School (Grades9−10)
1−2 years of experience	1.2845*	2.0165***	−0.0902	2.3827***	1.0650***	−1.0279**
	(1.91)	(5.85)	(0.20)	(3.89)	(2.60)	(2.29)
3–4 years of experience	1.8808***	2.6978***	−0.9703	3.0291***	0.7745	−2.1372***
	(2.63)	(6.23)	(1.61)	(3.56)	(1.30)	(3.36)
5–9 years of experience	0.9314	2.7568***	−1.4500**	2.7685***	1.4181*	−3.2668***
	(1.10)	(5.27)	(2.01)	(3.27)	(1.89)	(3.98)
10–14 years of experience	1.1908	2.7441***	−2.2923***	2.6621**	1.9337**	−4.2250***
	(1.14)	(4.20)	(2.91)	(2.49)	(2.11)	(3.42)
15–24 years of experience	0.9918	3.7083***	−3.7233***	3.4969***	2.2986**	−5.0352***
	(0.91)	(4.44)	(3.68)	(2.77)	(1.96)	(3.14)
25+ years of experience	0.0498	4.2256***	−4.9821***	2.7036*	1.6815	−5.0576***
	(0.04)	(4.21)	(4.02)	(1.66)	(1.08)	(2.77)
In-service Hours$_{t}$	−0.0035*	0.0013	−0.0007	−0.0044*	−0.0027*	−0.0003
	(1.74)	(0.79)	(0.27)	(1.85)	(1.69)	(0.10)
In-service Hours$_{t-1}$	0.0006	0.0051**	0.0002	−0.0005	−0.0065***	0.0034
	(0.24)	(2.87)	(0.07)	(0.22)	(3.48)	(1.47)
In-service Hours$_{t-2}$	−0.0048*	−0.0011	0.0045	−0.0044**	−0.0047*	−0.0012
	(1.93)	(0.64)	(1.62)	(2.18)	(1.88)	(0.38)
In-service Hours$_{t-3}$	−0.0017	0.0016	0.0043*	−0.0029	−0.0029	−0.0040
	(0.64)	(1.10)	(2.21)	(1.37)	(1.63)	(1.33)
In-service Hours$_{t-4}$	−0.0001	0.0044**	0.0037*	−0.0041**	−0.0081***	0.0012
	(0.03)	(2.46)	(1.89)	(1.97)	(4.09)	(0.40)
In-service Hours$_{t-5}$	−0.0018	0.0043**	0.0009	0.0004	0.0007	−0.0003
	(0.54)	(2.44)	(0.38)	(0.14)	(0.29)	(0.08)
Number of schools attended	−0.4953	−1.0714***	−0.8187	−0.9304***	−1.1218**	1.0553
	(1.14)	(3.84)	(1.59)	(2.75)	(2.37)	(1.38)
Structural move	0.1352	−0.8415	2.3111***	2.6824*	−0.3708	−0.1352
	(0.08)	(3.50)	(6.11)	(1.84)	(1.04)	(0.24)
Non-structural move	0.9764***	−0.2561	2.5744***	1.1833***	0.6646**	0.1688
	(3.09)	(1.21)	(7.54)	(5.29)	(2.25)	(0.38)
Class size	−0.1356***	−0.0413***	−0.0195	−0.0867***	−0.0404***	−0.0367*
	(3.96)	(3.75)	(1.29)	(3.41)	(2.79)	(1.83)
Fraction of peers – female	−0.4995	0.5011	−0.3516	−0.2777	−1.2492*	0.0465
	(0.38)	(1.23)	(0.61)	(0.27)	(1.91)	(0.05)
Fraction of peers – Black	4.4849***	6.7867***	3.7716***	4.0835***	5.3932***	1.6323
	(2.96)	(11.96)	(5.20)	(2.86)	(7.06)	(1.60)
Fraction of peers – changed schools	0.3427	−0.2874	−2.8906***	−2.1951**	−1.0275	−2.0931**
	(0.41)	(0.55)	(5.20)	(1.96)	(1.28)	(2.53)
Fraction of peers – structural move	0.3842	−1.9661***	−1.4339**	−0.1986	−0.6117	−1.6996**
	(0.14)	(3.48)	(2.21)	(0.07)	(0.70)	(2.14)
Peer Mean age in Sept. (months)	−0.2132***	0.2111***	−0.2909***	−0.1541**	0.0933***	−0.0240
	(3.05)	(9.13)	(12.31)	(2.30)	(2.57)	(0.52)
New school	−4.0335	−1.4962	−0.0888	3.9341	0.5640	−1.3762
	(0.96)	(0.59)	(0.04)	(1.08)	(0.30)	(0.36)
Principal's administrative experience	−0.0727	0.0289	0.0537	−0.1266	−0.1197	−0.0617
	(0.55)	(0.40)	(0.66)	(1.33)	(1.10)	(0.54)
Principal's administrative experience^2	0.0004	−0.0011	−0.0022	0.0044	0.0028	0.0022
	(0.09)	(0.58)	(0.85)	(1.54)	(0.89)	(0.69)
New principal at school	−0.7713***	−0.0603	0.2748	−0.3616	0.5383**	−0.2232
	(2.66)	(0.38)	(1.42)	(1.51)	(2.31)	(0.83)
Number of students	216,903	295,930	223,793	240,327	192,633	161,961
Number of observations	439,786	683,064	501,640	487,485	436,052	325,796

Models include student, teacher, and school fixed effects as well as grade-by-year and repeater-by-grade dummies. Absolute values of bootstrapped *t*-statistics, based on 50 repetitions, appear in parentheses. * Indicates statistical significance at the .10 level and ** indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test.

Experience enhances the productivity of both elementary and middle school teachers, but not high school teachers. Some prior research suggests that experience effects occur exclusively over the first few years on the job. Here, we find that the bulk of the experience effects are indeed in the early years, but there are still marginal effects even after 10 years of experience. Overall, the results indicate that experience effects in elementary and middle school are quantitatively substantial, ranging from 1.1 to 2.4 scale score points for the first 1–2 years of experience to as much as 2.3–3.7 scale score points for 15–24 years of experience. This translates to 0.04 to 0.09 of a standard deviation in achievement gains or 0.03 to 0.06 of a standard deviation in the achievement level for the first couple of years of experience and as much as 0.16 of a standard deviation in achievement gains for a teacher with 15–24 years of experience (relative to a first-year teacher). In contrast, more experienced high school teachers are generally less productive than when they were rookie high school instructors.

We also find contemporaneous professional development (PD) is associated with either no change or a reduction in teacher productivity. This is unsurprising for several reasons noted in past research (Goldhaber and Anthony (2007); Harris and Sass (2009)). PD takes time away from classroom instruction and preparation time. In addition, if substitute teachers are hired so that the PD can take place during school hours, and if the substitutes are less effective or unable to maintain the continuity of instruction in the permanent teacher's absence, then this too may reduce measured teacher value-added.

Finally, because some portion of the school year has typically passed when PD occurs, and lesson plans for later in the year may already be established, teachers may have little opportunity to immediately incorporate what they learn.

If PD boosts the human capital of teachers, we would expect to see positive effects on student achievement after the professional development training is acquired and put to use in the classroom, perhaps with some diminution over time if the knowledge depreciates or becomes less relevant. However, we only observe positive effects of prior PD for middle and high-school math teachers. For high school math, PD acquisition in the current year has a negative effect, but the effect on student achievement becomes positive and is marginally significant after two years. Third and fourth-lagged PD is positively correlated with current student achievement gains, but fifth-lagged PD has no significant effect on high school math achievement. For middle school math, the positive and statistically significant effects of prior PD occur with once, fourth and fifth-lagged PD. For each of the other subject-grade level combinations, there are mainly insignificant and sometimes even negative correlations of past PD with current student achievement gains.

Even where PD does seem to show consistent positive effects, they appear modest in size relative to experience. Given the average number of hours of professional development coursework per year is approximately 50, the difference between having no professional development versus the average professional development is at most about ($0.005\times 50=0.25$) scale score points or about one-fifth the difference between a rookie middle school math teacher and one with a year of experience. Also, we estimate the cost of 50 h of PD is $3,175, yielding an effect-to-cost ratio of 0.0019 (with costs in thousands of dollars).[29] In comparison, Harris (2009) finds the comparable ratio for class size reduction is roughly 0.026–0.086 and even this ratio is considered small relative to alternative education programs such as peer tutoring. Nor would a ratio of this size pass a cost–benefit test (Harris (2009), Krueger (2003)). Thus, the effects of PD seem small by any measure.

Perhaps the most striking pattern in Table 3 is the considerable variation in estimated effects of experience and PD across grade levels and subjects. While the divergent estimates may reflect real differences in training impacts, there are also a number of measurement/estimation issues that affect each grade/subject combination in distinctive ways: (a) math may be less prone than reading to the influence of non-school activities that students participate in (e.g., reading for leisure, although if constant over time this should be captured in the student fixed effects); (b) course content is more closely linked to test content in middle school than in high school, where math courses become more diverse and specialized and English/language arts courses include subjects like literature[30]; (c) middle (and high) school teachers generally teach multiple classes per year and thus have many more student observations per teacher than do elementary teachers, reducing the sampling error in teacher productivity estimates; (d) the fact that elementary teachers typically teach one class per year makes it more difficult to identify teacher effects from classroom/peer effects and (e) achievement gains can be measured for three grades in middle school (6, 7 and 8), but only two grades each in elementary and high school, thereby reducing sampling error in the estimated student fixed effects.

The combination of these five factors points to middle school math as the grade/subject combination in which teacher productivity will be most precisely estimated. Middle school math is also the only grade/subject combination where the lagged effects of PD follow anything like the predicted pattern. If we were to focus only on middle school math, the results would be more positive with regard to the efficacy of teacher experience and professional development and also (as we show below), more positive for the effects of advanced degrees and undergraduate coursework on teacher productivity. However, as noted above, even for middle school math the magnitudes of the PD effects are relatively small.

Given the potential role of outside factors on reading, it might seem surprising that teacher experience in elementary school appears to matter more in reading than in math. However, there is evidence that elementary teachers are more likely to work to develop their skills by seeking out help from colleagues and other sources for reading, whereas there is significantly less communication among teachers with regard to mathematics (Spillane (2005)).[31] If this is true then elementary teachers might be more likely to develop and improve their reading practices over time, thus producing greater productivity gains from experience in reading relative to math.

The same is not true in middle and high school where there is greater specialization of content and where the subject-matter department (e.g., math department) is the primary organizational unit for teachers in all subjects (Grossman and Stodolsky (1994)). We might therefore expect the role of teacher experience to be similar over time across subjects in middle and high schools, which is what we find.

In Tables 4–7, we present results from estimating a variety of alternative specifications for elementary math, elementary reading, middle school math, and middle school reading, respectively. The sensitivity analysis for high school can be found in Appendix A.[32] Column (1) in each table replicates the results presented in Table 3. Column (2) presents estimates from a model that excludes teacher effects. In column (3) we replace student fixed effects with a set of time-invariant and quasi-time-invariant indicators: race/ethnicity, gender, free-lunch status, gifted status, limited-English proficiency status, and mental, physical, emotional and other disabilities. In column (4), we no longer assume complete persistence in prior achievement and instead assume a decay rate of 0.4 (persistence of 0.6).[33] Column (5) reduces the number of lags in the effects of teacher PD to three. Finally, column (6) reports estimates from a model that distinguishes between content-oriented and other PD and adds the receipt of an advanced degree as a (time-varying) covariate.[34]

The results are generally insensitive to specification, with a few notable exceptions. The estimated effects of teacher experience in elementary are generally larger and more precise when we omit teacher fixed effects. This suggests that there is some "positive selection" with respect to teachers leaving the sample–that is, less

[29] Effects on test scores in education research are often reported by dividing by the standard deviation of the test score level. The 0.25 scale score points translates to 0.006 standard deviations. The average hourly wage (plus fringe benefits) of teachers in 2007 was $37 (Harris, 2009). Thus, the opportunity cost of average annual PD is 50*$37 = $1850. There are other costs such as PD development, the time and preparation of PD instructors, and facilities. According to the Florida House of Representatives (2008), these direct costs were approximately $1150 per teacher in 2002/03 or about $1325 in 2007 dollars, yielding $3175 in total costs per teacher. The costs from Harris (2009) for class size are expressed in thousands. So, the comparable ratio for PD is $0.006/3.175 = 0.0019$.

[30] Consistent with this hypothesis, Clotfelter et al. (2010), which employs end-of-course exams to assess teacher productivity, finds stronger effects of teacher credentials than do other studies of secondary school teachers that utilize general subject-area achievement exams.

[31] In one study, 82% of teachers reported that the "school has primary expertise" over reading and literacy while only 13% of the same teachers said the same thing about math; elementary teachers are instead more likely to rely on external groups such as professional organizations and government agencies for guidance about math instruction (Spillane (2005)).

[32] Given the potential problems with attributing teachers responsible for subject-relevant instruction and mismatch between course and exam content, as well as then need to conserve space, we omit analysis of alternative specifications for high school achievement. Estimates of alternative specifications for the high school sample are included in the Appendix Tables A8, A9, A12, A13 and A24.

[33] Estimates from elementary and middle-school models with persistence of 0.2, 0.4, 0.6, 0.8 and 1.0 are provided in Appendix Tables A10 and A11.

[34] To save space, only the estimated coefficients on the content-oriented PD variables are presented. Full results, with estimates for both the content-oriented and non-content oriented coefficients are presented in Appendix Table A9.

Table 4
Effects of teacher experience, in-service training and advanced degrees on elementary math achievement.

	(1)	(2)	(3)	(4)	(5)	(6)
1–2 years of experience	1.2845*	1.5166***	0.7016**	0.9905**	1.2859*	1.2766*
	(1.91)	(4.35)	(2.09)	(2.20)	(1.91)	(1.83)
3–4 years of experience	1.8808***	1.5664***	0.9489**	1.6261***	1.8771***	1.8661**
	(2.63)	(4.21)	(2.28)	(3.19)	(2.63)	(2.34)
5–9 years of experience	0.9314	1.6925***	0.3361	1.1913*	0.9245	0.9255
	(1.10)	(4.69)	(0.72)	(1.94	(1.09)	(0.93)
10–14 years of experience	1.1908	2.2137***	0.4974	1.5249**	1.1843	1.1843
	(1.14)	(5.85)	(0.84)	(2.03)	(1.13)	(0.87)
15–24 years of experience	0.9918	2.1364***	0.6766	1.5570*	0.9881	0.9828
	(0.91)	(5.56)	(0.96)	(1.87)	(0.91)	(0.68)
25+ years of experience	0.0498	1.3558***	−0.4906	1.1304	0.0508	0.0393
	(0.04)	(3.43)	(0.57)	(1.13)	(0.04)	(0.03)
In-service Hours$_t$	−0.0035*	−0.0023	0.0001	−0.0017	−0.0035*	−0.0031
	(1.74)	(1.54)	(0.10)	(1.04)	(1.72)	(0.74)
In-service Hours$_{t-1}$	0.0006	0.0011	0.0022*	0.0013	0.0006	0.0023
	(0.24)	(0.79)	(1.74)	(0.60)	(0.25)	(0.62)
In-service Hours$_{t-2}$	−0.0048*	−0.0021	−0.0020*	−0.0025	−0.0048**	−0.0103***
	(1.93)	(1.62)	(1.65)	(1.25)	(2.06)	(2.63)
In-service Hours$_{t-3}$	−0.0017	0.0034***	0.0002	0.0003	−0.0016	−0.0008
	(0.64)	(2.64)	(0.19)	(0.11)	(0.69)	(0.19)
In-service Hours$_{t-4}$	−0.0001	0.0026	0.0009	0.0009		−0.0012
	(0.03)	(1.50)	(0.67)	(0.41)		(0.27)
In-service Hours$_{t-5}$	−0.0018	−0.0010	0.0002	−0.0015		−0.0043
	(0.54)	(0.60)	(0.15)	(0.56)		(0.97)
Advanced degree						−0.3783
						(0.74)
Student fixed effects	Yes	Yes	No	Yes	Yes	Yes
Teacher fixed effects	Yes	No	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Input rate of decay	0.0	0.0	0.0	0.4	0.0	0.0
Type of PD	All	All	All	All	All	Content
Number of Students	216,903	216,903	216,901	216,903	216,903	216,893
Number of observations	439,786	439,786	439,782	439,786	439,786	439,766

See notes to Table 3. Specification in Column (6) also includes measures of non-content oriented PD.

Table 5
Effects of teacher experience, in-service training and advanced degrees on elementary reading achievement.

	(1)	(2)	(3)	(4)	(5)	(6)
1–2 years of experience	2.3827***	2.4054***	1.1497***	1.8937***	2.3883***	2.4032***
	(3.89)	(7.08)	(3.62)	(4.49)	(3.90)	(4.52)
3–4 years of experience	3.0291***	2.5362***	1.3080***	2.5083***	2.9893***	3.0681***
	(3.56)	(6.89)	(3.28)	(4.35)	(3.51)	(4.07)
5–9 years of experience	2.7685***	2.8183***	1.1127**	2.4719***	2.7234***	2.8562***
	(3.27)	(7.99)	(2.46)	(4.21)	(3.22)	(3.23)
10–14 years of experience	2.6621**	2.8837***	1.0320*	2.6644***	2.6033**	2.7758**
	(2.49)	(7.83)	(1.80)	(3.72)	(2.43)	(2.24)
15–24 years of experience	3.4969***	3.9212***	1.6391**	3.4014***	3.4490***	3.6045***
	(2.77)	(11.59)	(2.41)	(4.06)	(2.73)	(2.67)
25+ years of experience	2.7036*	3.5050***	0.4823	3.0436***	2.6560	2.7657*
	(1.66)	(10.37)	(0.57)	(2.87)	(1.63)	(1.73)
In-service Hours$_t$	−0.0044*	−0.0005	0.0000	−0.0031	−0.0039	−0.0087**
	(1.85)	(0.34)	(0.01)	(1.60)	(1.62)	(2.54)
In-service Hours$_{t-1}$	−0.0005	0.0030*	0.0005	−0.0005	0.0003	−0.0028
	(0.22)	(2.02)	(0.38)	(0.23)	(0.13)	(0.60)
In-service Hours$_{t-2}$	−0.0044*	−0.0014	−0.0017	−0.0024	−0.0034*	−0.0104**
	(2.18)	(1.04)	(1.49)	(1.55)	(1.65)	(2.40)
In-service Hours$_{t-3}$	−0.0029	0.0014	−0.0007	−0.0015	−0.0018	0.0020
	(1.37)	(1.00)	(0.64)	(0.84)	(0.85)	(0.53)
In-service Hours$_{t-4}$	−0.0041**	0.0032**	−0.0015	−0.0028		−0.0039
	(1.97)	(2.38)	(1.17)	(1.61)		(0.98)
In-service Hours$_{t-5}$	0.0004	−0.0004	0.0010	0.0006		0.0002
	(0.14)	(0.18)	(0.66)	(0.23)		(0.04)
Advanced degree						−0.3016
						(0.54)
Student fixed effects	Yes	Yes	No	Yes	Yes	Yes
Teacher fixed effects	Yes	No	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Input rate of decay	0.0	0.0	0.0	0.4	0.0	0.0
Type of PD	All	All	All	All	All	Content
Number of students	240,327	240,327	240,321	240,327	240,327	240,317
Number of observations	487,485	487,485	487,473	487,485	487,485	487,465

See notes to Table 3.

Table 6
Effects of teacher experience, in-service training and advanced degrees on middle school math achievement.

	(1)	(2)	(3)	(4)	(5)	(6)
1–2 years of experience	2.0165***	1.1587***	1.4639***	1.4549***	2.0148***	2.0403***
	(5.85)	(6.44)	(7.48)	(5.75)	(5.85)	(7.80)
3–4 years of experience	2.6978***	2.1615***	2.0489***	2.0039***	2.7114***	2.7132***
	(6.23)	(10.86)	(8.07)	(6.51)	(6.24)	(6.58)
5–9 years of experience	2.7568***	2.2538***	2.4295***	1.9141***	2.7903***	2.7718***
	(5.27)	(11.77)	(8.39)	(4.98)	(5.33)	(6.87)
10–14 years of experience	2.7441***	2.4703***	2.4772***	1.8534***	2.7489***	2.7773***
	(4.20)	(12.32)	(6.53)	(3.73)	(4.21)	(4.32)
15–24 years of experience	3.7083***	2.4081***	3.3999***	2.6186***	3.7370***	3.7720***
	(4.44)	(13.57)	(7.10)	(4.17)	(4.47)	(5.09)
25+ years of experience	4.2256***	2.5446***	3.6979***	2.7677***	4.234***	4.3061***
	(4.21)	(11.23)	(6.16)	(3.80)	(4.20)	(4.65)
In-service Hours_t	0.0013	0.0028**	0.0000	0.0007	0.0007	0.0018
	(0.79)	(2.54)	(0.04)	(0.48)	(0.42)	(0.69)
In-service Hours_{t-1}	0.0051**	0.0037***	0.0032***	0.0037**	0.0043**	0.0071**
	(2.87)	(3.26)	(3.11)	(2.55)	(2.45)	(2.26)
In-service Hours_{t-2}	−0.0011	−0.0028***	−0.0002	−0.0010	−0.0023*	0.0067*
	(0.64)	(2.96)	(0.19)	(0.72)	(1.37)	(1.83)
In-service Hours_{t-3}	0.0016	0.0015	0.0021**	0.0018	0.0006	0.0031
	(1.10)	(1.54)	(2.10)	(1.42)	(0.40)	(0.89)
In-service Hours_{t-4}	0.0044**	0.0000	0.0023**	0.0032**		−0.0000
	(2.46)	(0.04)	(2.32)	(2.21)		(0.01)
In-service Hours_{t-5}	0.0043**	0.0052***	0.0042***	0.0036**		0.0044*
	(2.44)	(4.84)	(3.61)	(2.44)		(1.66)
Advanced degree						0.7778**
						(2.18)
Student fixed effects	Yes	Yes	No	Yes	Yes	Yes
Teacher fixed effects	Yes	No	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Input rate of decay	0.0	0.0	0.0	0.4	0.0	0.0
Type of PD	All	All	All	All	All	Content
Number of students	295,930	295,930	295,918	295,930	295,930	295,824
Number of observations	683,064	683,064	683,040	683,064	683,064	682,705

See notes to Table 3.

effective teachers are more likely to leave. This seems especially true for teachers with more than four years of experience where the results switch from being statistically insignificant to significant. In contrast, removing teacher effects from the middle school models does not consistently raise the significance of the estimated experience effects and produces smaller point estimates, on average, relative to the model which includes teacher fixed effects. Changing the persistence assumption seems to have roughly the same influence on the results as dropping the teacher fixed effect, but it is unclear why this is the case. The results for experience are insensitive to other model specification choices.

The results for PD are somewhat more sensitive to model specification, though the resulting patterns are still mixed and frequently insignificant. Once again, the results are generally most sensitive to the inclusion/exclusion of teacher fixed effects with the impact of PD generally more positive with the omission of teacher fixed effects.

In the final column of Tables 4–7 we control for formal education in the form of attainment of advanced degrees and differentiate content-oriented PD from PD designed to address other components of a teacher's human capital, such as classroom management techniques and pedagogy. Since our model includes teacher fixed effects, post-baccalaureate degrees earned prior to the period of analysis wash out when we demean the data. Thus our approach measures the impact of *changes* in the possession of an advanced degree (for a given teacher) during the period of study.[35] Our results indicate that obtaining an advanced degree during one's teaching career is positively related to teacher productivity only in the case of middle school math. For middle school reading the effect is actually negative and for all others the effects are insignificant. The insignificant results might be because graduate degrees include a combination of pedagogy and content and our other evidence suggests that only the latter has a positive influence on teacher productivity.

Other explanations for the graduate degree results arise from issues of methodology. Our use of teacher fixed effects approach imposes the implicit assumption that the receipt of the graduate degree reflects a sudden infusion of new preparation. In reality, the receipt of the degree is the culmination of several years of graduate courses whose influence may already be reflected in the teacher effects, especially for those teachers who take graduate courses over many years before receiving a graduate degree. Another possibility is that teachers load up on courses in the academic year preceding the receipt of the degree and therefore have less time to devote to their students. We found evidence above of such a contemporaneous decline in productivity when we considered the effects of in-service professional development coursework. Again, the most striking variation in the results is across grades and subjects rather than across specifications.

5.2. Endogeneity and additional sensitivity analysis

Appendix A provides additional information and sensitivity analyses regarding dynamic selection of students to teachers and teacher to training, alternative controls for student heterogeneity, varying degrees of persistence in past schooling inputs, ceiling effects in the test instrument, variation in the number of lags of prior PD included in the model, pooling across grade levels (rather than estimating elementary, middle, and high school in separate

[35] The estimated coefficient on the advanced-degree variable measures the average difference in productivity between the time before and the time after receipt of the degree. Before the degree is received some knowledge may have already been acquired through coursework already completed, thus biasing the estimated effect toward zero. However, work toward an advanced degree may take away from time available for class preparation and other teaching-related activities, which would tend to lower productivity before receipt of the degree and upwardly bias the estimated impact of the degree.

Table 7
Effects of teacher experience in-service training and advanced degrees on middle school reading achievement.

	(1)	(2)	(3)	(4)	(5)	(6)
1–2 years of experience	1.0650***	0.1298	0.5019*	0.9234***	1.0725***	0.9963**
	(2.60)	(0.48)	(1.89)	(3.01)	(2.63)	(2.47)
3–4 years of experience	0.7745	−0.0480	0.0762	0.8771*	0.6946	0.7310
	(1.30)	(0.19)	(0.20)	(1.91)	(1.16)	(1.18)
5–9 years of experience	1.4181*	0.4289*	0.1787*	1.0883*	1.2202	1.3324*
	(1.89)	(1.79)	(0.38)	(1.93)	(1.62)	(1.88)
10–14 years of experience	1.9337**	0.4815*	0.5562*	1.6655**	1.6979*	1.8517*
	(2.11)	(1.91)	(0.89)	(2.41)	(1.84)	(1.70)
15–24 years of experience	2.2986**	0.6882***	0.7485	1.8506**	2.0976*	2.194*
	(1.96)	(2.63)	(0.98)	(2.15)	(1.77)	(1.66)
25+ years of experience	1.6815	0.7962***	0.6696	1.3766	1.3815	1.4971
	(1.08)	(2.76)	(0.71)	(1.19)	(0.88)	(0.83)
In-service Hours_t	−0.0027*	0.0017	−0.0004	−0.0019	−0.0017	−0.0010
	(1.69)	(1.36)	(0.33)	(1.32)	(1.04)	(0.37)
In-service Hours_{t-1}	−0.0065***	−0.0018	−0.0033***	−0.0051***	−0.0049***	−0.0038
	(3.48)	(1.30)	(2.69)	(3.31)	(2.64)	(1.11)
In-service Hours_{t-2}	−0.0047*	0.0004	−0.0027**	−0.0041**	−0.0024	−0.0025
	(1.88)	(0.29)	(2.22)	(2.01)	(1.02)	(0.55)
In-service Hours_{t-3}	−0.0029	−0.0000	−0.0021*	−0.0023	−0.0008	−0.0075**
	(1.63)	(0.04)	(1.76)	(1.60)	(0.46)	(2.14)
In-service Hours_{t-4}	−0.0081***	−0.0022*	−0.0041***	−0.0058***		−0.0092**
	(4.09)	(1.93)	(3.20)	(3.44)		(2.47)
In-service Hours_{t-5}	0.0007	0.0037**	−0.0010	0.0006		−0.0076**
	(0.29)	(2.49)	(0.65)	(0.31)		(2.07)
Advanced degree						−1.2715*
						(1.93)
Student fixed effects	Yes	Yes	No	Yes	Yes	Yes
Teacher fixed effects	Yes	No	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Input rate of decay	0.0	0.0	0.0	0.4	0.0	0.0
Type of PD	All	All	All	All	All	Content
Number of students	192,633	192,633	192,618	192,633	192,633	192,595
Number of observations	436,052	436,052	436,019	436,052	436,052	435,908

See notes to Table 3.

regressions), collinearity between PD and advanced degrees, spillovers across teachers, the role of student peers, and other issues. In each case, the basic findings presented in Table 3 are found to be robust.[36] We highlight a few particular issues below.

We consider the effects of altering the estimation sample on our results. The results presented in Tables 3–7 are based on separate samples of elementary, middle and high school students. In addition to emphasizing the differential effects of teacher training across grade groups, this approach maximizes the number of districts we can include in our sample because for many districts, evidence of dynamic student–teacher matching is found at some grade groups within a district and not others. The advantage of pooling across grade groups is an increase in the number of observations per student and thus a potential improvement in the identification of student and teacher effects. In Appendix A we present estimates of the achievement model, pooled across all grades with grade-group interactions to allow for differential slopes. Similar to the results from grade-group subsamples, we find that contemporaneous PD has no significant effects on either math or reading achievement and prior PD has virtually no positive effects on reading achievement. Lagged PD for math teachers continues to have some positive effects on current-year achievement gains.

While the above results focus on the direct effects of own-PD on individual teachers, we also considered the possibility of spillovers across teachers. If teachers share ideas and teaching techniques they learn in professional development courses, then there could be positive spillovers to the productivity of the colleagues of the teacher who obtained the training. To test for such positive externalities, we re-estimated the achievement model, including measures of the number of hours of PD obtained by other teachers at the same school. The results show almost no indication of positive externalities (only one coefficient is positive and significant), while several are negative and significant. The general absence of positive spillovers is unsurprising given the weak effects of own-PD on productivity.

The results in Tables 3–7 also assume that the effects of PD and experience are independent of one another. If on-the-job training acquired through experience is a substitute for PD and formal education, then the marginal effects of PD and advanced degrees should decline as a teacher acquires experience. The results provide at best modest support for this hypothesis, however. We find no clear pattern to the magnitude of the PD coefficients with respect to experience.

5.3. Pre-service training effects

In this section we use the estimated teacher effects as the dependent variable in order to analyze the effects of pre-service training on the time-invariant component of a teacher's productivity. Table 8 displays the results of estimating Eq. (7), measuring a teacher's pre-service formal education, $\mathbf{U}_k$, by the number of credits earned in various types of education and subject-relevant non-education courses. Pre-college ability is taken into account by including the SAT-equivalent entrance exam scores of future teachers.[37] This allows us to distinguish between the human capital effects of college coursework and the self-selection of individuals into particular types of courses.[38] Unfortunately, our sample size drops

[36] Consistent with our findings, Koedel and Betts (2010) show that ceiling effects are generally not a significant source of bias in achievement models.

[37] To maximize the amount of college entrance exam information available we include data from the state university system, the community college system and a database on applicants for a state merit-based scholarship known as “Bright Futures.” ACT scores as well as community college placement exam scores were converted to SAT-equivalent scores using concordance tables.

[38] College entrance exam scores are a noisy measure of teacher ability. If unmeasured teacher ability is correlated with course-taking patterns (e.g. individuals with strong oral communication skills take fewer pedagogy-related education courses, our estimated effects of coursework on subsequent teacher performance could be biased.

Table 8
Estimates of the effects of college course work and college entrance exam scores on a teacher's "value-added" to student math and reading achievement.

	Math			Reading		
	Elementary (Grades 3–5)	Middle (Grades 6–8)	High School (Grades 9–10)	Elementary (Grades 3–5)	Middle (Grades 6–8)	High School (Grades9–10)
Gen. educ. theory credits	−0.0869	−0.8196	1.1453	−0.4010	−0.4837	−1.0065
	(0.08)	(0.77)	(0.56)	(0.40)	(0.33)	(0.66)
Pedagogical – instructional credits	−1.1218	0.6119	−1.0557	−2.5472***	1.0158	−0.8428
	(1.13)	(0.61)	(0.64)	(2.79)	(0.74)	(0.57)
Pedagogical – management credits	2.4152	2.8635		−7.2584	14.3415	71.6291***
	(0.21)	(0.37)		(0.64)	(1.30)	(3.29)
Pedagogical – content credits	1.0890	0.9704	0.4393	0.1044	−1.9764	2.1757
	(1.29)	(1.39)	(0.28)	(0.12)	(1.54)	(1.23)
Professional development credits	0.7413	−0.0286	−1.1643	1.0954	−0.7975	0.7980
	(0.84)	(0.03)	(0.69)	(1.36)	(0.71)	(0.65)
Classroom observation credits						−3.8180
						(0.22)
Classroom practice credits	−0.4917	1.0370	−1.4072	0.6530	−1.6411	1.8573
	(0.49)	(1.00)	(0.95)	(0.69)	(1.33)	(1.27)
Subject content credits	−1.7832	0.0184	−2.9352	−3.5663*	0.1281	−0.6339
	(0.64)	(0.01)	(0.94)	(1.79)	(0.04)	(0.09)
Mathematics credits	−0.3690**	−0.0277	−0.6689			
	(2.32)	(0.11)	(1.60)			
Statistics credits	−0.7864	5.0413**	−1.9945			
	(0.34)	(2.52)	(0.80)			
English literature credits				−0.0482	0.2681	0.4477
				(0.16)	(0.59)	(0.88)
Math education credits	1.9152	−0.2972	0.6520			
	(1.51)	(0.38)	(0.58)			
Language arts educ. credits				0.2666	0.9251	−1.7901*
				(0.18)	(0.97)	(1.70)
SAT total score	−0.0012	−0.0020	−0.0062	−0.0012	0.0020	0.0002
	(0.30)	(0.40)	(1.09)	(0.34)	(0.39)	(0.03)
R-squared	0.0140	0.0176	0.0281	0.0100	0.0215	0.0455
Number of observations	1,125	616	325	1,307	419	368

The dependent variable is the (within-school) teacher fixed effect estimated from a model of student achievement using all Florida public school students in the relevant grades. Missing cells in the top half of the table reflect collinearity among variables. Observations are weighted by the square root of the number of students per teacher. Absolute values of *t*-statistics appear in parentheses. * Indicates statistical significance at the .10 level and ** indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test.

significantly because our data only contain information on a teacher's college coursework if she attended a public university in Florida in 1995 or later. In addition, even for the subsample of future teachers with transcript information, there are a substantial number of missing observations on the college entrance exam. In Appendix A we present results for a larger sample that does not control for college entrance exam scores.

With the lone exception of courses on classroom management taken by future high school reading teachers, we find no evidence that the quantity of any particular type of college coursework is associated with greater productivity.[39] This finding is corroborated by an analysis of college majors, presented in Appendix A. There we find that education majors are no more or less productive than teachers whose initial bachelor's degree was in another discipline. We also find that teacher college entrance exam scores are not associated with teacher productivity.[40]

6. Summary and conclusions

Our study differs in several important ways from other recent contributions to the rapidly expanding literature on teacher training. First, ours is the first multi-district study to simultaneously control for unobserved student, teacher and school heterogeneity through the use of multiple levels of fixed effects. Coupled with tests for dynamic student–teacher sorting, we argue this approach significantly attenuates selection biases due to non-random assignment of students to teachers and teachers to training. Second, while most recent research has focused on teacher experience and attainment of advanced degrees, ours is the first analysis to concurrently estimate the impacts of experience, post-baccalaureate degrees, in-service professional development and pre-service undergraduate education on the productivity of teachers. Further, we are able to measure the various forms of training at a finer level than in most prior studies, including distinguishing specific types of coursework at the undergraduate level and differentiating between different kinds of professional development training received while teaching. Finally, ours is the first study to distinguish between the quality of undergraduate training and the innate ability of future teachers by including individual-specific college entrance exam scores.

Our main conclusion is that elementary and middle school teacher productivity increases with experience (learning by doing), but formal training acquired while teaching generally does not enhance the ability of teachers to boost student achievement. Like much of the prior literature, we find that early-career experience significantly enhances teacher productivity. However, in contrast to previous work, we find that the effects of experience extend beyond the first few years in some subjects and grades. Our results also reinforce evidence from prior literature which finds that attainment of advanced degrees does not improve teacher productivity. Middle school math is the only subject/grade-level combination where we find obtaining an advanced degree promotes the ability of a teacher to boost student achievement. We also find that in-service professional development has little or no effect on the ability of teachers to improve student

[39] It is important to keep in mind that we are only analyzing teachers who teach math, reading or both subjects. It is quite possible that in other areas, such as science in secondary schools, that content knowledge acquired through undergraduate coursework could be an important determinant of teacher productivity.

[40] The only previous study to include entrance exam scores, Ferguson and Ladd (1996), utilizes school-level average composite scores on the American College Test (ACT), rather than the scores of individual teachers.

achievement, with the possible exception of middle school math. The observed ineffectiveness of professional development coursework is consistent with three recent studies on the subject (Jacob and Lefgren (2004); Garet et al. (2008); Garet et al., 2010)). However, our results apply to the broader context of PD generally acquired by teachers rather than the relatively narrow contexts (PD associated with accountability sanctions and with early reading and middle-school math in high-poverty schools) that have been the subject of prior research.

We find almost no evidence that specific undergraduate coursework in education affects an individual's later productivity as a teacher. Similarly, only in the case of high school reading do teachers whose first undergraduate degree was in education significantly outperform non-education majors. When pre-college ability is taken into account by college entrance exam scores, even this differential becomes insignificant. Our results are consistent with findings from the only other recent studies to include data on college majors, Aaronson et al. (2007) and Betts et al. (2003), as well as studies comparing teachers who completed traditional teacher preparation programs with “alternatively certified” teachers (Boyd et al. (2006), Kane et al. (2006), Xu et al. (2008), Constantine et al. (2009)).

The apparent ineffectivness of formal pre-service and in-service training may be because teacher productivity is context-specific (e.g., the specific curriculum and/or types of students) and formal training may be too standardized to account for these differences; alternatively, teacher education programs might not be focused on the types of skills that generate student achievement. This is also consistent with the fact that experience, which is less standardized, does seem to improve productivity.

Although our findings provide little support for the general efficacy of formal training for teachers, there is considerable variation in our results by grade level and subject. While the variation across grade levels and subjects may reflect real differences in training impacts, there are several reasons to think that these patterns may reflect methodological/measurement issues, especially the varying degree to which the actual content of courses match the content of the exams, as well as varying external influences on certain exam scores. We are arguably best able to measure a teacher's impact on student learning in middle school math and this where the effects of experience and formal training seem most positive.

Although much work remains to fully understand the ways in which training affects the ability of teachers to promote student learning, our analysis, in combination with other recent research, does offer some tentative suggestions for shaping future policy. First, our finding (and that of others) that experience greatly enhances the productivity of elementary and middle school teachers early in their careers indicates that policies designed to promote retention of young teachers can yield significant benefits over and above avoiding the cost of hiring new teachers. Second, our finding (consistent with prior research), that advanced degrees are uncorrelated with the productivity of elementary school teachers suggests that current salary schedules, which are based in part on educational attainment, may not be an efficient way to compensate teachers in primary school. Third, our evidence that only content-oriented professional development coursework taken by middle and high-school math teachers appears effective suggests that relatively more resources ought to be put into content-focused training for teachers in the upper grades and that changes are warranted in PD at the elementary level and in pedagogical in-service training generally. Finally, given we find scant evidence that the amount of undergraduate coursework in education effects future productivity and our work and that of others does not find education majors are significantly more productive as teachers than non-education majors, it seems worthwhile to rethink the structure of traditional preparation programs and continue experimentation with so-called “alternative certification” programs that facilitate the entry of non-education majors into teaching.

Acknowledgements

We wish to thank the staff of the Florida Department of Education's K-20 Education Data Warehouse and the Bureau of Educator Recruitment, Development and Retention for their assistance in obtaining and interpreting the data used in this study. The views expressed is this paper are solely our own and do not necessarily reflect the opinions of the Florida Department of Education. Valuable research assistance was provided by Li Feng, John Gibson, Micah Sanders and Cynthia Thompson. This research is part of a larger project assessing teacher quality funded by grant R305M040121 from the U.S. Department of Education.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.jpubeco.2010.11.009.

References

Aaronson, Daniel, Barrow, Lisa, Sander, William, 2007. Teachers and student achievement in the Chicago public high schools. Journal of Labor Economics 25, 95–135.
Andrews, Martyn, Schank, Thorsten, Upward, Richard, 2006. Practical fixed effects estimation methods for the three-way error components model. Stata Journal 6, 461–481.
Angrist, Joshua D., Lavy, Victor, 2001. Does teacher training affect pupil learning? Evidence from matched pairings in Jerusalem public schools. Journal of Labor Economics 19 (2), 343–369.
Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster and Josh Kinsler. 2005. “Estimating Spillovers in the Classroom with Panel Data”. Unpublished manuscript.
Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster and Josh Kinsler. 2007. “Estimating Spillovers in the Classroom with Panel Data, With an Application to the Classroom”. Unpublished manuscript.
Bergquist, Constance C. 2006. “Encouraging Systematic Changes in Professional Development: A Short Summary of Florida's Evaluation Protocol System”. Unpublished report.
Betts, Julian R., Zau, Andrew C., Rice, Lorien A., 2003. Determinants of student achievement: new evidence from San Diego. San Diego: Public Policy Institute of California.
Boardman, Anthony E., Murnane, Richard J., 1979. Using panel data to improve estimates of the determinants of educational achievement. Sociology of Education 52, 113–121.
Bonesronning, Hans, 2004. The determinants of parental effort in education production: do parents respond to changes in class size? Economics of Education Review 23, 1–9.
Boyd, Donald, Grossman, Pamela, Lankford, Hamilton, Loeb, Susanna, Wyckoff, James, 2006. How changes in entry requirements alter the teacher workforce and affect student achievement. Education Finance and Policy 1 (2), 176–216.
Boyd, Donald, Pamela, Grossman, Hamilton, Lankford, Susanna, Loeb, James, Wyckoff, 2008. “Teacher Preparation and Student Achievement.” NBER Working Paper No. 14314.
Cameron, A. Colin, Gelbach, Jonah B., Miller, Douglas L., 2008a. “Robust Inference with Multi-way Clustering.” NBER Technical Working Paper No. 327.
Cameron, A. Colin, Gelbach, Jonah B., Miller, Douglas L., 2008b. Bootstrap-based improvements for inference with clustered errors. Review of Economics and Statistics 90, 414–427 August.
Clotfelter, Charles T., Ladd, Helen F., Vigdor, Jacob L., 2006. Teacher–student matching and the assessment of teacher effectiveness. Journal of Human Resources 41 (4), 778–820.
Clotfelter, Charles T., Ladd, Helen F., Vigdor, Jacob L., 2007. “How and why do teacher credentials matter for student achievement?” CALDER Working Paper #2.
Clotfelter, Charles T., Ladd, Helen F., Vigdor, Jacob L., 2010. Teacher credentials and student achievement in high school: a cross-subject analysis with student fixed effects. Journal of Human Resources 45 (3), 655–681.
Constantine, Jill, Player, Daniel, Silva, Tim, Hallgren, Kristen, Grider, Mary, Deke, John, 2009. An evaluation of teachers trained through different routes to certification. U.S. Department of Education, Washington, DC.
Dee, Thomas S., 2004. Teachers, race and student achievement in a randomized experiment. Review of Economics and Statistics 86 (1), 195–210.
Dickens, William T., Katz, Lawrence F., 1986. “Interindustry wage differences and industry characteristics.” NBER Working Paper #2014.
Dickens, William T., Ross, Brian A., 1984. “Consistent estimation using data from more than one sample.” NBER Technical Working Paper #33.
Ding, Weili and Steven F. Lehrer. 2005. “Accounting for Unobserved Ability Heterogeneity Within Education Production Functions”. Unpublished manuscript.
Eberts, Randall W., Stone, Joe A., 1984. Unions and public schools: the effect of collective bargaining on American Education. Heath, Lexington.
Feng, Li., 2009. Opportunity wages, classroom characteristics, and teacher mobility. Southern Economic Journal 75, 1165–1190.

Ferguson, Ronald F., Ladd, Helen F., 1996. How and why money matters: an analysis of Alabama schools. In: Ladd, Helen F. (Ed.), Holding schools accountable: performance based reform in education. Brookings Institution, Washington.
Florida Department of Education, Bureau of Educator Recruitment, Development, and Retention. 2007. “Professional Development Report – 06/07”. Unpublished data.
Florida House of Representatives, Schools and Learning Council. 2008. “Teacher Professional Development Programs in Florida”. Interim project report.
Garet, Michael S., Cronen, Stephanie, Eaton, Marian, Kurki, Anja, Ludwig, Meredith, Jones, Wehmah, Uekawa, Kazuaki, Falk, Audrey, Bloom, Howard S., Doolittle, Fred, Zhu, Pei, Sztejnberg, Laura, 2008. The impact of two professional development interventions on early reading instruction and achievement. U.S. Department of Education, NCEE, Washington, DC. 2008–4031.
Garet, M., Wayne, A., Stancavage, F., Taylor, J., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., Doolittle, F., 2010. Middle school mathematics professional development impact study: findings after the first year of implementation. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, NCEE, Washington DC. 2010–4009.
Goldhaber, Dan, Anthony, Emily, 2007. Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. Review of Economics and Statistics 89 (1), 134–150.
Grossman, Pamela L., Stodolsky, Susan S., 1994. Considerations of content and the circumstances of secondary school teaching. Review of Research in Education 20 (1994), 179–221.
Hanushek, Eric A., 1986. The economics of schooling: production and efficiency in public schools. Journal of Economic Literature 24 (3), 1141–1177.
Hanushek, Eric A., Kain, John F., O'Brien, Daniel M., Rivkin, Steven G., 2005. “The market for teacher quality.” NBER Working Paper #11154.
Harcourt Assessment. 2002. “SAT-10 to SAT-9 Scaled Score to Scaled Score Conversion Tables”. Unpublished data.
Harris, Douglas N., 2009. Toward policy-relevant benchmarks for interpreting effect sizes: combining effects with costs. Educational Evaluation and Policy Analysis 31 (1), 3–29.
Harris, Douglas and Tim R. Sass. 2006. “Value-Added Models and the Measurement of Teacher Quality”. Unpublished manuscript.
Harris, Douglas N., Sass, Tim R., 2009. The effects of NBPTS teachers on student achievement. Journal of Policy Analysis and Management 28 (1), 55–80.
Hill, Heather, Rowan, Brian, Ball, Deborah Loewenberg, 2005. Effects of teachers' mathematical knowledge for teaching on student achievement. American Educational Research Journal 42 (2), 371–406.
Houtenville, Andrew J., Conway, Karen S., 2008. Parental effort, school resources and student achievement. Journal of Human Resources 43 (2), 437–453.
Hoxby, Caroline, 2000. “Peer effects in the classroom: learning from gender and race variation.” NBER working paper #7867.
Jacob, Brian A., Lefgren, Lars, 2004. The impact of teacher training on student achievement: quasi-experimental evidence from school reform efforts in Chicago. Journal of Human Resources 39 (1), 50–79.
Jepsen, Christopher, 2005. Teacher characteristics and student achievement: evidence from teacher surveys. Journal of Urban Economics 57 (2), 302–319.
Kane, Thomas J., Jonah E. Rockoff and Douglas O. Staiger, 2006. “What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City.” Working Paper #12155.
Koedel, Cory. 2009. “An Empirical Analysis of Teacher Spillover Effects in Secondary School”. Unpublished manuscript.
Koedel, Cory and Julian Betts. 2009. “Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique”. Unpublished manuscript.
Koedel, Cory, Betts, Julian, 2010. Valued-added to what? How a ceiling in the testing instrument influences value-added estimation. Education Finance and Policy 5 (1), 54–81.
Krueger, Alan B., 2003. Economic considerations and class size. Economic Journal 113, 34–63.
Monk, David H., 1994. Subject area preparation of secondary math and science teachers and student achievement. Economics of Education Review 13 (2), 125–145.
Murnane, Richard J., Phillips, Barbara R., 1981. What do effective teachers of inner-city children have in common? Social Science Research 10 (1), 83–100.
Nye, Barbara, Konstantopoulos, Spyros, Hedges, Larry V., 2004. How large are teacher effects? Educational Evaluation and Policy Analysis 26 (3), 237–257.
Rice, Jennifer K., 2003. Teacher quality: understanding the effectiveness of teacher attributes. Economic Policy Institute, Washington.
Rivkin, Steven G., Hanushek, Eric A., Kain, John F., 2005. Teachers, Schools and Academic Achievement. Econometrica 73 (2), 417–458.
Rockoff, Jonah E., 2004. The impact of individual teachers on student achievement: evidence from panel data. American Economic Review 94 (2), 247–252.
Rothsein, Jesse, 2010. Teacher quality in educational production: tracking, decay and student achievement. Quarterly Journal of Economics 125 (1), 175–214.
Sass, Tim R., 2006. Charter schools and student achievement in Florida. Education Finance and Policy 1 (1), 91–122.
Spillane, James, 2005. Primary school leadership practice: how the subject matters. School Leadership and Management 25 (4), 383–397.
Todd, Petra E., Wolpin, Kenneth I., 2003. On the specification and estimation of the production function for cognitive achievement. Economic Journal 113 (485), F3–F33.
Wayne, Andrew J., Youngs, Peter, 2003. Teacher characteristics and student achievement gains. Review of Educational Research 73 (1), 89–122.
Wilson, Suzanne, Floden, Robert, 2003. Creating effective teachers: concise answers for hard questions. AACTE Publications, New York.
Wilson, Suzanne, Floden, Robert E., Ferrini-Mundy, Joan, 2001. Teacher preparation research: current knowledge, gaps, and recommendations. Center for the Study of Teaching and Policy, Seattle, WA.
Xu, Zeyu, Jane, Hannaway, Colin, Taylor, 2008. “Making a difference? The effects of teach for America in high school.” CALDER working paper #17.