Program on Education Policy and Governance Working Papers Series

It's Easier to Pick a Good Teacher than to Train One: Familiar and New Results on the Correlates of Teacher Effectiveness

Matthew M. Chingos
Program on Education Policy and Governance
Harvard University

and

Paul E. Peterson
Program on Education Policy and Governance
Harvard University

PEPG 10-22

Program on Education Policy and Governance
Harvard Kennedy School
79 JFK Street, Taubman 304
Cambridge, MA 02138
Tel: 617-495-7976 Fax: 617-496-4428
www.hks.harvard.edu/pepg/

It’s Easier to Pick a Good Teacher than to Train One:

Familiar and New Results on the Correlates of Teacher Effectiveness

by

Matthew M. Chingos

and

Paul E. Peterson

Paper prepared for A Symposium sponsored by the *Economics of Education Review*

December 10, 2010

It's Easier to Pick a Good Teacher than to Train One:

Familiar and New Results on the Correlates of Teacher Effectiveness

Matthew M. Chingos and Paul E. Peterson

Abstract

Neither holding a college major in education nor acquiring a master's degree is correlated with elementary and middle school teaching effectiveness, regardless of the university at which the degree was earned. Teachers generally do become more effective with a few years of teaching experience, but we also find evidence that teachers may become less effective with experience, particularly later in their careers. These and other findings with respect to the correlates of teacher effectiveness are obtained from estimations using value-added models that control for student characteristics as well as school and (where appropriate teacher) fixed effects in order to measure teacher effectiveness in reading and math for Florida students in fourth through eighth grades for eight school years, 2001–02 through 2008–09.

In recent years, it has become the conventional wisdom that teachers vary substantially in their effectiveness at lifting student classroom achievement, as measured by their performance on standardized tests.[1] Despite that variability, it has been difficult for scholars to identify types of training that correlate well with teacher effectiveness. Teacher classroom performance is correlated neither with the type of certification a teacher has earned, nor with the acquisition of an advanced degree, nor with the selectivity of the university a teacher attended. Only on-the-job training that comes with each year of experience in the classroom has been regularly identified as a correlate of teacher effectiveness.

The emerging consensus depends upon a limited number of studies, however, so it is worth continuing to scrutinize available information to see whether findings can be replicated as

[1] Support for this research was received from the William Simon Foundation and the Searle Freedom Trust. The data were generously provided by the Florida Department of Education. We are grateful to staff members Tammy Duncan and Jeff Sellers for their assistance in this regard as well as to Antonio Wendland for administrative assistance and to Ashley Inman for research support. This is a revised version of a paper originally prepared for the Conference on "Merit Pay: Will it Work? Is it Politically Viable?" sponsored by Harvard's Program on Education Policy and Governance, Taubman Center on State and Local Government, Harvard's Kennedy School, 2010. Helpful comments were provided by Brian Gill, Eric Hanushek, Brian Jacob, Matthew Kraft, John Papay, Steven Rivkin, Guido Schwerdt, Martin West, and conference participants at Harvard University.

well as to explore certain lacunae in the literature. Due to data limitations, most prior studies of pre-service training, for example, have relied upon crude indicators of the type of training a teacher has received—whether or not the teacher is certified, or whether or not the teacher has attended a selective university. Important effects produced by specific university training programs could be masked by lumping pre-service training into such broad categories. Similarly, estimates of the impact of acquiring a master's degree have usually been unable to adjust for the fact that only some teachers pursue an advanced degree. If those pursuing advanced training do so in order to make up for teaching deficits, prior research may have under-estimated the benefits from that training. Also, the impact of specific masters' degree programs has not previously been estimated.

Prior estimations of on-the-job training (years of experience) also suffer from certain limitations. Most do not take into account the fact that more experienced teachers are a selected population. And those that do adjust for that fact (by conditioning on teacher fixed effects) have usually been unable to track the effects of experience of individual teachers over eight years of employment. In short, the effects of on-the-job training over the teaching life cycle have yet to be precisely estimated.

Finally, the literature has yielded inconstant findings with respect to certification by the National Board of Professional Teaching Standards (NBPTS). Some studies show positive impacts, while others show little impact. Further research on this question is clearly merited, especially since the program is growing in size.

In this paper we provide additional information on the correlates of teacher effectiveness, as estimated by the additional amount of learning taking place by students in their classroom (adjusted for student background characteristics). In addition to confirming propositions conventionally believed to be true, we find that effectiveness does not vary by the specific university one has attended. We also show that the value of additional on-the-job training decays after several years of teaching and turns negative after a decade or more of teaching, particularly when measured by low-stakes tests. Finally, we show that teachers certified by NBPTS are more effective than those not so certified, but the certification process yields no value-added in teacher effectiveness, except perhaps for teachers of elementary school math.

These findings come from data made available by the Florida Department of Education from its K–20 Education Data Warehouse (EDW). These data include test score performance in reading and math for students in grades four through eight for the years 2002 to 2009. (All years are identified by the year in which the school year ends, the point in time when standardized tests are administered.) Administrative data on student characteristics, teacher characteristics, school characteristics, and the specific teacher each student had each year were also made available.

Prior Research

In recent years, research on the value of pre-service, in-service, and on-the job training has begun to accumulate rapidly, as many school districts and states have made available test score information to members of the scholarly community.

Pre-Service Teacher Preparation

Nearly every state requires teachers to have earned both a bachelor's degree and taken a certain number of courses in the field of education in order to receive a teaching license from the state. The practice has become so pervasive that the 2002 federal law, No Child Left Behind, required that by 2006 all teachers in schools that received compensatory education funding be "highly qualified," defined as "a bachelor's degree as well as a state teaching license and demonstrated competence in the academic subject(s) he or she teaches" (Murnane and Steele

2007, p. 24). The number of required courses for certification is usually around 30 credits in the subject matter and for the specific grade (elementary, middle, or high school) being taught, about the same number that is required for a college major in the subject. School districts, in emergency situations, are able to hire temporary teachers without certification, and in some states teachers may be hired provided they begin a set of courses leading toward what is known as alternative certification (Hess et al. 2004; Nadler and Peterson 2009).

Prior econometric research has generally failed to detect positive impacts of pre-service teacher preparation programs on student learning. In their review of the literature, Murnane and Steele (2007, p. 24) conclude that "in general, empirical studies find little or no difference in average effectiveness between those teachers who are traditionally licensed and those who enter the profession through alternative routes." In recent studies of high quality, Clotfelter et al. (2006) found no benefits from certification in North Carolina and Kane et al. (2006) found no differences in the effectiveness of certified, non-certified, Teach for America (TFA), or teachers recruited through a special New York City initiative.

In Florida alternative training programs have been developed for teachers (Llaudet 2006, pp. 38–39; Moe 2006), and the state's department of education (2009) found that teachers who received their training in such programs were no less effective than traditionally certified teachers. However, the methodology employed in this study makes a number of highly restrictive assumptions. In another Florida study, using data similar to ours but for an earlier time period, Harris and Sass (2008) report no correlations between majoring in education and classroom effectiveness. They do find that math training may be associated with effectiveness in teaching high school math, but they caution the reader to interpret these results cautiously because of the sample size limitations they faced (Harris and Sass 2008, p. 55).

As mentioned above, it is possible that a particular teacher preparation program could be particularly effective, despite these negative findings about certification programs in general. If one could identify each degree program separately, one might find some to be effective and others to be ineffective.

University Selectivity

It has been generally assumed that graduates from more selective institutions of higher education are superior teachers. Unless "teaching requires a very idiosyncratic set of skills," Ballou and Podgursky (1997, p. 10) have argued, it is likely that students from selective colleges outperform others, inasmuch as studies of employees in other occupations "have found a positive relationship between earnings and the quality of the college attended." Hoxby and Lehigh (2004) make a similar assumption when they conclude that the quality of the teaching force has declined after showing a drop in the proportion of women teachers graduating from selective colleges. In 1990, Teach for America (TFA), a non-profit organization, began recruiting teachers from the pool of graduates of highly selective colleges, apparently on the assumption that those talented enough to win admission to Ivy League or other selective colleges will be more effective teachers—or that education at such colleges provides better preparation for teaching.

Early studies provided some support for these claims, as a number found positive correlations between quality of university attended and teacher effectiveness (Ehrenberg and Brewer 1994; Summers and Wolfe 1977). However, Kane et al.'s (2006) study of New York teachers found little differences between teachers recruited from highly selective colleges by TFA and other teachers. However, those recruited by TFA were not certified, so the comparison is not definitive. Clotfelter et al.'s (2006, 2007a) North Carolina studies found no relationship between college selectivity and teacher effectiveness in elementary and middle schools. However, Clotfelter et al. (2007b) did find a significantly positive relationship between college selectivity and the effectiveness of high school teachers. Boyd et al. (2008) find at most a weak correlation between selectivity and effectiveness in New York City (controlling for SAT scores).

A report issued by the Center for Education Policy Research at the Harvard Graduate School of Education (2010) found considerable variation in the effectiveness of teachers in the Charlotte-Mecklenberg, North Carolina school district, depending on the higher educational institution at which they were trained. However, the report does not report its analytic strategy in detail, does not identify the institutions of higher education included in its analysis, and is unable to ascertain whether the observed variation could be due to patterns of teacher selection particular to this specific district. The State of Florida (2009), in an analysis yet to be formally presented, also found variable college training impacts on teacher effectiveness, but its estimate of teacher effectiveness has important limitations. In short, there is no state-of-the-art, statewide study of the relative effectiveness of specific university teacher preparation programs.

On- the-Job Training (Teacher Experience)

It is conventionally believed that on-the-job teacher training is effective, because students learn more from teachers with additional years of experience, especially in the first two or three years (for a review of the early literature, see Rockoff 2004, p. 2). But as Figlio (1997) points out, many estimates of effectiveness returns to on-the-job training (years of teaching experience) may be upwardly biased because the estimations do not account for the probable attrition from the teaching force of less effective teachers (Clotfelter et al. 2006; Rivkin et al. 2005). Even these studies find few positive benefits from experience beyond the initial years, however. Rivkin et al. (2005, p. 449) conclude that “there is little evidence that improvements continue after the first three years" of teaching, Clotfelter et al. (2006, p. 28) conclude that the "benefit [to fifth graders in North Carolina] from having a highly experienced teacher is approximately one tenth of a standard deviation on reading and math test scores," but admit that "roughly half of this return occurs for the first one or two years of teaching experience."

To avoid selection bias, Rockoff (2004) estimated effectiveness returns to experience on math computation, math concepts, vocabulary and reading comprehension in two New Jersey school districts after conditioning on teacher fixed effects. He found marginal returns to the initial years of experience—overall, the linear effect varies between insignificant and 0.07 standard deviations, depending on the subject—but little in the way of additional returns after five years, except, perhaps, in reading comprehension. He did not estimate experience effects beyond eleven years of teaching (his specification assumed that there are not any returns at that point). Papay and Kraft (2010) do find returns to experience after the tenth year in one large urban school district.

Harris and Sass (2008), also employing a teacher fixed effects model, estimate on-the job training effects on the performance of all students in grades three through ten in Florida. On the whole, its estimations reveal effects half the size of the effects reported for North Carolina fifth graders (which were at risk of selection bias). Harris and Sass (2008, p. 19) report that "the bulk of the experience effects occur in the first year, with subsequent experience yielding diminishing increases in teacher productivity." Their analysis is limited to impacts on performance on Florida's high-stakes test. All comparisons are between rookie teachers and those with more experience, making it difficult to ascertain whether returns to experience turn in a negative direction. The Harris and Sass (2008) study is of particular interest here, because it draws upon data from Florida's EDW for the five years 2000–2004, while the results reported below are for the eight-year period 2002 to 2009. The data sets thus overlap but are not identical, and results reported here track teachers over at least three additional years (Florida only began testing all students in grades three to eight in 2001). We compare and contrast findings from the two studies in more detail below.

Despite the array of studies that have estimated the effectiveness returns for teachers to on-the-job training, certain avenues have not been fully explored. No previous study has detected a point at which the returns to experience turn downward, a point of some policy interest since teacher salary schedules generally reward teachers for additional year of experience (see Table 1 for the average salary schedule for 47 out of the 67 school districts in Florida).[2]

Teacher Examinations

In recent years a number of states began to ask teachers to pass an examination before they can be granted a license to teach. Such policies are supported by early studies that identified a correlation between verbal ability and teacher effectiveness (Ferguson and Ladd 1996; Hanushek 1971). A more recent review of the literature by Hanushek and Rivkin (2003, Table 2) identified positive effects of tests on performance that were statistically significant in only two of nine estimations. Specifically, Harris and Sass (2008) found no correlation between SAT performance and classroom effectiveness in Florida. Boyd et al. (2008) found a positive correlation with math SAT scores, but a negative correlation with verbal SAT scores in New York City. Clotfelter et al. (2006), however, did find a correlation (albeit a very weak one) between licensure test performance and classroom effectiveness in North Carolina.

Certification by the National Board of Professional Teaching Standards

[2] Table includes all Florida districts for which information was available on district websites in June 2010.

The NBPTS, established in response to a 1986 Carnegie report, has gradually established itself as a small but vibrant private, non-profit certifying agency. Many states provide additional compensation to those teachers who receive Board certification. To become certified, applicants must take appropriate tests, display lesson plans and other evidence of their teaching methods, and provide videos of their classroom instruction. Roughly half the applicants receive certification. As of December 2009, 82,000 teachers nationwide had received National Board certification; about six percent of the teachers in our sample had National Board certification. In Florida, teachers may receive an additional 10 percent of the average salary in the state annually if they are certified by the National Board, though that amount may be less if the state budget is constrained. (In 2009, they received an additional $3,800.)

The value of NBPTS certification remains uncertain, as results have varied with the study undertaken. In Los Angeles, Cantrell et al. (2008) identified a correlation between effectiveness and scoring high on an application for NBPTS certification, and in a study of Board-certified teachers in North Carolina, Ladd et al. (2007) found the correlation between certification and performance in that state to be quite substantial.[3] However, in Florida these same analysts found only a small correlation in reading and an insignificant relationship in math for students in grades four and five during the years 2001 to 2004.

Master's Degree

Seventeen states require that school districts augment the salaries of teachers who hold a master's degree. Elsewhere, many school districts bind themselves to provide such increments in collective bargaining agreements with local teacher unions. The amounts involved are often

[3] Goldhaber and Anthony (2007), in a similar study using data from North Carolina, also find a correlation between NBTPS certification and performance. They also present evidence that completing the NBTPS certification process does not increase teacher effectiveness.

substantial. For example, in 2010 a fifth-year Florida teacher with a master's degree earned seven percent more than an equally experienced colleague with a bachelor's degree (Table 1). In response to such monetary incentives, almost half of elementary and secondary public-school teachers have acquired a master's degree or better (Murnane et al. 1991, p. 117; U.S. Department of Education 2009, table 68). However, little or no impact of an advanced degree on student learning has been detected (Clotfelter et al. 2006; Coleman et al. 1966; Goldhaber 2002; Hanushek 1986; Harris and Sass 2008; Rivkin et al. 2005). However, most studies do not estimate the impact of the acquisition of an advanced degree by comparing individual teacher performances before and after the year the degree was acquired. Harris and Sass (2008) do provide such estimations but must rely on a data from a five-year time period and find inconsistent results.

Data

Our extract from Florida's Education Data Warehouse (EDW) contains observations of every student in Florida who took the state assessments from the school years 1999 to 2009, with each student linked to his or her courses (and corresponding teachers) for the school years ending in 2001 through 2009. Test scores include results from both the Florida Comprehensive Assessment Test (FCAT), which is the state accountability system's "high-stakes" test, and the Stanford Achievement Test, which is a nationally norm-referenced test that was administered to students together with the FCAT until 2008. Beginning in 2001, students in grades three through ten took both tests each year in math and reading.[4] Thus annual gain scores can be calculated for most students in grades four through ten beginning in 2002. The data also contain information on the demographic and educational characteristics of each student, including gender,

[4] Florida stopped administering the Stanford Achievement Test in 2009, so results that use this test are based on one less year than our FCAT results.

race/ethnicity, free or reduced-price lunch eligibility, limited English proficiency status, special education status, days in attendance, and age. The data allow for tracking of individual teachers for up to 8 years, the first time this has been possible on a statewide data set that provides links between teachers and the performance on standardized tests of students in their classes.

The EDW data also contain detailed information on individual teachers, including their demographic characteristics, experience, and the Florida college from which they received their degree and their major field of study. No information on college attended is available for teachers who earned their degree in another state or for teachers who earned their degree in Florida prior to 1995. Tables 2A and 2B show that the subset of teachers for whom we have college attended are younger (and thus less experienced) than the full sample of all teachers, as we would expect. However, their other characteristics are largely similar to the full sample of teachers, and both groups teach similar student populations. Data for teachers born in 1973 or later (who would have graduated from college in 1995 or later) indicate that about half of these teachers attended a public university in Florida. Estimates of the relationship between effectiveness and teacher characteristics among teachers who attended a public university in Florida are general similar to those based on teachers who went to college elsewhere.[5]

We constructed a file based on course enrollment data (that matches students and teachers) in order to identify the students for which any given teacher was responsible in a

[5] These results are available from the authors upon request. Estimates of the returns to experience (not controlling for teacher fixed effects) are similar for the two groups of teachers, as are the estimates of the relationship between National Board Certification and teacher effectiveness. The one exception is the estimate of the difference in effectiveness between teachers with M.A. and B.A. degrees, which is positive and statistically significant (with a magnitude of 0.01–0.02 standard deviations) for elementary school teachers that did not attend a Florida public university, but close to (and statistically insignificant from) zero for teachers that attended a public university in Florida. This difference is not observed for middle school teachers.

particular year.[6] The teacher experience variable we construct reflects all years the teacher has spent in the profession, including both public and private schools in both Florida and other states.

Florida's EDW has made available information on the college from which teachers received their bachelor's degrees, provided that degree was earned from a Florida state college or university between 1995 and 2008. Twenty-two percent of the teachers included in our analysis graduated from one of the eleven Florida public universities listed in Table 3. As that table shows, the universities differ from one another in the degree of selectivity of their admissions policies as well as in many other characteristics.

The selectivity of colleges and universities in Florida is based upon the rankings given by *U.S. News and World Report's* ranking of institutions of higher education (2009). In our analysis, the eleven public institutions of higher education within the state are divided into three categories: very selective (*U. S. News's* "tier one"), selective (*U. S. News's* "tier 3"), and less-selective (*U. S. News's* "tier 4" and "non-competitive"). Peterson's Guide to Colleges and University's (2009) says that "entrance difficulty" into the three schools *U. S. News* identifies as "tier one" schools is "very difficult," while it says that entering the other colleges and universities is only "moderately difficult" and denotes St. Petersburg College as "non-competitive."

[6] For elementary students, the course files do not always clearly identify the student's regular classroom teacher. In order to match the maximum number of students to their teachers, we examined students' general (e.g., self-contained classroom), math, and reading teachers and matched them to the one or two teachers with whom they spent at least 40 percent of their academic (general, math, and reading) time. We then dropped students who were matched to two teachers and students who were not matched to any teachers (32 percent of all students in these grades). A large and increasing number of middle school students in Florida appear to have more than one regular classroom teacher, perhaps reflecting an increase in team teaching. We match middle school students to their primary reading and math teachers in a similar fashion. For each subject, we linked each student to the teacher(s) with whom they spent at least 40 percent of their time in that subject and dropped students who were matched to two teachers in a given subject as well as those who were not matched to any teachers in that subject (16 percent of students in math and 38 percent of students in reading).

The selectivity of the university is correlated with a variety of other university characteristics. As can also be seen in Table 3, very selective and selective universities typically admit students with higher SAT scores and grade point averages than do less selective colleges and universities. For the most part, they ask in-state students to pay higher tuition. The size of the teacher preparation programs varies substantially—from as few as 92 students in 2009 at Florida A&M to 1,002 at the University of Central Florida. The percentage minority is similar, except a historically black school, Florida A&M, has a comparatively large African American student body, and Florida International University has a comparatively large Hispanic student body.

In short, higher educational institutions in Florida vary in many measurable respects. Given the range of variation, one may hypothesize that some colleges and universities are more successful in turning out graduates who will be more effective teachers.

Methodological Approach

To estimate a teacher's effectiveness, we estimate the gain or value-added in student performance over the previous year for those students who were in that teacher's classroom that year, controlling for a wide range of student characteristics as well as for school (and, at times, teacher) fixed effects. Such value-added estimations are said, by some, to be biased by the fact that students and teachers are not randomly assigned to classrooms. The more able students are assigned to teachers perceived to be effective (perhaps as the result of pressure by well-informed parents). As a result, teachers appear to be adding value even though they are simply working with students inclined to learn on their own. Low-performing children are disproportionately likely to be asked to study under less effective teachers (Rothstein 2010).

The critique is not implausible, but the value-added models have survived a variety of validity tests. Most importantly, their estimations of teacher effectiveness have been shown to resemble the estimations obtained from experiments in which students and teachers have both been randomly assigned to classrooms (Kane and Staiger 2008; Nye et al. 2004). Other external validity checks strengthen the case for the validity of the value-added approach. Harris and Sass (2007) and Jacob and Lefgren (2008) find a correlation between principal evaluations of teacher effectiveness and those generated by value-added models. In another validity test, Chingos and West (2010) find positive correlations between the effectiveness of teachers and their earnings after they left the profession.

It is worth mentioning, however, that value-added models appear to have the greatest validity when the estimations control for a variety of student background and classroom characteristics while also controlling for school fixed effects so that all comparisons of teacher effectiveness compare the teacher with others at the same school.[7] For this reason, the preferred models presented in this paper include both school fixed effects and multiple controls for student and classroom characteristics. Student-level controls include prior-year test scores (in both subjects, along with squared and cubed terms), gender, race/ethnicity, free and reduced lunch, English Language Learners, special education, non-structural move, structural move, days absent previous year, and whether repeating the grade. Classroom-level controls include the previous variables aggregated to the classroom level (with the exception that prior-year test scores are replaced with percent proficient on the FCAT in reading and math in the prior year) and class

[7] Estimates that condition on school fixed effects may still suffer from bias that results from the non-random matching (on unobservable characteristics) of teachers and students within the same school. One empirical strategy to deal with non-random student-teacher matching within school-grade-year cells is to aggregate the data to the school-grade-year level and estimate models that condition on school-grade and school-year fixed effects. However, this is not feasible in our analysis given the substantial amount of missing data on postsecondary institutions attended by teachers.

size. Teachers in classes in which 25 percent or more of the students were in special education were excluded from the analysis.[8]

The model, then, is

$$A_{ist} = \alpha_1 Selective_t + \ \alpha_2 VerySelective_t + \omega A_{i,t-1}, + \beta X_{it} + \ \gamma C_{ist} + \ \delta_s + \pi_{it} + \epsilon_{ist},$$

where A_{it} is the test score of student *i* in school *s* in year *t* (standardized by grade and year to have a mean of zero and standard deviation of one); $Selective_t$ and $VerySelective_t$ are dummy variable indicating whether the teacher attended a selective of very selective university (as compared to a less selective university); $A_{i,t-1}$ includes the student's prior-year test scores in both subjects (and their squared and cubed terms); *X* and *C* are student- and classroom-level characteristics; δ_s is a vector of school fixed effects; π_{it} is a vector of grade-by-year fixed effects, and ϵ_{ist} is a standard zero-mean error term adjusted for clustering at the teacher level. We estimate this equation separately for reading and math performance for students in grades four and five, which are referred to as elementary grades, as well as for students in sixth through eighth grades, which are referred to as middle school grades, as most Florida students attend such schools at these grade levels. We also estimate versions of this equation that include teacher characteristics, such as experience, college major, race/ethnicity, and gender.

Most of our estimates of the relationship between classroom effectiveness and various teacher characteristics should not be interpreted as causal, as teachers choose their college major and the university they attend.[9] In the absence of random assignment to these conditions, we

---

[8] We obtain qualitatively similar results when we include classes with a large share of special education students.

[9] In addition to the selection process that matches potential teachers to universities, schools also select teachers in the hiring process. Our findings could be biased if less effective potential teachers are less likely to find employment at any school. However, that is unlikely both because it is difficult to identify ineffective teachers before hiring them and because the demand for teachers in Florida was rising rapidly during the time period covered by this study (as the result of steep school-age population growth and the need for more teachers required by the implementation of a constitutionally mandated reduction in class sizes).

know only whether the characteristic is associated with classroom effectiveness, not whether any association is causal.[10] However, causal interpretations may be inferred from findings with respect to the acquisition of a master's degree, National Board certification and teacher experience, if the models include teacher fixed effects, provided one can assume teachers do not change in ways correlated with unobservables before and after the change in these teaching-relevant characteristics.

Results

Our results are reported in standard deviations. In interpreting the size of the effects reported in the tables, it is worth bearing in mind that the effectiveness in math and reading of the elementary grade teachers included in our sample has a standard deviation of 0.11 and 0.05 student-level standard deviations, respectively. In other words, students taking a class from a teacher whose effectiveness is one standard deviation above the average teacher will learn 0.11 standard deviations more in math and 0.05 standard deviations more in reading. Among middle school teachers, the standard deviation of effectiveness is smaller—0.07 and 0.03 standard deviations for the two subjects, respectively, perhaps because no one teacher in these grades has as sustained a relationship with the students in middle school as in elementary school.[11]

---

[10] We interpret our value-added estimates of teacher impacts on student performance as causal, even though our data is observational, not experimental, because, as discussed above, our value-added models have been validated by experimental estimations.

[11] The standard deviations in teacher effectiveness reported in the text—which are corrected for measurement error—are smaller than those reported in other studies of teacher effectiveness. Hanushek and Rivkin (2010) report results from ten studies as showing an average of 0.15 and 0.11 standard deviations in math and reading, respectively. It is not stated whether the studies are of elementary or middle school teachers. It is possible that the rigorous Florida accountability system restricts the variability in teacher practice in that state. The method we use to calculate a standard deviation in teacher effectiveness is described in Chingos and West (2010).

Certification

Table 4 shows the relationship between teacher effectiveness and many of the teacher characteristics previously examined in the literature. Our results are broadly consistent with this literature. We find no difference in the classroom effectiveness of those with an education major and those with a major in another subject (which in the absence of a master's degree means the person is not certified or has alternative certification). We do find that teachers who are certified in education, but outside the field of elementary education, are less effective as teachers of students in the elementary grades, but that finding should be interpreted cautiously as only four percent of the teachers at those grade levels with available data had an education degree other than one in elementary education.[12]

State Examinations and National Board Certification

We also find a fairly strong positive relationship between certification exam performance and classroom effectiveness (except for the reading performance of elementary students). But the policy significance of that finding is diminished considerably when one considers that only two percent of all the teachers in our sample found a way to fail the test one or more times. Certification by the National Board is correlated with achievement in math and reading in both elementary and middle school. The average differences in effectiveness between teachers with and without this credential ranges from 0.02 to 0.03 student-level standard deviations, which is about 25-30 percent of a standard deviation in teacher effectiveness in the elementary grades and 40-60 percent of a standard deviation in the middle school grades. These differences are not

[12] For teachers of middle-school students, we find a slight negative impact on reading performance of having a middle-school education major (relative to teachers with a degree in English education), and a slight positive impact of having an elementary school major (relative to those having a math education major) on math performance, but the effects are too modest and too inconsistently observed to be taken as a guide to policy.

trivial, but are also not particularly large considering that only 5-7 percent of teachers in Florida are certified by the National Board.

So, contrary to Ladd et al.'s (2007) findings for Florida, our findings lend some support for the policy of giving extra compensation to Florida teachers who pass the National Board examination.[13] However, the findings do not demonstrate that participation in the National Board certification process enhances teacher effectiveness. In a separate analysis, reported in Table 5, we use a teacher fixed effects model to identify differences in a teacher's performance before and after acquiring certification.[14] We observe only one significant increase in the effectiveness of these teachers upon certification, which is in elementary school math. However, this significant result does not appear on the low-stakes Stanford Achievement Test (results not shown).

Advanced Degrees

Table 4 also shows that advanced degrees—both masters' degrees and doctoral degrees—do not identify better teachers.[15] That finding replicates prior studies, which also simply compare teachers who have advanced degrees with those who do not. However, that estimate may under-estimate the value of an advanced degree if those who are lower performing teachers

---

[13] Ladd et al. (2007) find significant correlations between National Board certification and effectiveness in North Carolina. Why their Florida results do not yield the same results as ours for elementary math teachers is unclear, because our data sets overlap, though we have access to five more recent years of data than were available at the time of the Ladd et al. (2007) study (the results of which are based on data from 2001 through 2004, whereas our results are based on data from 2002 through 2009). It may be that the National Board procedures have improved with time. Or it may be that estimations are sensitive to model differences when the percentage of teachers having National Board certification is only six percent of the total.

[14] The results in Table 5 are only based on teachers that we observe in the same school both before and after obtaining the listed credential, so the sample sizes are much smaller than in the results that include teachers that we only observe with or only without that credential.

[15] We do find a statistically significant relationship between holding a master's degree and middle school gains in reading achievement, but the correlation is so small (.003 standard deviations) as to be substantively meaningless, especially given the fact that all other estimations are statistically insignificant.

try to improve their effectiveness by pursuing additional training. To see whether the acquisition of a master's degree lifts teacher effectiveness, we included teacher fixed effects in our model. If this credential provides some sort of training that enhances teachers' classroom effectiveness, we would expect individual teachers to become more effective after receiving the credential. Table 5 indicates that this is not the case; estimates are small and statistically insignificant.

On-the-job-training

Table 4 shows a positive relationship between student achievement and the number or years of service a teacher has provided. However, these effects likely reflect bias introduced by the higher attrition rate of less effective teachers. This bias can be ameliorated by taking teacher fixed effects into account, but teacher effects cannot simply be added to a model that includes year effects because experience and year are highly correlated within teachers over time. There is no ideal solution to this problem, so we present results from three different models that each make a different key assumption.[16] We follow the terminology and conventions used in Papay and Kraft (2010) to describe and estimate these models. The "simple experience model" assumes that there are no year effects by omitting them from the model. This assumption is not particularly plausible in Florida over this time period, as student achievement was rising fairly steadily.[17] The "censored growth model" (originally proposed by Rockoff [2004]) assumes that there are no returns to experience after a certain period of time (we chose 20 years) and uses

[16] We do not observe any teachers for more than eight years, so all of our estimates of the entire effectiveness-experience relationship also require making transitivity assumptions (e.g., that one can extrapolate by comparing the change between six and 11 years of teaching with the change between one and seven). Returns are thus estimated by making within-teacher comparisons among teachers who remain in the classroom at different points in their teaching careers. For example, our estimates of the effectiveness-experience relationship beyond 30 years of experience are based on teachers who remained in the classroom after they may have been eligible for retirement.

[17] Throughout our analysis we standardize test scores by grade and year, but there still can be year effects because of our inclusion of control variables. In other words, we have standardized the test scores to have the same mean each year, but the average test score can still be higher than we would expect in a given year based on the student characteristics included in the models.

teachers with experience above this cutoff to estimate the year effects.[18] The results of this model will be biased if there are returns to experience beyond 20 years.

Finally, we implement a model recently proposed by Papay and Kraft (2010), which makes a different assumption than the other two models. The idea behind their "two-stage model" is that the estimates of the year effects do not need to condition on teacher fixed effects if the assumption is made that the teacher fixed effects are uncorrelated with the year effects after controlling for teacher experience. In other words, this model assumes that there are no changes in average teacher quality (conditional on experience) over time. In the first stage of the model, the year effects are estimated conditional on all of the control variables used in the other two models with the exception of teacher fixed effects. In the second stage, which includes teacher fixed effects as well as the control variables from the first stage, the coefficients on the year effects only are constrained to their first-stage values (all other coefficients are estimated).[19]

The advantage of this model is that it can estimate returns to experience beyond the cutoff year imposed by the censored growth model (20 years in our case) while also controlling for the year effects that are ignored by the simple experience model. The disadvantage of the two-stage model is that it will produce biased results if there are changes in average teacher quality over time, which may have occurred in Florida during the time period covered by our data, particularly because many new teachers were being hired to implement a class-size reduction policy (Chingos 2010). Below we quantify the changes in average teacher quality over this time period.

[18] We estimate this model by including dummies for the first 20 years of experience (top-coding teachers with more than 20 years experience) and controlling for grade-by-year effects and teacher*school fixed effects.

[19] In practice, we use a full set of grade-by-year effects in place of year effects and teacher*school effects in place of teacher fixed effects (the first stage includes school fixed effects only).

The results are presented separately by grade range and subject in Figures 1a through 2b. The results for the elementary grades are quite sensitive to the model chosen. All three models suggest an initial bump in effectiveness after the first year in both math and reading, but the effectiveness-experience profiles diverge after that (Figures 1a and 1b). The simple experience model indicates that the effectiveness-experience relationship turns negative after a few years in math and after about 15 years in reading. The censored growth model indicates that effectiveness peaks after about 5 years in both math and reading, with declines thereafter in both subjects. The two-stage model indicates more steady growth over a longer period of time in both subjects in these grades, with a leveling off not occurring until about 25 years of experience.

In the middle school grades, the simple experience model indicates steady declines in effectiveness after the first year of teaching in both reading in math (Figures 2a and 2b). The censored growth and two-stage models produce fairly similar results to each other in reading, indicating an essentially flat experience-effectiveness profile (with the two-stage model showing a decline in effectiveness after about 15 or 20 years). In math, the censored growth model does not show any returns to experience after a few years, whereas the two-stage model shows more steady (but still modest) returns until about 20 years of experience.

Which of these three models is to be preferred? As indicated above, all three models make assumptions that may be violated by events in Florida during the period covered by our data. The absence of any year effects implied by the simple experience model seems highly implausible given the gains in student achievement made in Florida over this time period, so this model is our least preferred. The assumption made by the two-stage models—that there are no changes in average teacher quality conditional on experience over time—can actually be tested by calculating teacher value-added estimates and examining whether they change over time (controlling for experience and year effects).[20] Although there does not appear to be a strong linear trend in teacher effects over time (the correlation between value added and a continuous year term is quite small), a more flexible model shows modest (non-linear) variation in average value added over time in the elementary grades. Compared to its average value in 2002, average math value added in grades was about the same in 2003, 0.02 standard deviations lower in 2004 through 2008, and 0.01 lower in 2009. Results for reading value added in the elementary grades are similar.[21] In the middle school grades, average value added in math varied over time but in a somewhat less consistent fashion: compared to its average value in 2002, average value added was about the same in 2003, 2004, and 2006; 0.02 standard deviations lower in 2005 and 2009, and 0.03 lower in 2007 and 2008. Value added in reading in these grades varied somewhat less.[22] This result is consistent with the censored growth and two-stage models producing more similar estimates in middle school reading than in middle school math and in the elementary grades.

For the middle school grades, these data provide clear evidence of no returns to on-the-job training after the first year or two, and raise the possibility of negative returns to experience over the course of teachers' careers. For the earlier grades, the results are much less consistent but two of the three models show no evidence of positive returns to experience beyond the first

[20] Specially, we regress student test scores on student and classroom controls, grade dummies, and school fixed effects, then aggregate the residuals to the teacher level (weighting each year equally). These are our value-added estimates. We then regress these estimates on year dummies (or a continuous variable for year), teacher experience, and school fixed effects.

[21] Compared to its average value in 2002, average reading value added in the elementary grades was 0.01 standard deviations lower in 2003, 0.03 lower in 2004, and 0.02 lower in 2005 through 2009.

[22] Compared to its average value in 2002, average reading value added in the middle schools grades was about the same in 2003, 2006, and 2009; 0.02 standard deviations lower in 2004 and 2007, and 0.01 standard deviations lower in 2005 and 2008.

five years of teaching, as well as declines in effectiveness in math later on. Results based on low-stakes test scores are qualitatively similar to the results using the high-stakes FCAT.

University Selectivity

The correlation between the selectivity of the university from which teachers received their degree and average gains in achievement in math and reading during the year students were in their classrooms is statistically insignificant, except in one instance where the relationship is the opposite sign from what one would expect. Table 4 shows that elementary students performed 0.017 standard deviations less well on the math examination if they had a teacher from a selective college rather than a less selective one. That accounts for about 15 percent of a standard deviation in teacher effectiveness in elementary math instruction, a non-trivial finding, but the coefficient is the only one that is significant out of eight estimated in Table 4, raising questions as to whether it occurred by chance, especially since the relationship seems perverse.

The model presented in Table 4 may be inappropriate for estimating the relationship between university selectivity and teacher effectiveness, however. For one thing, it may suffer from omitted variable bias in that it does not include controls for teacher race/ethnicity and gender, which could be correlated with university selectivity and effectiveness in the classroom. We exclude teacher race/ethnicity and gender from Table 4 in order to focus solely on qualifications, but it may be desirable to control for these characteristics when estimating the effects of university selectivity.

At the same time, a number of the variables presented in Table 4 may not be appropriately included in models estimating the connections between university selectivity and effectiveness on the grounds that they themselves may also be consequences of the higher education experience a student has had. Passage of the certification examination, the pursuit of a master's or doctoral degree, and NBPTS certification could all be a function of the college one attended. To include them in the selectivity model may alter the estimated relationship between selectivity and effectiveness.

To ascertain whether results are sensitive to the specific model that is estimated, five models are presented in Tables 6A and 6B.[23] Model 1, the naïve model, reports the simple relationship between selectivity and effectiveness with no controls for any teacher, student or school characteristics whatsoever (except for grade-by-year fixed effects). As can be seen in column 1 of these two tables, administrators who rely on a naïve model would conclude that they are recruiting better teachers if they obtain them from more selective institutions of higher education. All coefficients in the naïve model are statistically significant and some are very large, as much as 0.11 standard deviations.

In column 2, however, one finds estimations that control for student and classroom characteristics and school fixed effects.[24] Once these controls are introduced, the size of the coefficients dwindles to insignificant levels, with many of the signs becoming negative, one of which is significant (as discussed above). The prospective teacher coming from a well-regarded university appears to have first chance at preferred teaching positions; their apparent accomplishments appear to be due more to the situation in which they find themselves than the training they have received.

If all the other characteristics of a teacher are in some way influenced by the university a teacher attended, the models presented in the second column of these tables are the most

[23] These models are estimated based only on teachers for whom data on college attended are available, whereas Table 4 included all teachers and controlled for a dummy variable identifying those for whom data on college attended were not available.

[24] Throughout our analysis, we generally obtain qualitatively similar results if we replace school fixed effects with controls for school characteristics (results available upon request).

informative, as they do not includes any of the teacher characteristics presented in Table 4. In column 3, that strong assumption is relaxed somewhat by also including indicators of a teacher's years of experience. In column 4, results are presented from a model that controls for the teacher's college major—whether or not they are an education major as well as the other distinctions presented in Table 4. And in column 5 estimations include controls for teacher race/ethnicity and gender. Across all these models, the results remain essentially the same as those reported in the models whose results are reported in the second column. In almost all cases, estimations not only show no significant effects but the coefficients are also precisely estimated. Only a handful of significant effects are observed, and they run contrary to the selectivity hypothesis. One can thus rule out with a good degree of confidence any positive relationship between attendance at a selective Florida university and student performance on the FCAT.

Individual University Programs

Lumping together Florida's eleven different universities into three clusters based on selectivity may obscure a more specific connection between university training and classroom effectiveness. To explore this possibility, the relationships between teacher effectiveness and training at 11 specific universities in Florida is estimated. Results for both reading and math are reported in Table 7A for students in elementary grades and in Table 7B for students in the middle school grades. Estimations from the five models discussed previously are reported, and indicate the average difference in effectiveness between the listed university and the University of Florida. Note that, if one were to rely strictly on the naïve model presented in column 1, which does not take into account student characteristics and school fixed effects, one would conclude that students learn more from teachers who had attended Florida's flagship university, the University of Florida, the most selective of all public institutions of higher education in the state. Thirty-six out of 40 coefficients in the four columns headed by the label (1) in Tables 6A and 6B are negative, and 22 of those 36 are statistically significant. Thus the University of Florida appears to be providing the best teacher education in the state of Florida only if one ignores the context in which teachers subsequently find themselves.

But when adjustments are made for student characteristics and school fixed effects, as is done in the results from the models presented in columns headed by (2) in these same tables, the picture changes dramatically. Most of the statistically significant relationships vanish, and the number of coefficients that show lower effectiveness at other universities drops to nine, about one quarter of the 40 estimations. We are unable to reject the null hypothesis that all of the coefficients are equal to zero, with p-values of this F test in the 0.19–0.89 range. In other words, once student and school characteristics are taken into account, the University of Florida advantage disappears.

Nor does it regain a decisive edge when the other three models are estimated, as presented in columns headed 3, 4, and 5. True, teachers from Florida A&M appears to be less effective in middle school math than teachers from the University of Florida. But no other relationship passes the significance test, with the exception of Florida International University, which appears to produce modestly better elementary school teachers than the University of Florida (although the estimates are not statistically significant in all models). Because the numbers of observations are smaller when the effects of individual institutions are estimated, results are not as precisely estimated as when grouping institutions by their degree of selectivity. Yet the pattern of results is so consistent across models and across institutions that one can be reasonably confident that no institution of higher education in Florida produces teachers that systematically secure higher levels of student achievement on the FCAT. The potential exception—Florida International University—hardly gives comfort to those who prize attendance at a selective university, as Florida International is one of the state's less selective institutions.

Still, it is possible that some teachers are excellent instructors in ways not captured by the FCAT, the state accountability test which holds schools accountable. If some teachers are "teaching to the test" in ways counter-productive to learning math and reading more generally then our results could be misleading. To check to see whether results change when a norm referenced test not used for accountability purposes is used to measure teacher effectiveness, we estimated effects using model 5 on the Stanford Achievement Test, which was also administered to all Florida students in these grades until 2008. No consistent pattern of statistically significant relationships emerged.[25]

Admittedly, these results do not isolate the effectiveness of universities' teacher preparation programs, in part because they include students who did not receive a major in education (though models 4 and 5 do control for that). To check the possibility that differences among universities might appear if teachers with education majors are isolated, we also estimated model 5 using data for only those graduates who had a degree in education. Once again, almost all relationships were statistically insignificant.[26]

---

[25] These results are available from the authors upon request. It is worth noting that the apparent superiority of Florida International University, as evidenced by FCAT scores, vanishes when one examines low-stakes test scores.

[26] These results are also available from the authors upon request. They indicate that some programs appear to be relatively weak at turning out students in particular subjects for specific grade levels—the University of South Florida in math instruction of those in the younger grades, or the University of West Florida in reading instruction in the older grades. But whenever 40 relationships are estimated, two will appear to be statistically significant at the five percent level by chance alone, so we cannot rule out that these findings could have occurred by chance, especially since the findings are isolated to a specific subject in a particular set of grades.

We presented evidence above that teachers with masters' degrees are no more effective in the classroom than those without an advanced degree, and teachers do not appear to become more effective upon earning a master's degree. However, it could be that certain master's degree programs attract better teachers (and thus are a useful signal of teacher quality) or provide better training to teachers. The tuition cost of earning a master's degree varies widely across public universities in Florida, from $4,210 at the University of West Florida to $12,060 at the University of Florida (Table 3).[27] We examine this question in Table 8, using data on teachers that received their master's degree from a public university in Florida in 1995 or later. The first four columns show results that compare the effectiveness of teachers with MA degrees from various programs to teachers with an MA degree from the University of Florida. No single MA program stands out in more than a single grade-subject combination, with the exception of the University of Central Florida. (The large number of significant coefficients, all of roughly equal magnitude, in middle school math simply indicates that teachers with masters' degrees from the University of Florida were modestly less effective in this subject-grade combination than teachers with masters' degrees from the other nine universities.)

The final four columns of Table 10 condition on teacher fixed effects, and thus indicate to what extent the effectiveness of individual teachers improved (or declined) after receipt of a master's degree from a given institution. We do not find more than one statistically significant positive coefficient for any individual institution, and there are just as many significant negative coefficients are there are significant positive coefficients.

---

[27] Tuition data are from university websites, accessed June 2010.

Interpretations and Conclusions

In general, we find that it is easier to pick a good teacher than to train teachers to make them more effective. NBPTS certification identifies more effective teachers, but its process of selecting them (which involves extensive self-examination of teaching strategies) appears not to have additional value. Masters' degrees appear to have little impact. More generally, we find little difference in the apparent effectiveness of attending a more selective university or, indeed, in having majored in any specific Florida university teacher training program.

We also find that the on-the-job training that teachers receive with each year of experience on the job to be fairly modest and that it may even turn downward at some point later in their careers.

If it is easier to identify good teaching than to set up a training program that systematically enhances teacher quality, public policy is organized on the opposite principle. Districts are expected to recruit teachers who hold a state teaching license or to ask them to obtain such a license within a short time after beginning teaching. Teachers are compensated according to whether or not they have an advanced degree and according to the number of years of experience in the classroom they have had. In Florida, a teacher with 5 years of experience who also holds a M.A. degree receives, in the average school district, 7 percent more each year than the teacher who has a B.A. degree (for specific numbers, see Table 1).

Districts also compensate teachers for training that occurs on-the-job with each additional year of service, especially after teachers have been in service for ten years. Table 1 indicates that a teacher with 8 years of experience earns 10 percent more than a new teacher (less than one percent uncompounded growth per year), while a teacher with 20 years of experience earns 26 percent more than a teacher with 8 years (better than 2 percent uncompounded growth annually).

A revisit of the design of teacher recruitment and compensation policies seems to be in order, given what is currently know about the factors associated—and not associated—with effective teaching.

References

Ballou, Dale and Michael Podgursky. 1997. *Teacher Pay and Teacher Quality*. Kalamazoo, Michigan: W. E. Upjohn Institute for Employment Research.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, Jonah Rockoff, and James Wyckoff. 2008. "The Narrowing Gap in New York City Teacher Qualifications and its Implications for Student Achievement in High-Poverty Schools." *Journal of Policy Analysis and Management* 27(4): 793–818.

Cantrell, Stephen, Jon Fullerton, Kane, Thomas J. and Douglas O. Staiger. 2008. "Estimating National Board Certification and Teacher Effectiveness: Evidence from a Random Assignment Experiment." NBER Working Paper No. 14608. Cambridge, Mass.: National Bureau for Economic Research.

Chingos, Matthew M. 2010. "The Impact of a Universal Class-Size Reduction Policy: Evidence from Florida's Statewide Mandate." Program on Education Policy and Governance Working Paper 10-03.

Chingos, Matthew M. and Martin R. West. 2010. "Do More Effective Teachers Earn More Outside of the Classroom?" Program on Education Policy and Governance Working Paper 10-02.

Center for Education Policy Research. 2010. "Teacher Employment Patterns and Student Results in Charlotte-Mecklenberg Schools." Strategic Data Project, Harvard Graduate School of Education.

Coleman, James, et al. 1966. *Equality of Educational Opportunity*. Washington, D. C.: Government Printing Office.

Clotfelter, Charles T, Helen F. Ladd, and Jacob L. Vigdor. 2006. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." Cambridge MA: National Bureau of Economic Research. Working Paper No. 11936.

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor. 2007a. "How and Why Do Teacher Credetnials Matter for Student Achievement?" Working Paper No. 2. Washington, D. C.: CALDER.

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor. 2007b. "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects." Working Paper No. 11. Washington, D. C.: CALDER.

Ferguson, Ronald F. and Helen F. Ladd. 1996. "How and Why Money Matters: An Analysis of Alabama Schools." In Helen F. Ladd, ed., *Holding Schools Accountable: Performance-Based Education Reform*. Brookings Institution Press, pp. 265–98.

Ehrenberg, Ronald G. and Dominic J. Brewer. 1994. “Do School and Teacher Characteristics Matter? Evidence from high School and Beyond.” *Economics of Education Review* 13(1): 1–17.

Figlio, David N. 1997. “Teacher Salaries and Teacher Quality,” Economics Letters 55(2): 267–271.

Goldhaber, Dan. 2002. “The Mystery of Good Teaching.” *Education Next* 2 (2): 50–55.

Goldhaber, Dan, and Emily Anthony. 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.

Hanushek, Eric. 1971. “Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data.” *American Economic Review* 61(2): 208–88.

Hanushek, Eric A. and Steven G. Rivkin. 2003. “How to Improve the Supply of High Quality Teachers.” Paper prepared for the Brookings Papers on Education Policy. Washington, D. C.

Hanushek, Eric A. and Steven G. Rivkin. 2010. “Generalizations about Using Value-Added Measures of Teacher Quality.” *American Economic Review* 100(2): 267–271.

Harris, Douglas N. and Tim R. Sass. 2007. “What Makes for a Good Teacher and Who Can Tell?” Department of Economics, Florida State University.

Harris, Douglas N. and Tim R. Sass. 2008. “Teacher Training, Teacher Quality and Student Achievement.” Department of Economics, Florida State University.

Hess, Frederick M., Andrew J. Rotherham, and Kate Walsh. 2004. *A Qualified Teacher in Every Classroom: Appraising Old Answers and New Ideas*. Cambridge, Mass: Harvard Education Press.

Hoxby, Caroline M. and Andrew Leigh. 2004. “Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in the United States.” *American Economic Review* 94(2): 236–240.

Ingersoll, Richard M. 1999. “The Problem of Underqualified Teachers in American Secondary Schools.” *Educational Researcher*, 28(2): 26–37.

Jacob, Brian A., and Lars Lefgren. 2008. “Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education.” *Journal of Labor Economics*, 26(1): 101–136.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Stagier. 2008. “What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City.” *Economics of Education Review* 27(6): 615–631.

Kane, Thomas J. and Douglas O. Staiger. 2008. “Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation.” NBER Working Paper No. 14607. Cambridge, Mass.: National Bureau for Economic Research.

Ladd, Helen F., Tim R. Sass, and Douglas N. Harris, “The Impact of National Board Certified Teachrs on Student Achievement in Florida and North Carolina.” Evidence prepared for the National Academies Committee on the Evaluation of the Impact of Teacher Certification by the National Board for Professional Teaching Standards, 2007.

Llaudet, Elena. 2006. “Eight Years of Reform.” In Paul E. Peterson, ed. *Reforming Education in Florida.* Hoover Institution Press, pp. 25–48.

Moe, Terry, 2006. “Quality Teachers.” In Paul E. Peterson, ed. *Reforming Education in Florida*. Hoover Institution Press, pp. 135–148.

Murnane, Richard J., Judith D. Singer, John B. Willett, James J. Kemple, and Randall J. Olsen. 1991. *Who Will Teach? Policies that Matter.* Cambridge, MA: Harvard University Press.

Murnane, Richard J. and Jennifer L. Steele. 2007. “What is the Problem? The Challenge of Providing Effective Teachers for All Children.” In *The Future of Children: Excellence in the classroom*, 17(1): 15–43.

Nadler, Daniel and Paul E. Peterson, 2009. “What Happens when States have Genuine Alternative Certification?” *Education Next* 9 (Winter), 70–74.

Nye, Barbara, Spyros Konstantopoulos, and Larry V. Hedges. 2004. “How Large are Teacher Effects?” *Educational Evaluation and Policy Analysis* 26(3): 237–57.

Papay, John P. and Matthew A. Kraft. 2010. “Do Teachers Continue to Improve with Experience? Evidence of Long-Term Career Growth in the Teacher Labor Market.” Harvard University Graduate School of Education.

Peterson’s Guide. 2009. *Four-Year Colleges – 2010.* Lawrenceville, NJ: Peterson’s.

Rockoff, Jonah. 2004. “The Impact of Indivudal Teachers on Student Achievement: Evidence from Panel Data.” *American Economic Review* 94(2): 247–252.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. “Teachers, Schools, and Academic Achievement.” *Econometrica* 73: 417–458.

Rothstein, Jesse. 2010. “Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement.” *Quarterly Journal of Economics* 94(2): 247–252.

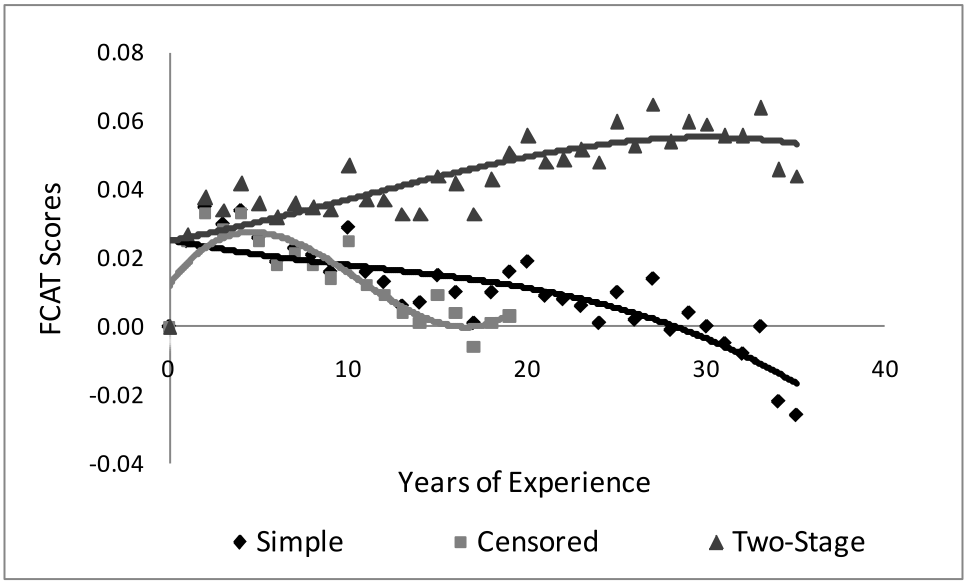
State of Florida, Department of Education. 2009. Teacher Effectiveness in Reading and Mathematics 2008-09. Report.

Summers, Anita and Barbara Wolfe. 1977. “Do Schools Make a Difference?” *American Economic Review* 67(4): 639–52.

U.S. Department of Education, Institute of Education, *Digest of Education Statistics*, 2009.

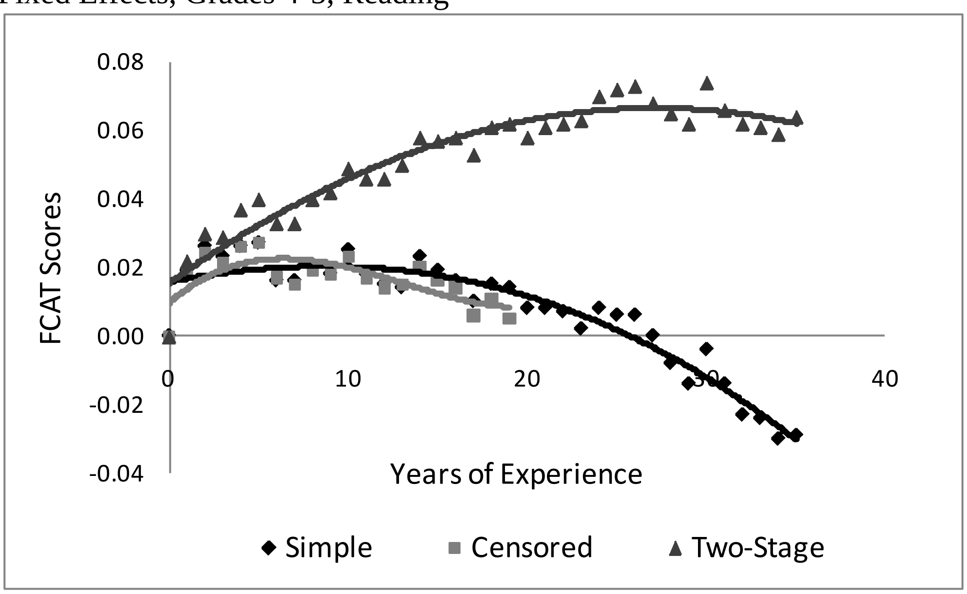
U.S. News & World Report. 2009. *Best Colleges 2010*.

Figure 1a. Teacher Effectiveness by Experience as measured by FCAT, with Teacher*School Fixed Effects, Grades 4-5, Math



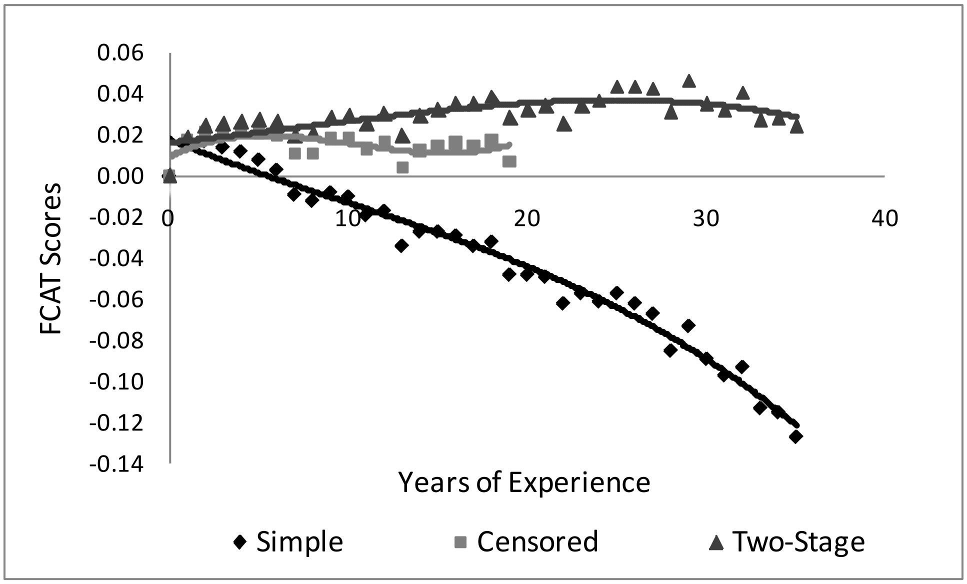
Notes: Regression lines are polynomial functions with squared and cubed terms. Data cover period from 2002 to 2009.

Figure 1b. Teacher Effectiveness by Experience as measured by FCAT, with Teacher*School Fixed Effects, Grades 4-5, Reading



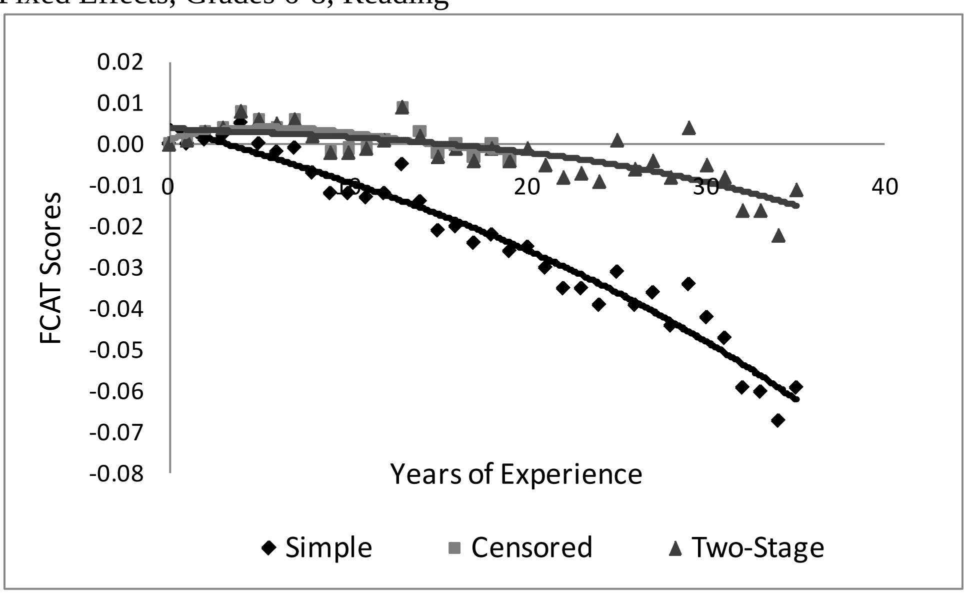
Notes: Regression lines are polynomial functions with squared and cubed terms. Data cover period from 2002 to 2009.

Figure 2a. Teacher Effectiveness by Experience as measured by FCAT, with Teacher*School Fixed Effects, Grades 6-8, Math



Notes: Regression lines are polynomial functions with squared and cubed terms. Data cover period from 2002 to 2009.

Figure 2b. Teacher Effectiveness by Experience as measured by FCAT, with Teacher*School Fixed Effects, Grades 6-8, Reading



Notes: Regression lines are polynomial functions with squared and cubed terms. Data cover period from 2002 to 2009.

Table 1. 2009-2010 Base Teacher Salary Schedule, Average of 47 Florida Counties

Experience	Bachelor's	Master's
0	$35,375	$37,771
1	$35,644	$38,060
2	$35,951	$38,402
3	$36,378	$38,858
4	$36,802	$39,305
5	$37,322	$39,844
6	$37,852	$40,385
7	$38,401	$40,956
8	$39,022	$41,587
9	$39,679	$42,264
10	$40,353	$42,961
11	$41,019	$43,659
12	$41,718	$44,375
13	$42,432	$45,113
14	$43,269	$45,041
15	$44,205	$46,926
16	$45,215	$47,954
17	$45,857	$48,828
18	$47,138	$49,948
19	$48,145	$50,968
20	$49,354	$52,219
Increase 0-8	10%	10%
Increase 8-22	26%	26%

Source: Various school district websites, accessed June 2010.

Table 2A. Descriptive Statistics, Teachers in Grades 4-5

	All Teachers	Teachers Matched to College			
		All	Less Selective	Selective	Very Selective
Teacher Characteristics					
BA in Elementary Education	-	80%	84%	82%	70%
BA in Other Education Field	-	4%	3%	5%	4%
BA in Non-Education Field	-	16%	13%	13%	26%
Failed certification exam at least once	2%	0%	1%	0%	0%
Certification exam score unknown	54%	36%	32%	35%	44%
National Board Certified	7%	6%	6%	6%	7%
Average Experience	8.9	2.6	2.6	2.8	2.5
Master's Degree or Higher	38%	27%	26%	21%	38%
Age	40.3	31.9	32.9	32.2	29.8
Male	13%	12%	13%	12%	11%
Black	15%	14%	23%	7%	11%
Hispanic	10%	13%	23%	7%	5%
Students' Characteristics					
Prior-year FCAT Math Scores	0.02	0.01	-0.03	0.02	0.04
Prior-year FCAT Reading Scores	0.03	0.01	-0.06	0.04	0.09
Share Black	25%	27%	33%	20%	28%
Share Hispanic	23%	24%	31%	22%	15%
Share Eligible for Free/Reduced Lunch	55%	56%	61%	54%	52%
Share Special Ed	15%	15%	14%	17%	16%
Share Limited English Proficient	7%	7%	8%	8%	5%
Class Size	20.7	20.4	20.8	20.3	20.0
Number of Teachers	37,774	9,140	3,393	3,654	2,093

Note: Data cover period from 2002 to 2009.

Table 2B. Descriptive Statistics, Math and Reading Teachers in Grades 6-8

	All Teachers	Teachers Matched to College			
		All	Less Selective	Selective	Very Selective
Teacher Characteristics					
BA in Math or English Education	-	8%	4%	11%	9%
BA in Middle or Secondary Education	-	3%	3%	1%	4%
BA in Elementary Education	-	21%	27%	21%	14%
BA in Other Education Field	-	28%	29%	34%	20%
BA in Math or English (Not Education)	-	16%	18%	9%	22%
BA in Other Non-Education Field	-	51%	51%	44%	60%
Failed certification exam at least once	2%	1%	2%	0%	0%
Certification exam score unknown	57%	44%	42%	40%	50%
National Board Certified	5%	4%	4%	4%	4%
Average Experience	8.6	2.6	2.5	2.9	2.3
Master's Degree or Higher	40%	27%	29%	23%	30%
Age	42.1	32.7	33.3	34.0	30.4
Male	23%	21%	21%	20%	21%
Black	20%	22%	34%	14%	19%
Hispanic	8%	10%	21%	6%	4%
Students' Characteristics					
Prior-year FCAT Math Scores	-0.27	-0.27	-0.33	-0.31	-0.17
Prior-year FCAT Reading Scores	-0.25	-0.25	-0.34	-0.29	-0.12
Share Black	28%	29%	35%	24%	30%
Share Hispanic	21%	22%	30%	20%	14%
Share Eligible for Free/Reduced Lunch	55%	56%	61%	55%	51%
Share Special Ed	26%	26%	24%	31%	21%
Share Limited English Proficient	6%	6%	7%	7%	4%
Class Size	11.5	11.7	12.0	10.6	12.7
Number of Teachers	46,257	9,533	3,110	3,544	2,879

Note: Data cover period from 2002 to 2009.

Table 3. Characteristics of Florida Institutions of Higher Learning

Tier and Institution Name	National Ranking	Entrance Difficulty	Avg. SAT Score	Avg. HS GPA	BA In-State Tuition (Yearly)	MA In-State Tuition (Total)	# Education BAs, 08-09	Percent Black	Percent Hispanic
Very Selective (Tier 1)									
University of Florida	47	Very difficult	1270	4.1	$4,373	$12,060	258	10%	14%
Florida State University	102	Very difficult	1200	3.7	$2,933	$10,488	515	11%	12%
University of North Florida	49	Very difficult	1115	3.5	$4,193	$11,689	328	10%	7%
Selective (Tier 3)									
Florida Gulf Coast University	Not ranked	Moderately difficult	1030	3.3	$4,437	$12,078	168	4%	12%
University of Central Florida	Not ranked	Moderately difficult	1180	3.7	$4,526	$9,978	1,002	9%	14%
University of South Florida	Not ranked	Moderately difficult	1140	3.7	$4,503	$9,954	498	12%	14%
Less Selective (Tier 4 and Non-competitive)									
Florida A&M University	Not ranked	Moderately difficult	910	2.9	$3,047	NA	92	94%	1%
Florida Atlantic University	Not ranked	Moderately difficult	920	3.4	$3,367	$10,582	494	18%	19%
Florida International University	Not ranked	Moderately difficult	1150	3.7	$4,580	$8,008	382	12%	64%
St. Petersburg College	Not ranked	Non-competitive	NA	NA	$2,596	NA	185	10%	6%
University of West Florida	Not ranked	Moderately difficult	1060	3.5	$3,649	$4,210	265	10%	5%

Notes: Tier and national ranking are from U.S. News and World Report (2009). Entrance difficulty is from Peterson's Guide (2009). Number and percent of Education BAs and percent black and Hispanic are from IPEDS. Averaged SAT score is calculated as the mean of the 25th and 75th percentiles reported by the Princeton Review. At two of the universities, more students take the ACT than the SAT, but using the average ACT score (converted to the SAT scale) would not change the average score by more than 10 points. MA tuition is from various university websites (accessed June 201). If MA tuition is different for elementary and secondary education, the average of two is listed. MA tuition data were not available for Florida A&M, and St. Petersburg College does not grant MA degrees.

Table 4. Relationship Between Teacher Characteristics and Student FCAT Scores

	Grades 4-5		Grades 6-8	
	Math	Reading	Math	Reading
BA College (relative to Less Selective)				
Selective	-0.017**	-0.006	0.009	-0.003
	[0.005]	[0.004]	[0.005]	[0.004]
Very Selective	-0.008	-0.008	0.011	-0.001
	[0.006]	[0.004]	[0.006]	[0.004]
BA Major (relative to Elementary Education [4-5] or Math/English Education [6-8])				
Other Education	-0.029*	-0.034**	0.000	-0.007
	[0.012]	[0.010]	[0.009]	[0.008]
Field other than Education	-0.008	-0.003	0.007	0.004
	[0.006]	[0.004]	[0.006]	[0.004]
Middle or Secondary Education			0.021	-0.019*
			[0.012]	[0.008]
Elementary Education			0.015*	0.012
			[0.007]	[0.008]
Math/English (Not Education)			0.003	-0.006
			[0.007]	[0.004]
Failed certification exam one or more	-0.036**	-0.001	-0.028**	-0.030**
times (relative to never failing)	[0.013]	[0.011]	[0.008]	[0.008]
National Board Certified	0.026**	0.016**	0.028**	0.019**
	[0.004]	[0.003]	[0.004]	[0.003]
Total experience (relative to 0 years)				
1-2 years experience	0.034**	0.025**	0.023**	0.005*
	[0.003]	[0.002]	[0.002]	[0.002]
3-5 years experience	0.043**	0.037**	0.028**	0.008**
	[0.003]	[0.002]	[0.003]	[0.002]
6-12 years experience	0.048**	0.042**	0.033**	0.012**
	[0.003]	[0.002]	[0.003]	[0.002]
13-20 years experience	0.062**	0.055**	0.043**	0.018**
	[0.004]	[0.003]	[0.003]	[0.003]
21+ years experience	0.062**	0.064**	0.039**	0.021**
	[0.004]	[0.003]	[0.003]	[0.003]
Highest degree (relative to BA)				
Master's	0.002	0.001	0.004	0.003*
	[0.002]	[0.002]	[0.002]	[0.002]
Doctorate	-0.013	0.000	-0.003	0.001
	[0.010]	[0.008]	[0.007]	[0.006]
Observations (Student*Year)	1,822,921	1,825,076	2,963,955	2,127,438
Observations (Unique Teachers)	36,378	36,382	14,972	18,898
R-squared	0.67	0.64	0.71	0.65

Notes: * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in brackets. All models include school fixed effects, a dummy variable identifying teachers for whom university attended and major are not available, and a dummy variable identifying teachers for whom certification exam data are not available. Student-level controls include prior-year test scores (in both subjects, along with squared and cubed terms), gender, race/ethnicity, free lunch, LEP, special education, migrant, non-structural move, structural move, days absent previous year, and whether repeating the grade. Classroom-level controls include these variables aggregated to the classroom level (with the exception that prior-year test scores are replaced with percent proficient on the FCAT in reading and math) and class size. Data cover period from 2002 to 2009.

Table 5. Relationship Between National Board Certification, Master's Degree Receipt, and Student FCAT Scores, With Teacher*School Fixed Effects

	Grades 4-5		Grades 6-8	
	Math	Reading	Math	Reading
National Board Certified	0.018*	0.001	0.011	0.010
	[0.009]	[0.008]	[0.008]	[0.009]
Observations (Student*Year)	81,253	81,351	101,396	97,956
Observations (Unique Teachers)	940	940	260	374
R-squared	0.70	0.65	0.73	0.63

	Grades 4-5		Grades 6-8	
	Math	Reading	Math	Reading
Has Master's Degree	0.002	0.008	-0.006	-0.005
	[0.012]	[0.011]	[0.011]	[0.010]
Observations (Student*Year)	39,996	40,026	77,346	55,437
Observations (Unique Teachers)	561	561	252	281
R-squared	0.69	0.65	0.71	0.65

Notes: * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in brackets. All regressions include teacher*school fixed effects and grade-by-year fixed effects. Student-level controls include prior-year test scores (in both subjects, along with squared and cubed terms), gender, race/ethnicity, free lunch, LEP, special education, migrant, non-structural move, structural move, days absent previous year, and whether repeating the grade. Classroom-level controls include these variables aggregated to the classroom level (with the exception that prior-year test scores are replaced with percent proficient on the FCAT in reading and math) and class size. Data cover period from 2002 to 2009.

Table 6A. Relationship Between Selectivity of Teacher's BA Institution and Student FCAT Scores, Grades 4-5

	Math				
	(1)	(2)	(3)	(4)	(5)
BA College (relative to Less Selective)					
Selective	0.013	-0.019**	-0.016*	-0.015*	-0.012
	[0.011]	[0.007]	[0.007]	[0.007]	[0.007]
Very Selective	0.032*	-0.010	-0.005	-0.002	0.001
	[0.014]	[0.007]	[0.007]	[0.007]	[0.007]
Include student, classroom, and school controls?	No	Yes	Yes	Yes	Yes
Control for Teacher Experience?	No	No	Yes	Yes	Yes
Control for Teacher BA Major?	No	No	No	Yes	Yes
Control for Teacher Race and Gender?	No	No	No	No	Yes
Observations (Student*Year)	442,774	403,804	392,748	392,748	392,728
Observations (Unique Teachers)	9,134	8,988	8,888	8,888	8,887
R-squared	0.00	0.66	0.66	0.66	0.66

	Reading				
	(1)	(2)	(3)	(4)	(5)
BA College (relative to Less Selective)					
Selective	0.074**	-0.012*	-0.011*	-0.010*	-0.009
	[0.011]	[0.005]	[0.005]	[0.005]	[0.005]
Very Selective	0.111**	-0.011*	-0.006	-0.006	-0.004
	[0.013]	[0.005]	[0.005]	[0.005]	[0.005]
Include student, classroom, and school controls?	No	Yes	Yes	Yes	Yes
Control for Teacher Experience?	No	No	Yes	Yes	Yes
Control for Teacher BA Major?	No	No	No	Yes	Yes
Control for Teacher Race and Gender?	No	No	No	No	Yes
Observations (Student*Year)	443,441	404,362	393,276	393,276	393,256
Observations (Unique Teachers)	9,136	8,990	8,890	8,890	8,889
R-squared	0.00	0.63	0.63	0.63	0.63

Notes: * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in brackets. All regressions include grade-by-year fixed effects. Student-level controls include prior-year test scores (in both subjects, along with squared and cubed terms), gender, race/ethnicity, free lunch, LEP, special education, migrant, non-structural move, structural move, days absent previous year, and whether repeating the grade. Classroom-level controls include these variables aggregated to the classroom level (with the exception that prior-year test scores are replaced with percent proficient on the FCAT in reading and math) and class size. School controls are school fixed effects. Data cover period from 2002 to 2009.

Table 6B. Relationship Between Selectivity of Teacher's BA Institution and Student FCAT Scores, Grades 6-8

	Math				
	(1)	(2)	(3)	(4)	(5)
BA College (relative to Less Selective)					
Selective	0.066**	0.004	0.006	0.008	0.004
	[0.023]	[0.007]	[0.007]	[0.007]	[0.007]
Very Selective	0.102**	0.008	0.010	0.012	0.008
	[0.025]	[0.006]	[0.006]	[0.006]	[0.006]
Include student, classroom, and school controls?	No	Yes	Yes	Yes	Yes
Control for Teacher Experience?	No	No	Yes	Yes	Yes
Control for Teacher BA Major?	No	No	No	Yes	Yes
Control for Teacher Race and Gender?	No	No	No	No	Yes
Observations (Student*Year)	553,800	498,477	482,477	482,477	482,477
Observations (Unique Teachers)	3,918	3,056	3,002	3,002	3,002
R-squared	0.00	0.70	0.70	0.70	0.70

	Reading				
	(1)	(2)	(3)	(4)	(5)
BA College (relative to Less Selective)					
Selective	0.071**	-0.002	-0.003	-0.004	-0.005
	[0.021]	[0.005]	[0.005]	[0.005]	[0.005]
Very Selective	0.093**	-0.000	-0.001	0.001	-0.000
	[0.022]	[0.005]	[0.005]	[0.005]	[0.005]
Include student, classroom, and school controls?	No	Yes	Yes	Yes	Yes
Control for Teacher Experience?	No	No	Yes	Yes	Yes
Control for Teacher BA Major?	No	No	No	Yes	Yes
Control for Teacher Race and Gender?	No	No	No	No	Yes
Observations (Student*Year)	501,119	450,971	435,925	435,925	435,925
Observations (Unique Teachers)	5,573	4,237	4,163	4,163	4,163
R-squared	0.01	0.64	0.64	0.64	0.64

Notes: * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in brackets. All regressions include grade-by-year fixed effects. Student-level controls include prior-year test scores (in both subjects, along with squared and cubed terms), gender, race/ethnicity, free lunch, LEP, special education, migrant, non-structural move, structural move, days absent previous year, and whether repeating the grade. Classroom-level controls include these variables aggregated to the classroom level (with the exception that prior-year test scores are replaced with percent proficient on the FCAT in reading and math) and class size. School controls are school fixed effects. Data cover period from 2002 to 2009.

Table 7A. Relationship Between Teacher's BA Institution and Student FCAT Scores, Grades 4-5

	Math					Reading				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
BA College (relative to University of Florida)										
Very Selective										
Florida State University	-0.018	0.006	0.002	0.004	0.003	-0.009	0.004	0.006	0.007	0.007
	[0.024]	[0.010]	[0.011]	[0.011]	[0.011]	[0.022]	[0.008]	[0.008]	[0.008]	[0.008]
University of North Florida	-0.115**	0.012	0.008	0.006	0.004	-0.025	0.017	0.015	0.015	0.015
	[0.025]	[0.013]	[0.013]	[0.013]	[0.013]	[0.023]	[0.010]	[0.010]	[0.010]	[0.011]
Selective										
Florida Gulf Coast University	-0.105**	-0.022	-0.015	-0.017	-0.020	-0.074*	0.004	0.012	0.012	0.011
	[0.031]	[0.018]	[0.018]	[0.018]	[0.018]	[0.029]	[0.013]	[0.013]	[0.013]	[0.013]
University of Central Florida	-0.017	-0.006	-0.007	-0.008	-0.009	-0.005	-0.001	-0.001	-0.001	-0.001
	[0.020]	[0.009]	[0.009]	[0.009]	[0.009]	[0.018]	[0.007]	[0.007]	[0.007]	[0.007]
University of South Florida	-0.095**	-0.003	-0.009	-0.010	-0.012	-0.085**	0.005	0.001	0.001	0.001
	[0.020]	[0.009]	[0.009]	[0.009]	[0.009]	[0.019]	[0.007]	[0.007]	[0.007]	[0.007]
Less Selective										
Florida A&M University	-0.288**	0.014	0.002	-0.001	-0.011	-0.286**	0.020*	0.014	0.014	0.010
	[0.027]	[0.013]	[0.013]	[0.013]	[0.014]	[0.025]	[0.010]	[0.010]	[0.010]	[0.010]
Florida Atlantic University	0.006	0.012	0.007	0.004	0.003	-0.079**	0.012	0.007	0.006	0.006
	[0.022]	[0.010]	[0.010]	[0.010]	[0.010]	[0.021]	[0.007]	[0.007]	[0.007]	[0.007]
Florida International University	-0.122**	0.035*	0.023	0.021	0.020	-0.186**	0.026*	0.021	0.021	0.019
	[0.024]	[0.014]	[0.014]	[0.014]	[0.014]	[0.023]	[0.010]	[0.011]	[0.011]	[0.011]
St. Petersburg College	-0.220**	0.002	0.025	0.022	0.019	-0.181**	0.010	0.028	0.029	0.028
	[0.045]	[0.020]	[0.020]	[0.019]	[0.020]	[0.041]	[0.016]	[0.016]	[0.016]	[0.016]
University of West Florida	-0.017	-0.003	-0.014	-0.015	-0.018	0.035	0.000	-0.003	-0.002	-0.003
	[0.026]	[0.017]	[0.016]	[0.017]	[0.016]	[0.024]	[0.012]	[0.012]	[0.012]	[0.012]
Joint significance of BA College (p-value)	0.000	0.193	0.276	0.275	0.209	0.000	0.204	0.267	0.294	0.410
Include student, classroom, and school controls?	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Control for Teacher Experience?	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Control for Teacher BA Major?	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Control for Teacher Race and Gender?	No	No	No	No	Yes	No	No	No	No	Yes
Observations (Student*Year)	442,774	403,804	392,748	392,748	392,728	443,441	404,362	393,276	393,276	393,256
Observations (Unique Teachers)	9,134	8,988	8,888	8,888	8,887	9,136	8,990	8,890	8,890	8,889
R-squared	0.01	0.66	0.66	0.66	0.66	0.01	0.63	0.63	0.63	0.63

Notes: * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in brackets. All regressions include grade-by-year fixed effects. Student-level controls include prior-year test scores (in both subjects, along with squared and cubed terms), gender, race/ethnicity, free lunch, LEP, special education, migrant, non-structural move, structural move, days absent previous year, and whether repeating the grade. Classroom-level controls include these variables aggregated to the classroom level (with the exception that prior-year test scores are replaced with percent proficient on the FCAT in reading and math) and class size. School controls are school fixed effects. Data cover period from 2002 to 2009.

Table 7B. Relationship Between Teacher's BA Institution and Student FCAT Scores, Grades 6-8

	Math (1)	Math (2)	Math (3)	Math (4)	Math (5)	Reading (1)	Reading (2)	Reading (3)	Reading (4)	Reading (5)
BA College (relative to University of Florida)										
Very Selective										
Florida State University	-0.002	0.007	0.005	0.006	0.006	-0.009	0.004	0.004	0.004	0.002
	[0.042]	[0.009]	[0.009]	[0.009]	[0.009]	[0.033]	[0.006]	[0.006]	[0.006]	[0.006]
University of North Florida	-0.112*	0.005	0.004	-0.008	-0.010	-0.096**	0.005	0.004	0.010	0.009
	[0.049]	[0.014]	[0.014]	[0.015]	[0.014]	[0.035]	[0.008]	[0.008]	[0.008]	[0.008]
Selective										
Florida Gulf Coast University	-0.139*	-0.008	-0.004	-0.019	-0.021	-0.199**	0.021	0.023	0.022	0.023
	[0.061]	[0.018]	[0.018]	[0.019]	[0.019]	[0.061]	[0.015]	[0.015]	[0.015]	[0.015]
University of Central Florida	-0.037	0.004	0.004	0.004	0.003	0.009	-0.002	-0.002	-0.005	-0.005
	[0.036]	[0.009]	[0.009]	[0.009]	[0.009]	[0.030]	[0.006]	[0.006]	[0.006]	[0.006]
University of South Florida	-0.072	-0.006	-0.008	-0.007	-0.007	-0.084**	0.001	-0.001	-0.003	-0.003
	[0.038]	[0.008]	[0.008]	[0.008]	[0.008]	[0.032]	[0.006]	[0.006]	[0.006]	[0.006]
Less Selective										
Florida A&M University	-0.294**	-0.024*	-0.029**	-0.031**	-0.024*	-0.311**	0.001	-0.001	-0.004	-0.005
	[0.044]	[0.011]	[0.011]	[0.011]	[0.012]	[0.047]	[0.008]	[0.008]	[0.009]	[0.009]
Florida Atlantic University	-0.025	0.000	-0.004	-0.007	-0.006	-0.054	0.003	0.004	0.002	0.002
	[0.042]	[0.010]	[0.010]	[0.010]	[0.010]	[0.036]	[0.007]	[0.008]	[0.008]	[0.008]
Florida International University	-0.194**	0.009	0.007	0.006	0.008	-0.106**	0.001	0.002	0.000	0.004
	[0.046]	[0.012]	[0.012]	[0.012]	[0.012]	[0.035]	[0.008]	[0.008]	[0.009]	[0.009]
St. Petersburg College	0.001	0.018	0.026	0.027	0.027	-0.343	0.021	0.024	0.020	0.020
	[0.123]	[0.025]	[0.025]	[0.027]	[0.027]	[0.212]	[0.028]	[0.029]	[0.030]	[0.030]
University of West Florida	-0.079	0.001	0.003	-0.010	-0.009	-0.091*	0.016	0.017	0.020	0.019
	[0.041]	[0.017]	[0.017]	[0.018]	[0.018]	[0.045]	[0.014]	[0.014]	[0.014]	[0.014]
Joint significance of BA College (p-value)	0.000	0.401	0.188	0.127	0.338	0.000	0.889	0.785	0.579	0.557
Include student, classroom, and school controls?	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Control for Teacher Experience?	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Control for Teacher BA Major?	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Control for Teacher Race and Gender?	No	No	No	No	Yes	No	No	No	No	Yes
Observations (Student*Year)	553,800	498,477	482,477	482,477	482,477	501,119	450,971	435,925	435,925	435,925
Observations (Unique Teachers)	3,918	3,056	3,002	3,002	3,002	5,573	4,237	4,163	4,163	4,163
R-squared	0.01	0.70	0.70	0.70	0.70	0.01	0.64	0.64	0.64	0.64

Notes: * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in brackets. All regressions include grade-by-year fixed effects. Student-level controls include prior-year test scores (in both subjects, along with squared and cubed terms), gender, race/ethnicity, free lunch, LEP, special education, migrant, non-structural move, structural move, days absent previous year, and whether repeating the grade. Classroom-level controls include these variables aggregated to the classroom level (with the exception that prior-year test scores are replaced with percent proficient on the FCAT in reading and math) and class size. School controls are school fixed effects. Data cover period from 2002 to 2009.

Table 8. Relationship Between Teacher's MA Institution and Student FCAT Scores, With and Without Teacher Fixed Effects

	Grades 4-5		Grades 6-8		Grades 4-5		Grades 6-8	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading
MA University								
Very Selective								
University of Florida					-0.010	0.004	0.097**	0.015
					[0.053]	[0.033]	[0.019]	[0.043]
Florida State University	-0.004	-0.008	0.047**	-0.024	0.010	0.002	0.022	0.024
	[0.022]	[0.017]	[0.017]	[0.014]	[0.038]	[0.033]	[0.026]	[0.014]
University of North Florida	0.012	-0.012	0.053	0.001	-0.000	0.013	0.021	0.001
	[0.025]	[0.020]	[0.029]	[0.018]	[0.029]	[0.023]	[0.033]	[0.024]
Selective								
Florida Gulf Coast University	-0.026	-0.038	0.041	0.026	-0.031	0.017	-0.007	0.064
	[0.032]	[0.025]	[0.028]	[0.017]	[0.033]	[0.041]	[0.047]	[0.042]
University of Central Florida	-0.017	-0.031	0.061**	0.031*	-0.026	0.001	0.019	0.016
	[0.021]	[0.016]	[0.020]	[0.013]	[0.027]	[0.016]	[0.032]	[0.018]
University of South Florida	0.001	0.017	0.067**	-0.002	0.031*	0.025	-0.026	0.005
	[0.019]	[0.015]	[0.017]	[0.011]	[0.016]	[0.013]	[0.015]	[0.014]
Less Selective								
Florida A&M University	-0.021	-0.019	0.035	0.033	0.012	0.031	0.037	-0.056*
	[0.032]	[0.023]	[0.027]	[0.021]	[0.046]	[0.058]	[0.032]	[0.028]
Florida Atlantic University	0.001	-0.023	0.066**	0.018	-0.014	-0.001	0.008	-0.067**
	[0.022]	[0.017]	[0.023]	[0.012]	[0.022]	[0.020]	[0.014]	[0.022]
Florida International University	0.022	0.008	0.052*	-0.013	-0.079**	-0.024	-0.013	0.021
	[0.023]	[0.018]	[0.023]	[0.014]	[0.025]	[0.025]	[0.028]	[0.017]
University of West Florida	-0.038	-0.012	0.044	0.030	0.018	0.070*	-0.049**	0.020
	[0.034]	[0.028]	[0.022]	[0.024]	[0.025]	[0.029]	[0.016]	[0.018]
Joint significance of MA University (p-value)	0.718	0.022	0.031	0.001	0.079	0.351	0.000	0.012
Teacher*School Fixed Effects?	No	No	No	No	Yes	Yes	Yes	Yes
Observations (Student*Year)	194,031	194,229	273,911	202,416	194,075	194,273	273,957	202,423
Observations (Unique Teachers)	3,806	3,806	1,359	1,717	3,807	3,807	1,360	1,717
R-squared	0.68	0.65	0.72	0.65	0.70	0.66	0.73	0.66

Notes: * significant at 5%; ** significant at 1%; robust standard errors adjusted for clustering at the teacher level appear in brackets. All regressions include grade-by-year fixed effects and the following controls: 1) Student-level controls include prior-year test scores (in both subjects, along with squared and cubed terms), gender, race/ethnicity, free lunch, LEP, special education, migrant, non-structural move, structural move, days absent previous year, and whether repeating the grade, 2) Classroom-level controls include these variables aggregated to the classroom level (with the exception that prior-year test scores are replaced with percent proficient on the FCAT in reading and math) and class size, 3) Teacher-level controls including experience, race/ethnicity and gender, and BA institution and major (including a dummy for those without BA institution/major data available). Time-invariant teacher-level controls are excluded from models that include teacher fixed effects. Models that do not include teacher*school fixed effects include school fixed effects. Data cover period from 2002 to 2009.