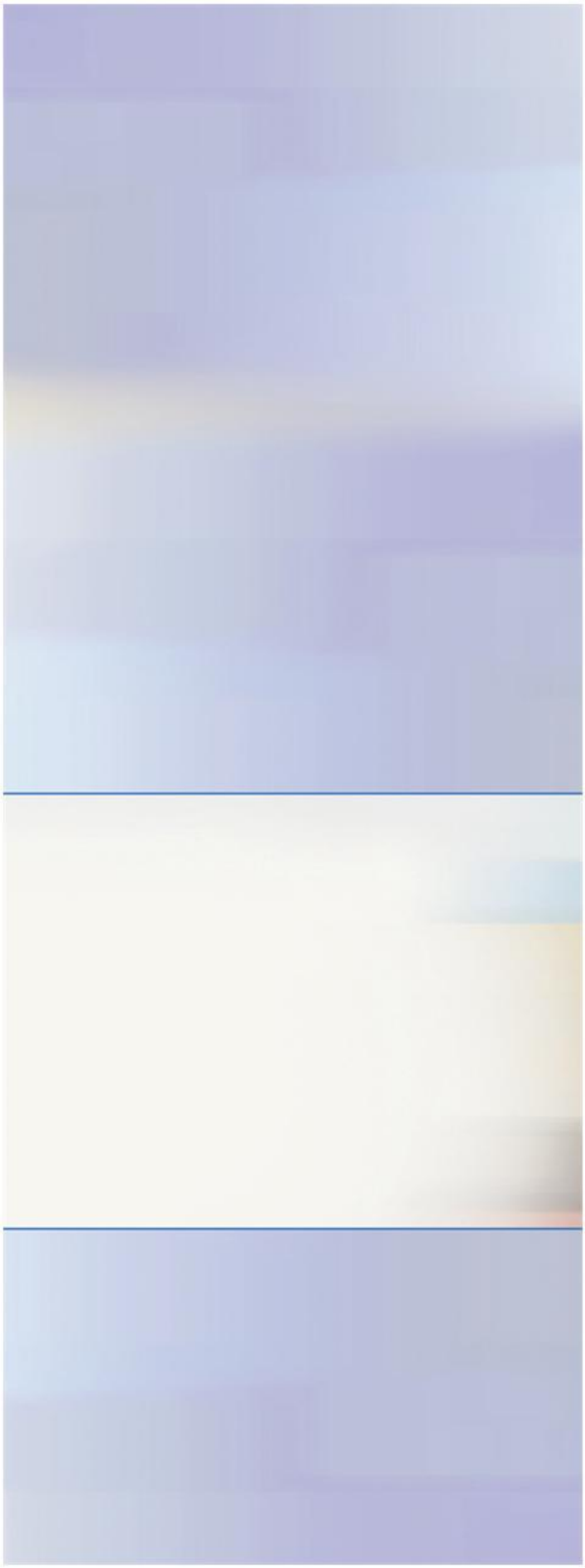
ABCTE Teachers in Florida and Their Effect on Student Performance

Final Report

September 4, 2009

Christina Clark Tuttle
Tara Anderson
Steven Glazerman

MATHEMATICA
Policy Research, Inc.



Mathematica Reference Number:
6215-400

ABCTE Teachers in Florida and Their Effect on Student Performance

Final Report

September 4, 2009

Christina Clark Tuttle
Tara Anderson
Steven Glazerman

Submitted to:
American Board for Certification of Teacher Excellence
1225 19th Street, NW, Suite 400
Washington, DC 20036
Project Officer: William Schimmel

Submitted by:
Mathematica Policy Research
600 Maryland Avenue, SW
Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763
Project Director: Steven Glazerman

MATHEMATICA
Policy Research, Inc.

CONTENTS

TABLES

Abstract

Do alternatively certified teachers produce student achievement gains on par with other teachers? In this study, we use propensity score matching to identify a comparison group of teachers in Florida to address this question for one alternative certification route, the American Board for Certification of Teacher Excellence (ABCTE). Recognized in nine states, ABCTE certification requires examinations but no specific course work or student teaching for initial teacher licensure. Using data on 30 ABCTE teachers in Florida over two years, we found no differences in student gains in reading between students of ABCTE and non–ABCTE teachers. However, students of ABCTE teachers scored lower than their counterparts on the state math test. The difference was equal to an effect size of 0.25 and was statistically significant, although the estimate varied slightly depending on how the matching was conducted and how the statistical model was specified. Limiting the sample to subgroups of teachers, such as novices or ABCTE teachers who scored high or low on certification exams (and their matched counterparts), did not change the overall conclusion either.

Introduction and Background

Introduction

In the United States today, alternative certification programs account for about one-third of all new teachers certified annually (Feistritzer 2007). These programs, including Troops to Teachers, Transition to Teaching, and the certification offered by the American Board for Certification of Teacher Excellence (ABCTE), represent a departure from the traditional path for entry into the teaching profession. Several factors have contributed to the growth of alternative teacher-certification programs, including efforts to reduce class size, difficulties in attracting high-quality minority candidates to the profession, and a shortage of individuals prepared to teach mathematics, science, and special education. Despite the large and growing role of alternative certification programs, much remains to be learned about the teachers certified through the programs and teachers' effects on the students in their classrooms.

In this report, we focus on ABCTE teachers. We address the question, **Do ABCTE-certified teachers produce student achievement gains on par with other teachers?** Using data from Florida, we compare the performance of students in classrooms taught by ABCTE teachers to that of students in matched classrooms.

To set a context for interpreting the findings of the study, we provide a brief history of the ABCTE program and related research in Section I. Section II describes the data and methods for the study, and Section III presents the major findings. The final section presents conclusions and limitations of the research.

History of ABCTE Certification

With its requirement that all teachers must be highly qualified in the subjects they teach, the No Child Left Behind (NCLB) Act of 2001 generated considerable interest in alternative certification in general and in subject-area certification in particular. As a result, the demand for streamlined certification also grew, including alternatives to the traditional systems (which may involve lengthy periods of coursework and student teaching). The increased demand provided opportunities for both out-of-field teachers and new teachers.

ABCTE was formed through a grant from the U.S. Department of Education (ED) to develop an alternative certification both for current teachers who needed to earn certification within a subject as well as for professionals seeking to change careers and become teachers. The goal was to provide an affordable, expeditious, high-quality route to certification that would be nationally recognized and portable. Unlike many traditional state-sanctioned teacher licensure programs and some alternative routes to certification, ABCTE does not require formal coursework, a portfolio, or student teaching as a condition of certification. Instead, certification centers on demonstrated mastery of subject matter and professional teaching knowledge as measured by performance on a set of examinations.

ABCTE began offering certification[1] in 2003 in Elementary Education (kindergarten through grade 6) and in English Language Arts (grade 6 through 12) and Mathematics (grade 6 through 12) one year later. To earn ABCTE certification, candidates were required to hold a bachelor's degree in any subject area, pass a federal background check, pass an ABCTE examination in the subject(s) they want to teach, and pass the ABCTE professional teaching knowledge examination at the time of our study. The tests were administered by computer at testing centers around the world, and for a fee of approximately $500, candidates had access to study materials and mentoring. The first enrollees in the program received teaching certificates in late 2004 and were eligible to teach in traditional public schools only in Idaho (as well as in charter and private schools elsewhere).

Currently, ABCTE offers certification in 9 areas. For teachers in kindergarten through grade 6, ABCTE offers certification in Elementary Education and Special Education. For grades 6 through 12, the following certifications are offered: English Language Arts, Mathematics, General Science, Biology, Chemistry, Physics, U.S. History, and World History.[2] ABCTE certification is recognized under the federal No Child Left Behind Act of 2001 as an approved way to demonstrate that a teacher is "highly qualified." As of the date of this report, it is an approved route to state licensure in nine states: Florida, Idaho, Mississippi, Missouri, New Hampshire, Pennsylvania, South Carolina, Utah, and Oklahoma. Each of these states recognizes one or more of the available certifications and may impose additional requirements on ABCTE certification holders who wish to teach there.

As of October 2007, the cutoff date for the most recent survey of ABCTE alumni, 609 individuals had successfully obtained ABCTE certification, of whom almost two-thirds were teaching in the United States in some capacity (Glazerman et al. 2008).

Related Research

This report continues a tradition of research comparing the classroom effectiveness of different types of teachers based on their qualifications, including certification route. The justification for creating alternative routes comes from empirical literature questioning the link between traditional certification and student achievement (Goldhaber and Brewer 2000; Ehrenberg and Brewer 1994) and suggesting that qualified applicants are discouraged from entering the profession (Ballou and Podgursky 1996).

Policy responses to the literature have ranged widely. They include interventions at several points along the route into the classroom--from recruitment and screening to training, placement, and, ultimately, licensure. For example, Troops to Teachers is a program that mainly provides placement services for uniformed service members seeking teaching positions. Teach For America is a program that recruits, screens, trains, and places recent college graduates from fields other than education into schools in very high-poverty areas where they commit to two years in the classroom and receive a summer training experience in lieu of the more traditional bachelor's degree from a school or college of education. Other programs labeled as certification offer more training and recruit candidates from different sources.

---

[1] Before 2009, this certification was termed the Passport to Teaching.

[2] A reading endorsement, although available to all certified teachers, satisfies the specific requirements of Florida and Idaho.

Research on these and related programs has compared student achievement growth for teachers from the different routes and demonstrated that teachers following alternative routes to the classroom have produced student achievement gains that were similar to those of traditionally prepared teachers (Boyd et al. 2008; Constantine et al. 2009) or slightly better (Glazerman et al. 2006; Kane et al. 2008).

The National Board for Professional Teaching Standards (NBPTS) is an organization that offers advanced certification for experienced teachers. The numerous, recent studies on National Board certification (Cavalluzzo 2004; Vandevoort et al. 2004; Goldhaber and Anthony 2007; National Research Council 2008; Cantrell et al. 2008) typically examine the student achievement gains for certified and non-certified teachers, including failed applicants and non-applicants to the program.

Past studies have varied in their approaches to the key challenge in this line of research, which is to isolate the true effect of the teacher certification route from other factors that affect student achievement. Some researchers have been able to randomly assign students to classrooms in order to produce equivalent groups of students that vary only in the teacher characteristics (Glazerman et al. 2006; Cantrell et al. 2008; Constantine et al. 2009). More typically, researchers have relied on statistical controls to identify impacts of teacher certification (e.g. Goldhaber and Anthony 2007; Kane et al. 2008; and Boyd et al 2008).

Unlike many of the other teacher credentials discussed above, ABCTE certification has not received much research attention. To date, the research on ABCTE has generally focused on the individuals who obtain certification rather than on the students they teach. Two earlier studies by Mathematica Policy Research fall in this category.

One of those studies reported on the career choices, teaching assignments, and careers of ABCTE certificate holders (Glazerman et al. 2008). The data came from a survey of all certificate holders at a point when they had been certified for one, two, or three years. It found that 59 percent of ABCTE certificate holders had obtained K-12 teaching positions, with 57 percent of those in teaching positions reporting having a mentor assigned to them. Of those respondents who ever taught since being certified by ABCTE, 86 percent were still teaching at the time of the Mathematica survey.

The other Mathematica study reported on the relationship between scores on the American Board examinations required for certification and the similar Praxis II examinations required by most states for certification (Chaplin et al. 2007). That study found that scores on ABCTE and Praxis II exams were correlated; the correlation coefficient was 0.73 for secondary mathematics, 0.73 for elementary subjects, and 0.30 for pedagogy. The report also compared pass rates on the two sets of exams and found the ABCTE elementary exams to be more difficult to pass in all states and the ABCTE math and pedagogy exams being more difficult to pass according to states with lower Praxis cut scores, but easier to pass in others.

Some suggestive evidence on ABCTE teachers' effectiveness was included in an early report on principals' attitudes toward the first cohort of ABCTE alumni who were teaching (Glazerman et al. 2006). Ninety percent of ABCTE-certified teachers were rated by their principals as average, above average, or substantially above average in terms of overall performance as compared to all teachers observed by those principals throughout their careers. Two limitations of the 2006 study, however, are (1) the subjective nature of performance ratings and (2) its focus on the first cohort of ABCTE

teachers, which may not be representative of the type of teacher who obtains the credential as the certification has become more popular and more widely recognized.[3]

To generate objective evidence on the classroom performance of ABCTE teachers, we obtained test score data from the Florida Department of Education and linked the student data to the ABCTE teachers responsible for those outcomes. The findings presented here represent an important first step in measuring the effectiveness of ABCTE teachers, motivating future research on a more ambitious scale. This may include larger-scale randomized trials of ABCTE certification, in which random assignment could be used to control both observable and unobservable determinants of student achievement that might otherwise be confounded with certification status.

Data and Methods

The study uses test score data provided by Florida school districts to the state Department of Education to measure the average achievement growth of students taught by ABCTE and comparable non-ABCTE teachers. We use propensity score matching methods, discussed below, to balance the ABCTE and comparison classrooms.

Florida was the third state to accept ABCTE certification (in June 2004). According to a recent survey, it accounts for the second-highest number of ABCTE alumni and exhibits the most rapid growth in ABCTE certification in the United States (Glazerman et al. 2008). The state also maintains one of the most sophisticated education data systems in the country, the Florida K-20 Education Data Warehouse (EDW), which contains longitudinally linked data on students, teachers, and schools. Given the state's data quality and ABCTE representation, we chose to focus on Florida for our analyses.

Although limiting the analyses to a single state potentially limits the external validity of the results, ABCTE teachers in Florida do not differ markedly from ABCTE teachers overall on most observable characteristics. In Table 1, we describe ABCTE teachers by using data collected by Mathematica in a survey of ABCTE alumni completed in 2008. (For more detail on the survey and its findings, see Glazerman et al. 2008.) We compare characteristics of respondents to the survey who are teaching kindergarten through grade 12 in the United States, limit that group to teachers in Florida, and then further limit the group to teachers in our analytic sample (those who taught grades and subjects covered by standardized tests). Some of the differences between Florida teachers and ABCTE teachers in general include a higher number of certifications in 2007 in Florida (since it was one of the first states to recognize the credential), a higher percentage in traditional public schools, and a higher percentage pursuing ABCTE certification to retain a current position rather than to enter the profession.

This difference was even greater when we examine the subset of teachers who were included in the analysis for this report. ABCTE teachers in our sample had lower scores on the ABCTE exam of professional teaching knowledge than ABCTE certificate holders nationally, but their scores on the subject matter exams were statistically indistinguishable from those of reported in the national survey data.

[3] For example, only 49 percent of the first cohort of ABCTE alumni obtained certification to enter the teaching profession as compared to 64 percent of the most recent cohort (Glazerman et al. 2008).

Table 1. Characteristics of ABCTE Teachers in Florida Versus the U.S. in 2007

Characteristics (percent unless otherwise indicated)	All ABCTE Teachers	Florida ABCTE Teachers	Analysis Sample
Female	65	72	86**
Age (in years)	39	38	36
Race			
White, non-Hispanic	85	78	82
Black, non-Hispanic	3	8**	7
Hispanic	2	2	4
Asian/Pacific Islander	5	2	0
Multiracial	2	3	4
Not specified	4	6	4
Reason for Seeking ABCTE Certification			
To enter teaching profession	49	37**	21*
To retain current position	29	56**	75**
To obtain another teaching position	15	3**	4
To advance career in education in a non-teaching position	2	2	0
Other	5	1*	0
ABCTE Examination Score			
Teaching knowledge	305	299**	298*
Multiple subjects (kindergarten through grade 6)	312	312	309
Mathematics content (grade 6 through 12)	301	290	292
English Language Arts (ELA) (grade 6 through 12)	305	303	303
Certification Subject Area			
Elementary Education	53	47	39
English Language Arts (grade 6 through 12)	20	23	43**
Mathematics (grade 6 through 12)	13	8	18
Teaching in Certified Area	86	90	93
Year Certified			
2004 or 2005	21	8**	14
2006	31	31	25
2007	48	61**	61
Teaching Experience			
1 year	23	15	7*
2 years	16	13	18
3 to 4 years	30	52**	54**
5 to 9 years	21	20	21
10 years or more	10	1	0
School Type			
Public, traditional	71	87**	100**
Public, charter	12	3**	0*
Private	15	8**	0*
Sample Size	**329**	**87**	**28**[a]

Source: Mathematica Policy Research, Alumni Survey 2007.

Notes: Statistically different from the mean for "All ABCTE Teachers" at the 0.05 (*) and 0.01 (**) levels.

[a]Two ABCTE teachers in the analysis sample did not complete the 2007 Alumni Survey and are therefore not represented in Table 1.

The Florida K-20 Education Data Warehouse

To measure the effect of ABCTE teachers on student performance, we collected data on ABCTE teachers, non-ABCTE teachers, and the students taught by both groups. While we were able to use survey data collected by Mathematica to identify the ABCTE teachers, we requested assistance from EDW in determining their counterparts—teachers of the same courses in the same districts—in both the 2006–2007 and 2007–2008 school years. EDW provided background characteristics and test score data for students linked, by course, to these teachers.

Outcomes: FCAT Scores

The Florida Comprehensive Assessment Test, or FCAT, is the statewide instrument used to measure students' and schools' academic achievement. It is a criterion-referenced test measuring benchmarks in reading, writing, mathematics, and science. In the years covered by this study, the FCAT was administered in reading and mathematics in grades 3 through 10 and in science in grades 5, 8, and 11. Performance is reported by using scale scores, which range from 100 to 500 for each grade level. Reliability is generally high, with an Item Response Theory (IRT) marginal reliability over 0.9 and Cronbach's Alpha over 0.85 for all grades in reading and mathematics (FCAT 2007).

To allow us to estimate some measure of value added by the ABCTE teachers, we converted students' scale scores to z-scores to be able to compare performance across subjects, grades, and years of administration. A z-score has the property that a positive number is higher than the state average (relative to a student's same-grade peers), a negative score is below average, and a score of zero is equal to the average. The magnitude can be translated into percentiles using a standard normal table. We calculated z-scores, defined as the difference between a student's score and the mean score divided by the standard deviation, by using statewide means and standard deviations by year, grade, and subject. Given that standard deviations tend to be larger at higher grades, z-scores tend to show less student growth at those levels compared to non-normalized scale scores. Our conclusions are not affected by this normalization, but we used it to facilitate interpretation.

We report on FCAT performance in reading and mathematics for students in grades 4 through 10. Grade 3 students are not included in the analyses that require pre-test data because no test is given in grade 2.

Defining the ABCTE Sample

Although we identified 94 ABCTE teachers in Florida through the 2007–2008 school year, some taught in private schools that do not administer the FCAT. Of the remainder, only a subset of public school teachers taught in subjects and grades with associated FCAT outcomes. Further, we limited the ABCTE group of teachers to those with at least 10 students contributing post-test scores in a given course and those teaching courses for which the teacher in question is primarily responsible for the FCAT score (e.g., English 10, but not playwriting).

Our analysis sample includes 30 ABCTE teachers spanning the 2006–2007 and 2007–2008 school years in Florida—25 with reading test scores and 18 with math scores. In Table 2, we present the disposition of teachers excluded from the sample for the reasons described. The largest number of excluded teachers (23 of 64, or 36 percent) were in grades in which the FCAT is not administered. Another 13 (20 percent) taught untested subjects, while 11 (17 percent) taught science in tested grades but the associated outcome data was not available for the analysis.

Table 2. Composition of Analysis Sample (number of teachers)

Status	2006	2007	Final Sample
In Analysis Sample	10	28	30
Not in Analysis Sample			
Untested grade	7	20	23
Untested subject			
Science	2	11	11
Other	5	13	13
Insufficient number of students	1	1	1
No teaching assignment (supplemental instruction)	1	3	2
Not matched	1	4	4
Teaching in private school	5	7	10
Total	**32**	**87**	**94**

Source: EDW; calculations by authors.

Note: Final status categories are mutually exclusive.

Defining the Comparison Group

We used propensity score matching to define the comparison group of teachers for the analysis sample. Propensity score techniques allow the researcher to match ABCTE and non-ABCTE classrooms on a variety of characteristics by reducing those characteristics to a single dimension: the predicted probability or "propensity" based on those characteristics to be an ABCTE teacher. Once the propensity score is calculated, there are several ways to use it to match study subjects. For the analyses presented here, we selected the five nearest neighbors (with replacement) for each ABCTE teacher. We tested other methods as well, but this approach produced the most balanced sample.

For each ABCTE teacher, we started with a potential comparison group consisting of all teachers who taught the same course and grade level. In this way, an individual teacher's potential comparison group would be different for his or her English 9 versus his or her English 10 class, if he or she taught both grade levels. The potential comparison group members were drawn from the same districts in Florida as the ABCTE teachers, but we did allow matches across districts. We controlled for district differences in the regression analysis.

For each teacher, both treatment and potential comparison, we calculated measures of average student characteristics: race, language, overage for grade, special education status, and mathematics and reading pre-test scores. As a measure of socioeconomic status, we calculated the percent of students eligible for free or reduced-price lunch (FRPL). We also included the teacher's years of experience in the prediction model.

We defined the propensity score as the probability that a given teacher in a given course would be an ABCTE teacher given the observable classroom characteristics. Across the 30 ABCTE teachers, there were 54 teacher/course/year combinations that contributed outcome data. To obtain a predicted value for ABCTE status, we estimated a logit model at the teacher-course level by regressing ABCTE status (whether or not a teacher was certified by ABCTE) on the set of covariates listed above with district fixed effects. Although our analyses were conducted at the student level, we matched at the teacher-course level to account for classroom or peer effects.

In Tables 3 and 4, we present mean values of baseline characteristics at the teacher-course and student levels for the analytical sample for each of several matching strategies. We matched teachers to their nearest neighbor, their nearest three neighbors, nearest five neighbors, and to all teachers who fell within a fixed distance. We implemented this latter approach, known as radius matching, with several cutoff distances, known as radius calipers. Finally, we used kernel density matching which allows for all of the potential comparison group members to be included in the analysis, but weights them according to the overall distance their propensity score lies away from the propensity score of each ABCTE teacher. We implemented these different matching techniques in order to seek out the one that yielded the most balanced sample. These different matching methods are similar, but they make slightly different tradeoffs between being inclusive, which provides a larger and richer comparison group, and being closely matched, which reduces bias.

Table 3. Baseline Characteristics, Teacher-Course Level

		Comparison Group					
Characteristic (percent unless otherwise indicated)	ABCTE	Nearest Neighbor	Nearest 3 Neighbors	Nearest 5 Neighbors	Radius (.0005)	Radius (.01)	Kernel Density
Race							
Percent black	18.9	17.8	18.5	19.2	21.5	21.5**	26.5**
Percent Hispanic	25.8	24.2	23.5	23.3	23.5	24.4	25.3
Percent white	47.9	52.1	51.0	50.5	48.1	47.2	41.8**
Percent Over Age for Grade	13.0	12.0	10.8	11.3	11.7	11.6	12.4
Percent Limited English Proficient	9.5	8.2	8.7	8.0	9.0	8.7	9.2
Percent Low Income (eligible for FRPL)	37.8	37.0	34.4	34.8	38.4	37.0	43.4**
Percent Special Education	8.2	5.4	6.9	7.2	7.4	8.7	13.4**
Pre-test Score (standard deviation units)							
Mathematics pre-test	18.4	29.1	26.9	27.6	19.8	21.2	7.1**
Reading pre-test	13.4	19.7	19.7	21.6	12.6	16.4	2.7**
Teacher's Years of Experience	1.7	1.9	1.9	1.9	1.7	2.6**	7.2**
Sample Size	**54**	**50**	**145**	**233**	**264**	**3343**	**5653**

Source: EDW; calculations by authors.

Note: Deviation from ABCTE mean is statistically significant at the 0.05 (*) 0.01 (**) level.

Table 4. Baseline Characteristics, Student Level

Characteristic (percent unless otherwise indicated)	ABCTE	Comparison Group					
		Nearest Neighbor	Nearest 3 Neighbors	Nearest 5 Neighbors	Radius (.0005)	Radius (.01)	Kernel Density
Race							
Percent black	26.9	17.4**	16.4**	20.5**	16.5**	22.5**	28.6**
Percent Hispanic	31.3	24.3**	23.0**	22.9**	20.9**	24.2**	26.0**
Percent white	39.8	45.5**	52.1**	48.6**	54.1**	46.0**	39.2**
Percent Over Age for Grade	10.7	14.4**	10.8	11.9*	9.2	14.0**	12.2**
Percent Limited English Proficient	9.3	10.1	9.3	8.9	10.2*	9.7*	9.2
Percent Low Income (eligible for FRPL)	40.8	47.9**	32.5**	35.2**	38.1**	33.8**	40.5
Percent Special Education	12.0	5.9**	5.0**	7.2**	4.3**	11.0**	11.0**
Pre-test Score (standard deviation units)							
Mathematics pre-test	1.3	29.0**	27.6**	23.3**	29.4**	20.6**	11.6**
Reading pre-test	-7.2	26.4**	22.3**	23.5**	26.8**	14.9**	8.5**
Sample Size	**2,578**	**2,001**	**6,004**	**9,651**	**9,948**	**128,378**	**213,318**

Source: EDW; calculations by authors.

Note: Deviation from ABCTE mean is statistically significant at the 0.05 (*) 0.01 (**) level.

We found no statistically significant differences between our sample of ABCTE and non-ABCTE teacher-courses for the nearest neighbor methods. On the baseline measure of the outcome variables (pre-test score), the sample of students taught by ABCTE teachers is equivalent to that taught by non-ABCTE teachers, and similar on other demographic characteristics. The magnitudes of the differences can be inferred from Table 3 for each type of comparison group. (Analyses are weighted to account for the possibility that some comparison group members match with more than one ABCTE teacher.) Only the more inclusive radius matching estimator (caliper = 0.01) and the kernel density estimator produced comparison groups whose mean teacher-course level characteristics differed from the ABCTE sample by a statistically significant margin. However, it should be noted that the more inclusive matching estimators are more likely to reject the balancing tests *ceteris parabus* because they result in larger samples and hence better enable the researcher to detect differences and label them as statistically significant. Overall, the differences in the baseline teacher-course characteristics make it somewhat difficult to select a single estimator. We focus on the nearest five neighbors matching estimators in the main analysis, but report on results for estimator separately.

When we compared *students* taught by ABCTE teachers in the study to students of teachers in the matched samples, we found differences of similar magnitudes, but the differences were more likely to be statistically significant because there is more statistical power at the student level for such a balancing test (see Table 4). Also, because we matched at the teacher-course level, there is reason to suspect that student level characteristics may be less balanced than teacher-course level. We chose to match at the level of intact classrooms because the phenomenon we are studying is teacher performance.

Analytic Model Specification

To measure the effects of ABCTE teachers on their students, we estimate the following student achievement growth model, generating robust standard errors that treat teacher-course combinations as clusters:

$$Y_i^t = aY_i^{t-1} + b'X_i + cABCTE_i + u_i$$

where:

Y	= test score for student i in district d in year t
X	= vector of student characteristics (race/ethnicity, income, special educations status)
$ABCTE$	= indicator set to 1 if student i is taught by an ABCTE teacher, or else 0
a, b, c	= parameters to be estimated
u	= error term, assumed to be independent and identically distributed across individuals

Our primary (or benchmark) model regresses post-test on a treatment indicator (whether the student is in a classroom taught by an ABCTE teacher) and a series of covariates, including a pre-test measure and a series of school year, grade, and district fixed effects. This is our most defensible model in terms of statistical assumptions, although we also test its sensitivity to alternative specifications. The alternative specifications included those that omitted selected variables, added non-linear transformations of the pre-test score, estimated separate models by school year, estimated the standard errors based on teachers rather than teacher-course combinations as the unit of clustering for the hierarchical model, and used non-normalized version of test scores (scale scores instead of z-scores).

Because this model includes a predictor variable (a standardized pre-test) that is measured with error, ignoring that error may lead to biased estimates. To test for this, we accounted for errors in two ways: by (1) using an errors-in-variables approach with measures of FCAT reliability by subject and grade level and (2) estimating the impact on the gain score (the difference between the post-test and pre-test scores). Our findings using either of these approaches were similar in scale and significance with those presented in our benchmark model. They are available on request from the authors.

Findings

Student Performance of ABCTE Versus Comparison Teachers

The main findings are shown in Table 5. We were not able to detect any differences between the reading scores of ABCTE teachers' students and those of non-ABCTE teachers' students. The average FCAT reading scores, accounting for differences in observable characteristics between the groups, were about 20 percent of a standard deviation above the state mean in ABCTE classrooms and 8 percent of a standard deviation above the mean in comparable non-ABCTE classrooms (see Table 5). The difference between ABCTE and comparison classrooms was not statistically significant, meaning that we could have observed such a difference purely by chance.

Table 5. Test Score Effects—Benchmark Model

Outcome	ABCTE Mean[a]	Comparison Mean[a]	Difference	Standard Error	Number of students	R-squared
Reading	0.204	0.080	0.124	0.080	9,791	0.837
Mathematics	-0.193	0.059	-0.252*	0.100	2,438	0.951

Source: EDW; calculations by authors.

Notes: Difference is statistically significant at the 0.05 (*) and 0.01 (**) levels.

Pre-test values are imputed when missing unless missing for entire class.

[a]Means are regression-adjusted, expressed here as a fraction of the statewide standard deviation in test scores, normalized within grade and subject.

In mathematics, however, the students in ABCTE teachers' classrooms scored 25 percent of a standard deviation lower than students in comparable classrooms. This difference was statistically significant, meaning that we are unlikely to have observed a difference of this magnitude purely by chance. Twenty-five percent of a standard deviation is equivalent to moving a classroom average nearly ten percentile points from the 53rd percentile (where the students in comparison classrooms scored) down to the 43rd percentile.

The results shown in Table 5 are based on a regression model that controls for the background characteristics of the students in the sample so as not to confound differences in these characteristics with the effect of having an ABCTE teacher. Using this as a benchmark model, we tested several alternative regression models to determine whether the results were sensitive to model specification, as discussed above.

The results of the alternative models (shown in Table 6) suggest that the findings are robust, although the control variables play an important role in reducing bias and increasing precision of the estimates. The regression coefficient on the reading score is not statistically significant under any of the six model specifications presented in the table. The size of the negative coefficient on math scores varied as we included more control variables to eliminate the influence of confounding factors, such as ABCTE teachers being over-represented in a particular year, grade, or district relative to the comparison teachers. Failure to control for district differences would have led to a smaller effect, a larger standard error, and the conclusion that the estimated negative effect was too small to be statistically significant. The most comprehensive model (column 6) suggests a difference that is statistically significant at the 0.05 level. This is the benchmark model that was reported in Table 5.

The general conclusion does not change when we use different matching methods. The benchmark model is based on a comparison group made up of five nearest neighbors for every ABCTE teacher-course combination. Table 7 presents the estimated effects on math and reading scores using the different matching methods discussed in the previous section. The three-neighbor estimator shows a significant positive relationship between ABCTE certification and reading scores (effect size = 0.23, which is similar in magnitude to the negative math effect under the benchmark model), but all of the other matching estimators suggest a smaller effect, below the threshold of statistical significance for this study.

Table 6. Test Score Effects (z-scores)—Alternative Sets of Control Variables

	Model					
Regression Results	1	2	3	4	5	6
ABCTE Coefficient, Reading	**-0.112**	**0.107**	**0.104**	**0.104**	**0.099**	**0.124**
(Standard error)	(0.282)	(0.081)	(0.085)	(0.084)	(0.079)	(0.079)
Number of students	9,791	9,791	9,791	9,791	9,791	9,791
R-squared	0.003	0.782	0.808	0.808	0.826	0.837
ABCTE Coefficient, Math	**-1.325****	**-0.297**	**-0.453***	**-0.389**	**-0.242**	**-0.252***
(Standard error)	(0.479)	(0.207)	(0.186)	(0.202)	(0.144)	(0.100)
Number of students	2,438	2,438	2,438	2,438	2,438	2,438
R-squared	0.236	0.814	0.874	0.878	0.933	0.951
Included Covariates						
Pre-test score		x	x	x	x	x
Student characteristics			x	x	x	x
Year fixed effects				x	x	x
Grade fixed effects					x	x
District fixed effects						x

Source: EDW; calculations by authors.

Notes: Statistically significant at the 0.05 (*) and 0.01 (**) levels.

Pre-test values are imputed when missing unless missing for entire class.

The estimated effect of ABCTE teachers on math scores is statistically significant and lies between -0.4 and -0.2 according to the sensitivity tests shown in Table 7. Only the radius matching with a caliper of 0.0005 results in an estimate that is not statistically significant at the 0.05 level.

Table 7. Sensitivity to Alternative Matching Procedures

Subject/Method	Coefficient	Standard Error	P-value	Sample Size
English/Language Arts Teachers				
Nearest neighbor	0.15	(0.11)	0.166	3,714
Nearest 3 neighbors	0.23**	(0.08)	0.006	6,926
Nearest 5 neighbors	0.12	(0.08)	0.120	9,791
Radius match (caliper=0.0005)	0.14	(0.10)	0.150	9,767
Radius match (caliper=0.01)	0.03	(0.07)	0.643	94,731
Kernel density	0.05	(0.07)	0.422	166,371
Mathematics Teachers				
Nearest neighbor	-0.42**	(0.03)	0.000	865
Nearest 3 neighbors	-0.30*	(0.13)	0.028	1,656
Nearest 5 neighbors	-0.25*	(0.10)	0.014	2,438
Radius match (caliper=0.0005)	-0.19	(0.12)	0.120	4,483
Radius match (caliper=0.01)	-0.27*	(0.11)	0.016	36,225
Kernel density	-0.27**	(0.09)	0.002	49,345

Subgroup Analyses

One way to interpret the findings is to examine subgroups of teachers. This analysis is useful for identifying whether one type of teacher might be driving the general results presented above or whether the full sample results might mask some important heterogeneity. But reduced sample size and statistical power limit our ability to make meaningful or generalizable inferences about the performance of one class of ABCTE teacher or another.

The estimates of ABCTE teachers' effect on test scores for subgroups of teachers are shown in Table 8. We conducted this analysis using nearest neighbor matching with one neighbor for practical reasons, but the result did not depend on this choice. The finding of no significant differences for reading scores held for all subgroups we analyzed, although the point estimates tended to be negative, favoring the comparison group. In other words, ABCTE teachers who had sought the certification in order to enter the teaching profession and those who scored above average on the certifying exams did not achieve higher test score gains than their matched comparison non-ABCTE teachers. In terms of math scores, the student test scores for every subgroup of ABCTE teachers were lower than the student scores for matched non-ABCTE classrooms. Some of these differences were statistically significant, but others were not, because of large standard errors associated with small sample sizes at the subgroup level. Because there are even fewer ABCTE math teachers in the sample than ELA teachers, the sample sizes for the subgroup performance in math are especially small and inadequate to draw general conclusions from these data.

We repeated the analysis by matching just within novice teachers, since many ABCTE teachers are career changers and/or new to the teaching profession. In so doing, we examined the effects of novice ABCTE teachers (less than five years of experience) to novice non-ABCTE teachers. We found that limiting the potential matches in the comparison sample to teachers with less than five years' experience did not change the story in a qualitative way. For reading the effects were still statistically insignificant. For math, the effect was negative, and despite being large (-2.1 standard deviations) was marginally insignificant ($p=0.056$) due to the small sample of 23 classrooms.

Conclusions

Summary of Findings

Did a group of ABCTE-certified teachers produce student achievement gains on par with their non-ABCTE counterparts? On reading measures, the evidence from the Florida data shows that student gains were statistically indistinguishable between students of ABCTE and non-ABCTE teachers. In mathematics, students of ABCTE teachers scored lower than students of non-ABCTE teachers. Controlling for previous performance, district differences, and other background characteristics, students of ABCTE teachers of mathematics scored 0.25 standard deviation units lower than their counterparts in non-ABCTE classrooms in the most plausible of the statistical models we used. The findings are consistent across model specifications and subgroups, although the size of the effect varies somewhat.

To help interpret these findings, in Table 9 we present characteristics of the analytic sample by subject area, compared with ABCTE teachers overall. Two main patterns emerge. First, mathematics teachers in our sample tend to have been certified more recently and were more likely to have obtained that certification to enter teaching. Correspondingly, they have less teaching experience than either the English/Language Arts (ELA) teachers in the sample or ABCTE teachers overall.

Also notable is that, while the differences are not statistically significant, mathematics teachers in the sample scored lower on both the teaching knowledge (297 versus 305) and mathematics content examinations (292 versus 301) than ABCTE teachers overall. Meanwhile, ELA sample teachers performed better than average on the ELA content exam (309 versus 306). These findings suggest that the ABCTE examinations may be a reliable predictor of performance, but more data would be needed to confirm this hypothesis.

Table 8. Subgroup Analysis of Test Score Effects

Subject/Subgroup	ABCTE Mean[a]	Comparison Mean[a]	Difference	P-value of Difference	Number of Classrooms	Number of Students
Reading						
Full Sample	0.22	0.07	0.15	0.166	81	3,714
Certification Area						
Elementary	-0.06	0.50	-0.55	0.470	39	1,862
Secondary mathematics	0.20	0.22	-0.02	0.952	75	3,640
Novice Teachers (fewer than 5 years' experience)	0.05	0.19	-0.14	0.672	77	3,502
Reason for Seeking ABCTE Certification						
To enter teaching profession	-0.47	0.00	-0.47	0.462	20	854
To retain current position	0.21	0.38	-0.17	0.620	90	4,520
ABCTE Test Performance						
Professional teaching knowledge exam						
Above average score	0.01	0.41	-0.40	0.276	47	1,840
Below average score	0.14	0.28	-0.14	0.734	68	3,587
Subject exam						
Above average score	-0.61	0.11	-0.72	0.090	46	2,037
Below average score	0.56	0.43	0.13	0.716	70	3,517
Mathematics						
Full Sample	-0.13	0.56	-0.42**	0.000	23	865
Certification Area						
Elementary	-0.72	-0.01	-0.71	0.177	16	363
Secondary mathematics	-0.29	2.25	-2.54	0.036	11	666
Novice Teachers (fewer than 3 years' experience)	-0.68	1.45	-2.13	0.056	23	865
Reason for Seeking ABCTE Certification						
To enter teaching profession	0.68	1.79	-1.11	0.268	12	368
To retain current position	-0.79	1.07	-1.86	0.084	17	757
ABCTE Test Performance						
Professional teaching knowledge exam						
Above average score	-1.23	-0.45	-0.78	0.141	7	186
Below average score	-0.26	1.76	-2.02	0.057	20	843
Subject exam						
Above average score	0.15	1.26	-1.11	0.315	15	545
Below average score	-0.91	1.46	-2.36*	0.035	14	580

Source: EDW; calculations by authors.

Note: Statistically significant at the 0.05 (*) and 0.01 (**) levels. All results are based on matching to one nearest neighbor.

[a] Means are regression-adjusted.

Table 9. Characteristics of All Teachers Versus Sample Teachers, by Subject Area

Characteristics (percent unless otherwise indicated)	All ABCTE Teachers	Sample Teachers by Certification		
		Elementary	ELA	Math
Female	65	85	100	60
Age (in years)	39	35	36	36
Race				
White, non-Hispanic	85	85	83	80
Black, non-Hispanic	3	0	8	20*
Hispanic	2	8	0	0
Asian/Pacific Islander	5	0	0	0
Multiracial	2	8	0	0
Not specified	4	0	8	0
Reason for Seeking ABCTE Certification				
To enter teaching profession	49	31	17*	40
To retain current position	29	69**	75**	60
To obtain another teaching position	15	0	8	0
To advance career in education in a non-teaching position	2	0	0	0
Other	5	0	0	0
ABCTE Examination Score				
Teaching knowledge	305	294	300	297
Multiple subject examination (kindergarten through grade 6)	312	309		
Mathematics content (grade 6 through 12)	301			292
ELA content (grade 6 through 12)	305		303	
Certification Subject Area				
Elementary Education	53	100**	0**	0**
English Language Arts (grade 6 through 12)	20	0	100**	0
Mathematics (grade 6 through 12)	13	0	0	100**
Teaching in Certified Area	86	75	100	100
Year Certified				
2004 or 2005	21	23	8	0
2006	31	38	25	20
2007	48	38	67	80
Teaching Experience				
1 year	23	8	8	0
2 years	16	17	8	40
3 to 4 years	30	58*	42	60
5 to 9 years	21	17	42*	0
10 years or more	10	0	0	0
School Type				
Public, traditional	71	100**	100*	100
Public, charter	12	0	0	0
Private	15	0	0	0
Sample Size	**329**	**13**	**12**	**5**

Source: MPR Alumni Survey 2007.

Note: Statistically different from all ABCTE teachers at the 0.05 (*) and 0.01 (**) levels.

Limitations of the Study

Although we were able to create a high-quality matched comparison group using a large pool of potential comparison teachers, there are some limitations of the study.

First, the study focuses on a particular group of ABCTE teachers for whom data are available--Florida teachers in tested grades and subjects during the two most recent years for which data were provided. These teachers may be more or less skilled relative to their peers than other ABCTE teachers in other states or teaching assignments; nonetheless, caution is advisable before extrapolating to other contexts.

Second, our comparisons control for easily measured confounding factors, but unmeasured factors remain, some of which could explain differences between test score outcomes in ABCTE and comparison classrooms. For example, school principals may assign less motivated or more behaviorally challenged students to ABCTE teachers than to other teachers. On the other hand, perhaps principals assign *more* motivated and *less* challenging students to ABCTE teachers. We have no evidence of either type of student assignment and therefore no evidence on the direction of bias from omitted variables. At the same time, we do not know if ABCTE teachers seek out positions or are hired into positions in hard-to-staff schools whose students and working conditions might have as much to do with test score outcomes as our estimated teacher effects.

Despite these caveats, the study provides new direct evidence on the performance of ABCTE teachers in the classroom, something that has not been available to policymakers to date. Before making high-stakes policy decisions regarding the value of ABCTE certification, more research would be useful to replicate the results for larger samples and in more contexts. The analytic approach used here is applicable to additional states, outcomes, and cohorts of ABCTE teachers. Further, experimental designs have been successfully implemented in other studies of alternative teacher preparation (Glazerman et al. 2006; Cantrell et al. 2008; Constantine et al. 2009) and should be considered for providing more definitive evidence in the future.

References

Ballou, Dale, and Michael Podgursky. *Teacher Pay and Teacher Quality*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1996.

Boyd, Donald, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. "Teacher Preparation and Student Achievement." NBER Working Paper No. 14314, 2008.

Cantrell, Steven, Jon Fullerton, Thomas J. Kane, and Douglas O. Staiger. "National Board Certification and Teacher Effectiveness: Evidence from a Randomized Experiment." NBER Working Paper No. 14608, 2008.

Cavalluzzo, Linda. "Is National Board Certification an Effective Signal of Teacher Quality?" Alexandria, VA: The CNA Corporation, 2004.

Chaplin, Duncan, Steven Glazerman, Gail Baxter, and Elizabeth Seif. "Teacher Licensure Exams: How Do They Compare?" Washington, DC: Mathematica Policy Research, 2007.

Constantine, Jill, Dan Player, Tim Silva, Kristin Hallgren, Mary Grider, and John Deke. *An Evaluation of Teachers Trained Through Different Routes to Certification, Final Report* (NCEE 2009-4043). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2009.

Cunningham, G.K., and J.E. Stone. "Value-Added Assessment of Teacher Quality as an Alternative to the National Board for Professional Teaching Standards: What Recent Studies Say." In *Value Added Models in Education: Theory and Applications*, edited by Robert Lissitz. Maple Grove, MN: JAM Press, 2005.

Ehrenberg, Ronald, and Dominic J. Brewer. "Do School and Teacher Characteristics Matter? Evidence from High School and Beyond." *Economics of Education Review*, vol. 14, 1994, pp. 1-13.

FCAT Reading and Mathematics Technical Report. San Antonio, TX: Harcourt, 2007.

Feistritzer, Emily. "Preparing Teachers for the Classroom: The Role of the Higher Education Act and No Child Left Behind." Testimony to the Committee on Education and Labor, Subcommittee on Higher Education, Lifelong Learning, and Competitiveness. U.S. House of Representatives, May 17, 2007.

Glazerman, Steven, Daniel Mayer, and Paul Decker. "Impacts of Teach For America on Student Achievement and Other Outcomes." *Journal of Policy Analysis and Management*, vol. 25, no. 1, Winter 2006, pp. 75-96.

Glazerman, Steven, Elizabeth Seif, and Gail Baxter. "Passport to Teaching: Career Choices and Experiences of American Board Certified Teachers." Washington, DC: Mathematica Policy Research, 2008.

Glazerman, Steven, and Christina Tuttle. "An Evaluation of American Board Teacher Certification: Progress and Plans." Washington, DC: Mathematica Policy Research, 2006.

Glazerman, Steven, Christina Tuttle, and Gail Baxter. "School Principals' Perspectives on the Passport to Teaching." Washington, DC: Mathematica Policy Research, 2006.

Goldhaber, Dan, and Anthony, Emily. "Can Teacher Quality be Effectively Assessed? National Board Certification as a Signal of Effective Teaching." *Review of Economics and Statistics,* vol. 89, no. 1, 2007, pp. 134-150.

Goldhaber, Dan, and Dominic J. Brewer. "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement." *Education Evaluation and Policy Analysis*, vol. 22, no. 2, 2000, pp. 129-145.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. "What Does Certification Tell us About Teacher Effectiveness? Evidence From New York City." *Economics of Education Review,* vol. 27, no. 6, 2008, pp. 615-631.

National Research Council. *Assessing Accomplished Teaching: Advanced-Level Certification Programs*. Committee on Evaluation of Teacher Certification by the National Board for Professional Teaching Standards. Milton D. Hakel, Judith Anderson Koenig, and Stuart W. Elliott, editors. Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press, 2008.

Vandevoort, L.G., A. Amrein-Beardsley, and D.C. Berliner. "National Board Certified Teachers and Their Students' Achievement." *Education Policy Analysis Archives*, vol. 12, no. 46, 2004. Available at [http://epaa.asu.edu/epaa/v12n46/].

Princeton, NJ · Ann Arbor, MI

Cambridge, MA · Chicago, IL

Oakland, CA · Washington, DC

Improving public well-being by conducting high-quality, objective research and surveys

Visit our website at www.mathematica-mpr.com

MATHEMATICA
Policy Research, Inc.