Institute for Federalism & Intergovernmental Relations

IFIR WORKING PAPER SERIES

The Value of a College Education: Estimating the Effect of Teacher Preparation on Student Achievement

Sharon Kukla-Acevedo
Eugenia F. Toma

IFIR Working Paper No. 2009-06

Acknowledgments. The authors are indebted to the Kentucky Education Professional Standards Board for the opportunity to use administrative data that are not publicly available and to Terry Hibpshman for his role in extracting the necessary data.

The Value of a College Education: Estimating the Effect of Teacher Preparation on Student Achievement

Abstract

Federal legislation currently holds institutions of higher education accountable for the quality of teachers that they produce. However research has yet to demonstrate that teacher preparation programs (TPPs) have differential effects on the quality of teachers they produce in terms of student achievement. This study uses data from a sample of 2,582 5th grade math students in an urban school district in Kentucky and a school fixed effects design to explore the variation in average TPP effects. The authors find that TPPs are differentially effective in training teachers, which in turn impacts student performance on 5th grade math scores. There is also some indication that these differential effects converge around teachers' fifth year of teaching.

Keywords: Student achievement; teacher preparation, teacher effects

Sharon Kukla-Acevedo*
Deparment of Political Science
Central Michigan University
Anspach Hall 246
Mount Pleasant, MI 48859
kukla1sa@cmich.edu

Eugenia F. Toma
Martin School of Public Policy
and Administration
University of Kentucky
437 Patterson Office Tower
Lexington, KY 40506
Eugenia.Toma@uky.edu

*Corresponding author

The Value of a College Education: Estimating the Effect of Teacher Preparation on Student Achievement

INTRODUCTION

Current federal legislation reflects teachers' critical role as the most important institutional factor in the student learning process. *The No Child Left Behind Act* mandates the placement of a highly qualified teacher in every classroom, while Title II of the *Higher Education Act* (HEA) requires that states hold institutions of higher education publicly accountable for the quality of the teachers they produce. Under Title II, each state must report annually on licensure requirements, pass rates on certification assessments, state performance evaluations of teacher preparation programs, and the number of teachers in the classroom on waivers.

In response to these major pieces of legislation, states began looking closely at the quality of their teacher preparation programs. The Ohio Teacher Quality Partnership and the Massachusetts Coalition for Teacher Quality and Student Achievement are statewide collaborations that are undertaking comprehensive efforts to create datasets and projects that will evaluate the relationship between teacher preparation and student achievement. The Louisiana Board of Regents is funding pilot efforts to determine whether Louisiana's existing student achievement, teacher, and curriculum databases can be used to assess teacher preparation programs in the state (Noell, 2006). The collaborations are still in developmental stages, but researchers in Louisiana have produced studies that look at the differential effectiveness of teacher preparation programs, in terms of student achievement gains (Noell, 2006). Similarly, a group of researchers using New York City data has begun looking not only at the variation

between individual TPPs, but also at the key components that these programs utilize to train teachers most effectively (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2008).

Existing state administrative data is used in this paper to test two hypotheses regarding the effects of math teachers' preparation program on student achievement. The study attempts to address whether teacher preparation programs are differentially successful in training teachers, and how these effects change as teachers gain experience in the classroom. Similar to the Louisiana studies, this project explicitly models the effect of individual pre-service teacher preparation programs. This is in contrast to prior work which groups programs into quality categories (Clotfelter, Ladd, and Vigdor, 2006, 2007; Summers and Wolfe, 1977; Ehrenberg and Brewer, 1994; Murnane and Phillips, 1981) or types of programs (Andrew and Schwab, 1995; Andrew, 1990; Good, McCaslin, Tsang, Zhang, Wiley, Rabidou Bozak, and Hester, 2006; Wenglinsky, 2000; Rebeck, 2004).

Literature Review

The college environment is a setting that provides substantial opportunities to change and develop intellectually. Most colleges familiarize students with diverse sources of knowledge, facilitate training in logic and critical thinking, and present alternative ideas and courses of action (Floden and Meniketti, 2005). In their review of over 3,000 studies that look at the effect of college on student outcomes, Pascarella and Terenzini (1991) find that college students gain knowledge over their course of study and the gains are larger in their focal areas. They find little evidence that students' cognitive skills are increased by the college experience. Rather, college effects lead to improvement in students' communication, ability to analyze and think critically, and ability to judge and respond appropriately to external events (Pascarella and Terenzini, 1991).

Post-secondary institutions are diverse in terms of size, selectivity, and affiliation. Programs employ different pedagogical methods and foci that they deem best facilitate their students' gains. These vast differences across colleges and universities lead to the reasonable assumption that colleges have differential effects in terms of student learning. In the context of teachers' pre-service training, these points indicate that teacher preparation programs help future teachers to gain knowledge regarding classroom techniques and pedagogy, as well as develop critical skills needed to deliver their specialized knowledge. However, there is no evidence indicating that all teacher preparation programs are created equal. Rather, the two studies that directly measure the efficacy of teacher preparation programs detect differences in teacher effectiveness, especially during the first one to three years of teaching (Boyd et al., 2008; Noell, 2006).

While the link between individual teacher preparation programs and student outcomes has not been studied in-depth, there have been research efforts that seek to determine whether students learn more from teachers who graduate from highly-rated institutions. These studies use Barron's or Gourman's ratings of colleges to serve as an indicator for the quality of training the teachers receive. Generally, the results of these studies find little or no relationship between quality of training and student achievement (Clotfelter et al., 2006, 2007; Ehrenberg and Brewer, 1994; Murnane and Phillips, 1981), although one study determined that college quality is a predictor of student achievement (Summers and Wolfe, 1977). Clotfelter and colleagues (2006, 2007) provide the most recent effort and their methodological designs are strong. Using a student-teacher matched dataset from the state of North Carolina, they find no impact when analyzing schools in which students and teachers appear to be randomly assigned (2006) or when estimating student fixed effects models (2007). Summers and Wolfe (1977) also use a student

fixed effects design, although their gain score model is less robust than that used in the North Carolina studies because the test was not uniform from year to year.

Geographic location, methodological nuances, and differences in time period can all account for the disparate findings on college ratings, yet this measure may be problematic on theoretical grounds. A rating that represents the quality of an entire undergraduate institution may have very little relevance to the quality of one program at that institution. It is quite feasible that high quality teacher preparation programs exist at low-rated undergraduate institutions, and vice versa. This aggregate measure masks important variation among teacher education programs, which could result in an apparent lack of relationship.

The bulk of the research on teacher preparation focuses on the implications that different pathways to teaching have for student achievement. These studies look at whether teachers who have been trained in undergraduate teacher education programs are more effective than teachers who received training outside of a traditional teacher education curriculum. Teach for America is a highly salient example of an alternative pathway into teaching, although it should be noted that many states have a form of provisional, temporary, or emergency entry into the teaching workforce. The results of these studies tend to support the traditional university pathway into teaching. Student gains are generally larger (Clotfelter et al., 2007; Goldhaber & Brewer, 2000; Laczko-Kerr, 2002; Hawk, Coble, & Swanson, 1985), graduates of these programs feel more prepared (Darling-Hammond, Chung, & Frelow, 2002; Jelmberg, 1996), and they have higher classroom performance than their alternatively-certified counterparts (Good et al., 2006; Houston, Marshall, & McDavid, 2003; Hawk and Schmidt, 1989). This result, while strong, should be viewed with some caution. Three recent, high-quality studies find mixed evidence regarding the effects of certification on student achievement (Betts, Zau, and Rice, 2003; Boyd,

Grossman, Lankford, Loeb, and Wyckoff, 2005; Darling-Hammond, Holtzman, Gatlin, and Heilig, 2005). They find that the effect of pathway varies according to teacher experience and the subject matter taught. There are at least two reasons why this occurs. First, as Boyd et al. (2005) note, variation in effectiveness is often greater within each pathway than between pathways. Second, Darling-Hammond, Berry, and Thoreson (2001) note that many researchers do not make an important distinction among fully certified teachers. Specifically, prior to *No Child Left Behind*, it was possible for teachers to teach outside of their subject of expertise.

Far less attention in the literature is paid to the variation within the traditional pathway to teaching. Teacher training programs often include both 4-year and 5-year options. Five-year programs are characterized by stricter entry requirements, fewer education pedagogy courses, and longer student-teaching internships. While no study assesses whether these variants of traditional pathway differentially affect student achievement, there are indications that graduates from the two types of programs have differential rates of success in the schools. Andrew (1990) finds that perceptions of training quality were higher among graduates of 5-year teacher programs than those of 4-year programs, while Andrew and Schwab (1995) report that graduates of extended programs have higher rates of leadership involvement.

The current paper begins to address an important gap in the literature. Rather than use aggregate proxies of program quality, the analysis directly assesses whether individual teacher education programs impact student achievement. The authors expect to find main results that are consistent with Noell (2006) and Boyd et al. (2008), namely that teacher preparation programs are differentially successful in training pre-service teachers. The authors also predict that the TPP effect changes differentially over time according to program. By replicating basic trends in a

different geographic region and using a different estimation strategy, this study provides important information to an area of research that is still very much in its initial stages.

METHOD

Sample

Like most states, Kentucky's education administrative data serve multiple purposes and the data are not collected for research priorities. Three levels of data (school, district, and state) are collected and coded separately by a variety of divisions within the Kentucky Department of Education (KDE), the Kentucky Education Professional Standards Board (EPSB), and the Kentucky Council on Postsecondary Education (CPE). For example, EPSB collects data on teacher assignments at the school level, teacher experience and salary at the district level, and teacher assessment at the state level. KDE collects data on student demographics at the school level and student assessment data at the state level. This silo-ed data collection system results in KDE, EPSB, and CPE collecting different data pieces that are required to complete the value-added student learning puzzle. This arrangement does not appear to be unique to Kentucky. In fact, in most states at this time, the student-teacher matches are not available in a centralized, state location. States typically retain individual student information and individual teacher information but not in a way that enables the researcher to match the two.

Kentucky has a relatively decentralized public school system with 175 school districts for its approximately 670,000 K-12 students. With the approval of the EPSB, one urban school district agreed to provide its 5th grade classroom rolls to enable researchers to match teachers to students. All 5th graders participate annually in the math portion of the Kentucky Core Content Test (KCCT). This is an important test area to study, given the current focus of the federal *No*

Child Left Behind Act on math performance. The participants in this paper are students who were in 5th grade in either the 2001-2002 or the 2002-2003 school years. EPSB compiled student level data for approximately 65 percent of the district's 5th graders for the two academic years. As described above, the multiple data sources complicate the student teacher match, but EPSB was able to match teachers to approximately 28 percent of the district's 5th graders. After accounting for missing information on all variables, the study sample consists of 2,582 students.

The rate of participation and the amount of missing data provide justification to look more closely at the study sample. Specifically, it is important to discern whether the study students differ appreciably from those in the entire dataset on important variables that will be used in the statistical analysis. Table 1 provides means and standard deviations of key variables for both groups of students, as well as the results of two-group mean comparison hypothesis tests, which provide statistical evidence of whether the two groups are different. With one exception, students in the study sample are not statistically different than those who appear in the dataset, but have missing information on key variables. Half of the students in the study are female, 64 percent are European American and 33 percent are African American. Latinos/as, Asian Americans and students of another race together make up about 4.5 percent of the study sample. About 55 percent of the students received either free or reduced-price lunch and nearly eight percent of the students have some sort of Individualized Education Plan (IEP). The t-test provides modest evidence that students differ slightly on their 4th grade reading tests. Students in the study sample scored slightly lower on the KCCT 4th grade reading test than students in the full dataset (0.031 vs. 0.052).

The t-tests do indicate that the teachers with missing information are different from the sample teachers. In the study sample, there is a higher proportion of teachers who are European

American and consequently a lower proportion of teachers that are African American than in the full dataset. Additionally, teachers have about one year less experience and have slightly higher GPAs than teachers in the full dataset. The two samples also differ on the proportion of teachers that graduated from each teacher preparation program (TPP). These differences should only affect the generalization of these results, not results themselves. The teacher study sample is nearly 88 percent female, 87 percent European American, and has 13.8 years of teaching experience. On average, study sample teachers entered college with a 21.8 ACT composite score and graduated college with a 2.946 overall GPA. Thirty three percent of the study sample teachers graduated from a single TPP, while the next largest TPP category is that of out of state programs (18 percent). Roughly half of the study sample teachers graduated from the remaining 11 TPPs.

Design

There are several estimation challenges that must be taken into consideration when evaluating TPPs based on student achievement. The first is bias introduced by the non-random sorting of students and teachers that may unfairly inflate or deflate coefficient estimates. The conventional argument is that selection bias occurs on account of at least two sources of non-random sorting among students and teachers. Families and teachers choose neighborhoods and schools based on certain preferences (Tiebout, 1956). Generally, when faced with relocation decisions, more affluent families choose to live in districts that allow them to send their children to higher performing schools. In a similar vein, the most highly employable teachers tend to choose to work in more desirable schools. Additionally, students are placed among classrooms within schools according to such characteristics as academic ability and behavior considerations (Clotfelter et al., 2006). A teacher may be assigned more challenging students because he or she

has demonstrated success in containing certain types of behaviors, or a teacher may be assigned higher achieving students as a reward for excellent service to the school or continued improvement on his or her students' test scores. This selection bias makes it difficult separate TPPs' causal effects from the effects of pre-existing differences among classrooms for which the TPP has no influence. To mitigate these sources of bias, researchers typically use gain scores as the outcome variable of interest and/or some combination of student, teacher, and fixed effects (Boyd et al., 2008; Clotfelter et al., 2006; Harris & Sass, 2006; Author, 2009).

Recent research, however, provides some indication that this type of value-added model may not be the most appropriate estimation strategy to model the effect of TPPs on student achievement. First, Rothstein (2008a) demonstrates that a gain score, which has been used to attribute a student's academic gain over the course of a year to his teacher, may be an unfair credit or discredit. He shows that students' gains over the course of multiple years are dynamic and subject to mean reversion. Specifically, a student who makes higher than average gains in 4th grade will more than likely make smaller than average gains in 5th grade. In a subsequent study, Rothstein (2008b) demonstrates that a more accurate model incorporates lagged scores as control variables, and additional lagged scores further mitigate bias in the estimates. The present study incorporates elements of Rothstein's findings by using two lagged test scores as control variables in the models. Additionally, Boyd, Grossman, Lankford, Loeb, and Wyckoff (2008) provide some rationale to eliminate the teacher fixed effect when estimating whether teachers from one TPP are more effective than teachers from another TPP. They argue that the relative success of programs may be partly due to their ability to recruit and retain college students.

Taking into account these multiple sources of bias, relationship between TPP and 5th grade math achievement is represented by the following cross-sectional model:

(1) $A_{ijmt}=\beta_0+\beta_1 A_{ijm(t-n)}+\beta_2 Stu_{it}+\beta_3 Tch_{jmt}+\beta_4 C_{ijmt}+\beta_5 TPP_j+\lambda_m+u_{ijmt}$

where A_{ijmt} is a standardized 5th grade KCCT math score, $A_{ijm(t-n)}$ is a vector of two lagged achievement scores, and TPP_j is a vector of indicator variables capturing the teacher's preparation program. Stu_{it} is a vector of student-specific characteristics, such as race, gender, and subsidized lunch eligibility; Tch_{jmt} captures teacher-specific characteristics, including gender, race, experience, ACT composite score and college GPA. The subscripts denote students (i), teachers (j), schools (m) and time(t), while λ_m is a school fixed effect and u_{ijmt} is a random error term. Of primary interest is the estimation of TPP, which, if correctly modeled, can be interpreted as the impact of teacher pre-service education on student math gains.

If differential effects of TPP are detected in these initial analyses, then a more nuanced approach may be warranted. A substantial number of studies report a positive relationship between teacher experience and student test scores (Rivkin, Hanushek, & Kain, 2005; Jepsen, 2005; Noell, 2005, 2001; Rockoff, 2004; Goldhaber & Anthony, 2007; Clotfelter, Ladd & Vigdor, 2006; Krueger, 1999; Goldhaber & Brewer, 1997; Sanders, Ashton & Wright, 2005). There is also ample evidence demonstrating that the effect is non-linear in nature. Substantial improvements in teaching skill occur during the first three to five years in the classroom with the effects generally tapering off around the fifth year (Rivkin et al., 2005).

Given this demonstrated effect of experience in the literature, the analysis also considers whether the role of experience operates uniquely over time for teachers from different TPPs. To do this, the authors multiply years of experience times TPP and insert these interaction terms into the model:

(2) $A_{ijmt}=\beta_0+\beta_1 A_{ijm(t-n)}+\beta_2 Stu_{it}+\beta_3 Tch_{jmt}+\beta_4 C_{ijmt}+\beta_5 TPP_j+\beta_6 Exp_{jmt}+\beta_7 TPP*EXP+\lambda_m+u_{ijmt}$

Measures

The outcome measure is the individual KCCT 5th grade math score. The KCCT is a criterion-referenced test that assesses individual student performance against a specified set of state educational goals and consists of both multiple-choice and open-response questions. The test scores are converted to grade-by-year Z-scores with a state mean of zero and standard deviation of one. The math achievement mean of students with complete teacher information is 0.107 with a standard deviation of 1.064, suggesting that this sample of students performs slightly higher than other 5th grade math students in the state. The models incorporate individual 4th and 3rd grade test scores to control for prior student performance. During the time period of study students were not tested in the same subject area in consecutive years, so the 4th grade KCCT scores are in reading. The reading scores are similarly converted to grade-by-year Z-scores and the sample students performed slightly higher than the statewide average performance. The third grade test score is the math subject test from the Comprehensive Test of Basic Skills (CTBS). CTBS is a nationally norm-referenced test that assesses students at the end of a given school year. The CTBS scores are similarly converted to grade-by-year Z-scores with a national mean of zero and standard deviation of one. This sample of Kentucky students performed at about 1/10th of a standard deviation lower than the national average.

Additional student variables are included in the models to control for demographics and, to some extent, family income. Dichotomous variables indicate whether the student is female, African American, Latino/a, Asian American, or another race not listed ("other"). Male and European American students are used as reference categories. An indicator variable designates those students who receive some form of federally subsidized lunch. Table 1 provides means and standard deviations for the student characteristics. The table indicates a racially diverse district

with 62.4 percent European American students and 33.2 percent African American students. Asian American and Latino/a students constitute only about one percent each but these are both growing segments of the population in this district. Female students make up 50.1 percent of the population and 55.3 percent of students receive some form of federally subsidized lunch. A series of indicator variables represent 12 in-state teacher preparation programs.

EPSB recognizes 30 institutions of higher education with teacher training programs in the state, but only eight are publicly funded institutions with substantial numbers of graduates annually. The majority of the programs are located in small, private institutions that produce a limited number of education graduates per year. Attempts to estimate program effects with small numbers of graduates would likely result in noise, so programs with fewer than 30 teacher-student observations or fewer than three teacher graduates overall are grouped into a category entitled "Other TPP." There were also a number of out of state TPPs for which limited information was available to the researchers. Out of state programs generally did not meet the criteria for inclusion as an indicator variable, so the 29 out of state TPPs were grouped into one category called "TPP Out of State." The remaining 10 indicator variables are labeled TPP A – TPP J. Table 1 lists summary statistics for the college variables. The largest group of teachers attended Reference TPP, which is used as the comparison group in the analyses.

Additional teacher variables are included in the models to control for demographics, college performance, and experience. Indicator variables designate teachers' gender and race. As is the case with students, male and European American teachers are used as the reference categories. Teachers' ACT composite scores are included to control for pre-TPP achievement. The individual ACT scores were available to the authors for about half of the teachers. In the case of the other half, the mean ACT composite score accepted at that TPP was substituted.

Taking advantage of the rich teacher data, controls are included for teachers' overall college GPA and years of experience, both measured as continuous variables. To account for a possible non-linear relationship between student test scores and teacher experience, models often include two variables to capture experience - years of experience and experience squared. However, there is no evidence of a non-linear effect of experience on student achievement in these data, so the squared term is not used in the model. In this sample, an overwhelming majority of teachers in the sample are European American and female. On average, teachers have about 13.38 years of experience, and have a 2.946 GPA upon graduation.

Many researchers agree that the composition of students in a classroom has implications for student learning, especially for certain groups of students (Hoxby, 2001; Author, 2000). To account for classroom composition, the models also include variables that control for classroom characteristics. These variables include the averages of all the student characteristics in the classroom, as well as their mean test scores in the prior year.

RESULTS

Table 2 presents estimated correlations of TPPs with 5th grade math scores, while incorporating the complete set of student, teacher, and classroom controls (equation 1). Consistent with the main hypothesis of this study, the table shows that TPPs vary in the effectiveness of the teachers they prepare, as measured by 5th grade math achievement. Graduates of TPP B and TPP C are less effective, in terms of helping students perform highly on the 5th grade KCCT math test, than graduates of the reference TPP. The table also presents marginal evidence suggesting that graduates from TPP F are more effective than graduates of the reference TPP.

Not all TPPs demonstrate a statistically significant effect on 5th grade math scores implying that they do not differ significantly from the reference TPP. Table 2 raises an additional interesting possibility. It provides evidence that suggests further exploration of the role of experience in the relationship between TPP and test scores. Table 3 presents coefficients and standard errors for the base terms and the interaction terms. The addition of interaction terms to the model substantially alters the interpretation of the relevant coefficients. In the previous analysis, the coefficients on TPPs are simply interpreted as the effect of TPP on 5th grade math scores relative to the reference TPP. This interpretation is no longer valid in the current analysis since the interaction indicates that the effect of TPP on the outcome variable varies according to teachers' years of experience. The interaction coefficients indicate whether teachers are increasingly or decreasingly effective as they gain experience, relative to what occurs over time for the teachers that graduated from the reference TPP.

The estimated coefficients for the statistically significant interaction terms in Table 3 indicate that teachers' relative effectiveness diminishes as they gain experience in the case of five TPPs (A, B, F, H and out of state) in comparison to the reference TPP. The base coefficients for these TPPs are all positive, while the interaction terms are negative. This suggests that these teachers are initially more effective than those from the reference category, but this effect decreases as the teachers gain experience. With each additional year, the effectiveness of these teachers approaches the effectiveness of the teachers from the reference TPP. The table also shows that there are three statistically significant interaction terms that are positive (E, G, and Other). In each of these cases, the base TPP coefficients are negative and large in magnitude. This suggests that in relation to the reference TPP, these teachers are much less effective in the

5th grade math classroom initially, but they make rapid increases in their effectiveness as they gain experience.

The unique effect of TPP on math scores incorporates not only the base coefficient of TPP, but also the experience coefficient and the interaction coefficient. Joint tests of hypotheses must be conducted to determine if the suite of variables containing the interaction term is jointly equal to zero instead of the more common case that concludes whether an individual coefficient is equal to zero. These tests will determine whether TPP has a statistically significant effect on 5th grade math achievement when taking into account the joint relationship with experience. Table 4 lists the p-values from the F-tests of joint significance and reveals that seven TPPs (A, E, F, G, H, Other, and out of state) have statistical relationships with the outcome variable. The remainder of the paper focuses only on TPPs A, E, F, G, and H because there is limited utility in interpreting categories with multiple TPPs.

To create a visual representation of the effects of TPPs on 5th grade math scores over time, the authors compute the partial effect of each statistically significant TPP. The partial effect is calculated by first differentiating the equation with respect to the TPP of interest and then inserting interesting values of experience. The partial effect is calculated for the first five years of teaching for two main reasons. First, the effect of any given TPP is expected to be larger for new teachers and then erode as teachers draw upon the expertise of their colleagues and supervisors for curriculum, instruction, and behavioral concerns. Second, research indicates that the important classroom skill building occurs in the first five years of a teacher's career and subsequently tapers off (Rivkin et al., 2005). Figure 1 charts the effects of these five TPPs on 5th grade math scores over time.

With one exception, the TPP effects approach convergence around the five year mark, or shortly thereafter. TPPs A, F, and H graduate new teachers that are relatively more effective in their first year of teaching than the reference TPP, but by roughly year five, the teaching effectiveness of the graduates from these programs is quite similar. The average 5th grade KCCT math score of TPPs A, F, and H, fall within 0.307 standard deviations of each other. This is less than half of the estimated spread in the teachers' first year of teaching. Teachers from TPP G are initially less effective than teachers in the reference category; however, they improve slowly over the years. These teachers' student scores do not converge until about year 10 of teaching. Teachers from TPP E undergo the most extreme changes in efficiency. In the first year of teaching, these graduates are two standard deviations less effective than the reference category teachers, but they make rapid improvements. By year five, teachers from TPP E are as effective as teachers from TPPs F and H.

DISCUSSION

Taken as a whole, the findings of this study suggest that differential effects of TPPs can be seen in the performance of 5th grade math students. The analysis indicates that some Kentucky TPPs supply more effective teachers into this school district. Furthermore, experience modifies the relationship between TPP and student achievement, with the result that they become roughly equally effective around year five of teaching. Teachers that are less effective in comparison to the reference TPP improve in their teaching effectiveness over the years. The opposite occurs with the teachers that are more effective in comparison to the reference TPP teachers.

Understanding the unique effects of TPPs on student achievement is important for policies relating to the training of teachers. If training programs have no independent effects on a

teacher's classroom effectiveness, then state and federal efforts to increase student achievement should be directed primarily at identifying characteristics in individuals that correlate most strongly with student learning and encouraging individuals with these characteristics to enter the teaching profession. If, on the other hand, TPP effects dominate innate characteristics of teachers, then states should focus on identifying best practices from the most effective TPPs. If the key to placing the most effective teachers in classrooms is some combination of the previous two scenarios, then states must focus not only on selecting the best teachers into TPPs, but also on identifying the key practices that ensure later success in the classroom.

Since very little research is currently able to inform these policy questions, the present study is an important contribution to the research base. The main result corroborates the findings of Boyd et al., (2008), which detect variation across TPPs in the average effectiveness of the teachers they supply to the New York City schools. The secondary result that examines the joint relationship between experience and TPP also provides some support to Noell (2006), which finds differential TPP effects within the first three years of teaching, but not thereafter.

Despite the consistency of the findings of these three studies, research on the effectiveness of TPPs is still in its infancy and the results should be viewed with some caution. All three of these studies are based on state or regional data, which poses two challenges to the researcher. The first, limited generalizability of the findings is familiar to the researcher. Teacher selection, which is not specific to Kentucky, receives less attention in the literature. Specifically, the administrative data systems do not have the ability to track graduates that leave the state of Kentucky to begin their teaching careers. If the best (or the worst) graduating teachers leave the state in search of a teaching job, then these TPP estimates will be biased. Nationally-representative data, while difficult to collect, would mitigate these two challenges and provide

important contributions to the evaluation of TPPs. Even so, the results presented here, in concert with similar research being conducted in other regions of the country, provide strong indications that the learning undertaken at TPPs has subsequent impacts on student achievement.

References

Andrew, M. (1990). Differences between graduates of 4- and 5-year teacher programs. *Journal of Teacher Education, 41*(2), 45-51.

Andrew, M. & Schwab, R.L. (1995). Has reform in teacher education influenced teacher performance? An outcome assessment of graduates. *Action in Teacher Education, 17*,43-53.

Betts, J. R., Zau, A., & Rice, L. (2003). *Determinants of student achievement: New evidence from San Diego*. San Francisco, Calif: Public Policy Institute of California.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy, 1*(2), 176-215.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2008). Teacher preparation and student achievement. (NBER Working Paper 14314). Cambridge, MA: National Bureau of Economic Research. Retrieved November 31, 2008, from http://www.nber.org/papers/w14314.

Clotfelter, C., Ladd, H., and J. Vigdor. (2007). How and why do teacher credentials matter for student achievement? (NBER Working Paper 12828). Cambridge, MA: National Bureau

of Economic Research. Retrieved September 21, 2007, from http://www.nber.org/papers/w12828.

_____. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources, 61*(4), 778-820.

Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the Evidence. *Educational Evaluation and Policy Analysis, 23*(1), 57-77.

Darling-Hammond, L., Chung, R., & Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education*, *53*(4), 286-302.

Darling-Hammond, L., Holtzman, J., Gatlin, S., & Heilig, J. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. Working Paper. Stanford University.

Ehrenberg, R. & Brewer, D. (1994). Do school and teacher characteristics matter? Evidence from High School & Beyond. *Economics of Education Review, 14*(1), 1-17.

Floden, R., & Meniketti, M. (2005). Research on the effects of coursework in the Arts and Sciences and in the Foundation of Education. In M. Cochran-Smith and K. Zeichner

(Eds.), Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education. Lawrence Erlbaum Associates, Inc.: New Jersey.

Goldhaber, D. & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, *22*(2), 129-145.

Good, T., McCaslin, M., Tsang, H., Zhang, J., Wiley, C., Rabidue Bozack, A., & Hester, W. (2006). How well do 1st year teachers teach? Does type of preparation make a difference? *Journal of Teacher Education, 57*(4), 410-430.

Grossman, P L. (1989). Learning to teach without teacher education. *Teachers College Record*, *91*(2), 191-208.

Hawk, P., Coble, C. R., & Swanson, M. (1985). Certification: It does matter. *Journal of Teacher Education, 36*(3), 13-15.

Hawk, B., & M. Schmidt. (1989). Teacher Preparation: A Comparison of Traditional and Alternative Programs. *Journal of Teacher Education, 40*(5), 53-58.

Houston, W.R., Marshall, F., & McDavid, T. (1993). Problems of traditionally prepared and alternatively certified first-year teachers. *Education and Urban Society, 26*(1), 78-89.

Jelmberg, J. (1996). College-based teacher education versus state-sponsored alternative programs. *Journal of Teacher Education, 47*(1), 60-66.

Author (2009).

Laczko-Kerr, I., & Berliner, D.C.. (2002, September 6). The effectiveness of "Teach for America" and other under-certified teachers on student academic achievement: A case of harmful public policy," *Education Policy Analysis Archives*, *10*(37). Retrieved October 2, 2007 from http://epaa.asu.edu/epaa/v10n37/.

Murnane, R. & Phillips, B. (1981). What do effective teachers of inner-city children have in common? *Social Science Research, 10*(1), 83-100.

Noell, G. (2006). Annual Report of Value added assessment of teacher preparation. *Unpublished manuscript*.

Noell, G. & Burns, J. (2006). Value-added assessment of teacher preparation: An illustration of emerging technology. *Journal of Teacher Education, 57*(1), 37-50.

Rivkin, S., Hanushek, E. & Kain, J.. (2005). Teachers, schools and academic achievement. *Econometrica, 73*(2), 417-458.

Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247-252.

Rothstein, J. (2008). Teacher quality in educational production: Tracking, decay, and student achievement. (NBER Working Paper 14442). Cambridge, MA: National Bureau of Economic Research. Retrieved November 31, 2008, from http://www.nber.org/papers/w14442.

_____. (2008). Student sorting and bias in value added estimation: Selection on observables and unobservables. Working Paper. Princeton University.

Summers, A., & Wolfe, B. (1977). Do schools make a difference? *The American Economic Review, 67*(4), 639-652.

Tiebout, C. (1956). A pure theory of local expenditures. *The Journal of Political Economy*, *64*(5), 416-424.

Author. (2000).

Table 1

Descriptive statistics comparing means and standard deviations (in parenthesis) of the study sample to those of all available data.

	Study Sample	All Available Data	T-test (Sig.)
Student Characteristics			
5th Math Score	0.107 (1.064)	0.115 (1.068)	
4th Reading Score	0.031 (0.861)	0.052 (0.853)	*
3rd Math Score	-0.130 (0.933)	-0.086 (0.922)	
% Female	0.501 (0.500)	0.501 (0.500)	
% European American	0.624 (0.484)	0.615 (0.487)	
% Asian American	0.015 (0.120)	0.014 (0.116)	
% African American	0.332 (0.471)	0.341 (0.474)	
% Latino/a	0.012 (0.109)	0.013 (0.111)	
% Other Race	0.018 (0.123)	0.017 (0.130)	
% Subsidized Lunch	0.553 (0.497)	0.546 (0.498)	
% IEP	0.079 (0.270)	0.078 (0.269)	
Teacher Characteristics			
% Female	0.879 (0.326)	0.868 (0.339)	
% European American	0.874 (0.332)	0.807 (0.395)	***
% African American	0.126 (0.332)	0.193 (0.395)	***
Years Experience	13.380 (8.234)	14.429 (8.341)	***
TPP A	0.081 (0.273)	0.057 (0.231)	***
TPP B	0.011 (0.104)	0.007 (0.083)	*
TPP C	0.097 (0.296)	0.064 (0.245)	***
TPP D	0.083 (0.276)	0.060 (0.238)	***
TPP E	0.022 (0.146)	0.015 (0.122)	**
TPP F	0.017 (0.128)	0.010 (0.101)	**
Reference TPP	0.334 (0.499)	0.304 (0.460)	***
TPP G	0.091 (0.287)	0.060 (0.237)	***
TPP H	0.012 (0.109)	0.008 (0.090)	*
TPP I	0.017 (0.128)	0.010 (0.101)	***
TPP J	0.035 (0.184)	0.022 (0.147)	***
TPP Other	0.016 (0.127)	0.010 (0.101)	**
TPP Out of State	0.180 (0.384)	0.354 (0.478)	***
ACT Score	21.823 (1.168)	21.831 (1.165)	
Overall GPA	2.946 (0.416)	2.910 (0.413)	***
N	2582	3714 – 4156	

Table 2

Estimates of TPP on students' fifth grade math achievement.

	Coefficient	Standard Error
Experience	0.005	(0.005)
TPP A	-0.161*	(0.095)
TPP B	-0.345***	(0.128)
TPP C	-0.278***	(0.069)
TPP D	-0.056	(0.118)
TPP E	-0.229	(0.151)
TPP F	0.277*	(0.163)
TPP G	-0.058	(0.234)
TPP H	-0.046	(0.137)
TPP I	-0.114	(0.131)
TPP J	0.087	(0.266)
Other	0.151	(0.100)
Out of State	0.136	(0.219)
R^2	0.5232	
N	2582	

All models contain controls for students (4^{th} grade test, 3^{rd} grade test, race, gender, lunch status, IEP status), teachers (ACT score, college GPA, gender, race), average classroom characteristics (4^{th} grade test, gender, race, lunch status, IEP status), and indicator variables for each school.

Table 3

Estimates of the joint relationship between experience and TPP on fifth grade students' math achievement.

	Coefficient	Standard Error
Experience (years)	0.012*	(0.006)
TPP A	0.339**	(0.170)
TPP B	0.462	(0.370)
TPP C	-0.174	(0.173)
TPP D	-0.445	(0.333)
TPP E	-2.007***	(0.535)
TPP F	0.601*	(0.361)
TPP G	-1.142***	(0.451)
TPP H	1.228***	(0.450)
TPP I	0.020	(0.253)
TPP J	0.489	(0.373)
Other	-6.717***	(1.240)
Out of State	0.517*	(0.283)
TPP A*Experience	-0.033***	(0.011)
TPP B*Experience	-0.036**	(0.016)
TPP C*Experience	-0.005	(0.010)
TPP D*Experience	0.012	(0.016)
TPP E*Experience	0.498***	(0.118)
TPP F*Experience	-0.030*	(0.018)
TPP G*Experience	0.107***	(0.037)
TPP H*Experience	-0.158***	(0.054)
TPP I*Experience	-0.070	(0.064)
TPP J*Experience	0.013	(0.026)
Other*Experience	0.334***	(0.062)

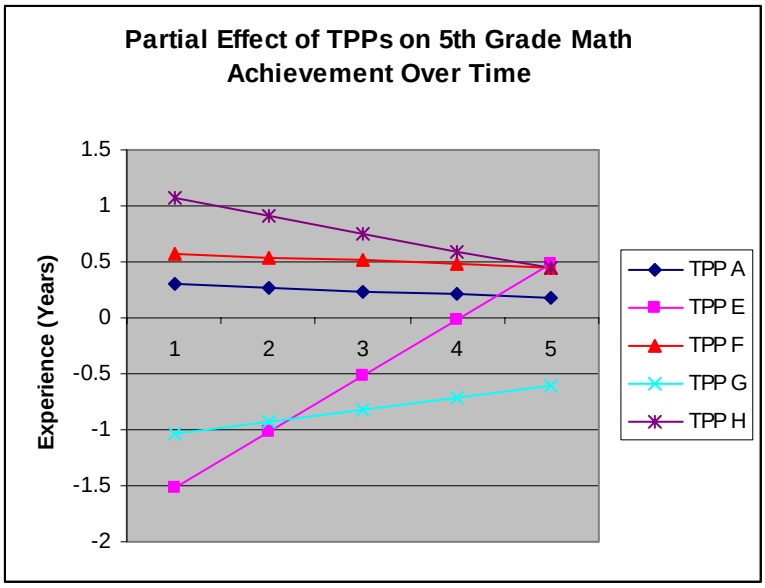
Out of state*Experience	-0.051**	(0.026)
R^2	0.5273	
N	2582	

All models contain controls for students (4^{th} grade test, 3^{rd} grade test, race, gender, lunch status, IEP status), teachers (ACT score, college GPA, experience, gender, race), average classroom characteristics (4^{th} grade test, gender, race, lunch status, IEP status), and indicator variables for each school.

Table 4

Joint tests of hypotheses of the effect of TPP and experience on fifth grade students' math achievement.

	p-value
Experience	0.002
TPP A	0.052
TPP B	0.221
TPP C	0.310
TPP D	0.190
TPP E	0.001
TPP F	0.094
TPP G	0.016
TPP H	0.008
TPP I	0.849
TPP J	0.149
Other	<0.001
Out of State	0.075


Partial Effect of TPPs on 5th Grade Math Achievement Over Time
Experience (Years)
1.5
1
0.5
0
-0.5
-1
-1.5
-2
1
2
3
4
5
TPP A
TPP E
TPP F
TPP G
TPP H

The IFIR Working Papers Series

Other titles in the series:

No. 2005-01: "MSA Location and the Impact of State Taxes on Employment and Population: A Comparison of Border and Interior MSA's," William H. Hoyt and J. William Harden.

No. 2005-02: "Estimating the Effect of Elite Communications on Public Opinion Using Instrumental Variables," Matthew Gabel and Kenneth Scheve.

No. 2005-03: "The Dynamics of Municipal Fiscal Adjustment," Thiess Buettner and David E. Wildasin.

No. 2005-04: "National Party Politics and Supranational Politics in the European Union: New Evidence from the European Parliament," Clifford J. Carrubba, Matthew Gabel, Lacey Murrah, Ryan Clough, Elizabeth Montgomery and Rebecca Schambach.

No. 2005-05: "Fiscal Competition," David E. Wildasin.

No. 2005-06: "Do Governments Sway European Court of Justice Decision-making?: Evidence from Government Court Briefs," Clifford J. Carrubba and Matthew Gabel.

No. 2005-07: "The Assignment and Division of the Tax Base in a System of Hierarchical Governments," William H. Hoyt.

No. 2005-08: "Global Competition for Mobile Resources: Implications for Equity, Efficiency, and Political Economy," David E. Wildasin.

No. 2006-01: "State Government Cash and In-kind Benefits: Intergovernmental Fiscal Transfers and Cross-Program Substitution," James Marton and David E. Wildasin.

No. 2006-02: "Decentralization and Electoral Accountability: Incentives, Separation, and Voter Welfare," Jean Hindriks and Ben Lockwood.

No. 2006-03: "Bureaucratic Advice and Political Governance," Robin Boadway and Motohiro Sato.

No. 2006-04: "A Theory of Vertical Fiscal Imbalance," Robin Boadway and Jean-Francois Tremblay.

No. 2006-05: "On the Theory and Practice of Fiscal Decentralization," Wallace E. Oates.

No. 2006-06: “The Impact of Thin-Capitalization Rules on Multinationals' Financing and Investment Decisions,” Thiess Buettner, Michael Overesch, Ulrich Schreiber, and Georg Wamser.

No. 2006-07: “Disasters: Issues for State and Federal Government Finances,” David E. Wildasin.

No. 2006-08: “Tax competition, location, and horizontal foreign direct investment,” Kristian Behrens and Pierre M. Picard.

No. 2006-09: “The effects of partisan alignment on the allocation of intergovernmental transfers. Differences-in-differences estimates for Spain,” Albert Solé-Ollé and Pilar Sorribas-Navarro.

No. 2006-10: “Reforming the taxation of Multijurisdictional Enterprises in Europe, "Coopetition" in a Bottom-up Federation,” Marcel Gerard.

No. 2006-11: “The Dilemmas of Tax Coordination in the Enlarged European Union,” Jens Brøchner, Jesper Jensen, Patrik Svensson, and Peter Birch Sørensen.

No. 2006-12: “Using a discontinuous grant rule to identify the effect of grants on local taxes and spending,” Matz Dahlberg, Eva Mörk, Jørn Rattsø, and Hanna Ågren.

No. 2006-13: “Size and Soft Budget Constraints,” Ernesto Crivelli and Klaas Staalz.

No. 2006-14: “On the Optimal Design of Disaster Insurance in a Federation,” Timothy Goodspeed and Andrew Haughwout.

No. 2006-15: “Fiscal Equalization and Yardstick Competition,” Christos Kotsogiannis and Robert Schwager.

No. 2007-01: “Disaster Policy in the US Federation: Intergovernmental Incentives and Institutional Reform,” David E. Wildasin.

No. 2007-02: “Local Government Finance in Kentucky: Time for Reform?” David E. Wildasin.

No. 2007-03: “Davis v. Department of Revenue of Kentucky: A Preliminary Impact Assessment,” Dwight Denison, Merl Hackbart, and Michael Moody.

No. 2007-04: “Medicaid Expenditures and State Budgets: Past, Present, and Future,” James Marton and David E. Wildasin.

No. 2007-05: “Pre-Emption: Federal Statutory Intervention in State Taxation,” David E. Wildasin.

No. 2007-06: "Think Locally, Act Locally: Spillovers, Spillbacks, and Efficient Decentralized Policymaking," Hikaru Ogawa and David E. Wildasin.

No. 2007-07: "Equalization Transfers and Dynamic Fiscal Adjustment: Results for German Municipalities and a US-German Comparison," Thiess Buettner.

No. 2008-01: "Civic Virtue, the American Founding, and Federalism," Stephen Lange.

No. 2008-02: "Public Finance in an Era of Global Demographic Change: Fertility Busts, Migration Booms, and Public Policy," David E. Wildasin.

No. 2009-01: "Decentralized Tax and Public Service Policies with Differential Mobility of Residents," William H. Hoyt.

No. 2009-02: "Business Incentives and Employment: What Incentives Work and Where?" William H. Hoyt, Christopher Jepsen and Kenneth R. Troske.

No. 2009-03: "Tax Limits, Houses, and Schools: Seemingly Unrelated and Offsetting Effects," William H. Hoyt, Paul A.Coomes and Amelia M. Biehl.

No. 2009-04: "The Taxpayer Relief Act of 1997 and Homeownership: Is Smaller Now Better?" Amelia M. Biehl and William H. Hoyt.

No. 2009-05: "Is the Grass Greener on the Other Side of the River?: The Choice of Where to Work and Where to Live for Movers," Ken Sanford and William H. Hoyt.

No. 2009-06: "The Value of a College Education: Estimating the Effect of Teacher Preparation on Student Achievement," Sharon Kukla-Acevedo and Eugenia F. Toma.

IFIR Working Papers contain original research contributed by scholars affiliated with the Institute for Federalism and Intergovernmental Relations at the Martin School of Public Policy and Administration at the University of Kentucky, Lexington, Kentucky. Visit the IFIR web site at http://www.ifigr.org to download IFIR Working Papers and for other information about IFIR.