Institute for Federalism & Intergovernmental Relations

IFIR WORKING PAPER SERIES

Evaluation of Teacher Preparation Programs: A Reality Show in Kentucky

Sharon Kukla-Acevedo, Ph.D.
Megan Streams, Ph.D.
Eugenia F. Toma, Ph.D.

IFIR Working Paper No. 2009-09

Evaluation of Teacher Preparation Programs: A Reality Show in Kentucky

Abstract

Title II of the *Higher Education Act* requires states to evaluate their teacher preparation programs (TPPs). In response, many states have introduced measures to evaluate TPPs similar to the ways in which they are evaluating K-12 schools. Some states, including Kentucky, have initiated pilot projects to assess the feasibility of statewide TPP evaluations. This paper stems from the Kentucky initiative and addresses methodological and data issues raised by the efforts to evaluate teacher preparation programs. This paper identifies some of the conceptual and empirical challenges of TPP evaluations. The purpose of this exercise is to serve as a model of learning for scholars interested in TPP evaluation and for policymakers and practitioners who are considering similar types of evaluations for their states.

KEY WORDS: Teacher preparation, evaluation, student achievement

Sharon Kukla-Acevedo, Ph.D. *
Central Michigan University
Department of Political Science
247 Anspach Hall
Mount Pleasant, MI 48859, USA
kukla1sa@cmich.edu

Megan Streams, Ph.D.
Tennessee State University
College of Public Service and Urban Affairs
330 N 10th Ave., PO Box 139
Nashville, TN 37203, USA
mstreams@tnstate.edu

Eugenia F. Toma, Ph.D.
University of Kentucky
Martin School of Public Policy & Administration
415 Patterson Office Tower
Lexington, KY 40506, USA
eugenia.toma@uky.edu

*Corresponding author

Evaluation of Teacher Preparation Programs: A Reality Show in Kentucky

INTRODUCTION

Most people link *No Child Left Behind (NCLB)* to the intense testing of students in K-12 schools. Increased accountability of K-12 schools is the hallmark of the federal legislation. But while less known to the public, another piece of legislation may be equally important to states and education policymakers. 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.

As a result of Title II, many states have introduced measures to evaluate teacher preparation programs (TPPs) much as they are evaluating K-12 schools. While some states such as Louisiana and Massachusetts have already begun statewide initiatives to address the Title II requirements, other states, including Kentucky, have initiated pilot projects to assess the feasibility of statewide TPP evaluations. This paper stems from the Kentucky initiative and addresses methodological and data issues raised by the efforts to evaluate teacher preparation programs. Administrative data from a sample of anonymous school districts within the state illustrate the challenges of this type of evaluation. The districts are both urban and rural and reveal distinct issues. This paper identifies some of the conceptual challenges of TPP evaluations and uses real-time data to illustrate issues that may have low-cost solutions, as well as those that cannot be so easily resolved. The purpose of this exercise is to serve as a model of learning for scholars interested in TPP evaluation and for policymakers and practitioners who are considering similar types of evaluations for their states. This pilot project brought a dose of reality to an abstract goal in Kentucky.

PREVIOUS WORK

Evaluation of education production has been extensive over the past thirty years. The effect of individual inputs into K-12 production have been examined to such a degree that an entire academic subdiscipline has arisen. The effectiveness of teacher salaries, of class size, of school size, of overall expenditure levels, of the source of school revenues, and increasingly of various aspects of teacher quality on student learning have been analyzed and reviewed by academics and shared with the policy world. These evaluations have proven very useful as they questioned many default assumptions of those in the policy arena, such as overall per pupil spending as a source of school quality variance. The studies have revealed that teacher quality appears to be one aspect of education production that not only significantly affects student learning, but that also may be subject to policy manipulation.

While the initial studies on teachers identified teacher effects but not which dimensions of teacher quality matter (Rivkin, Hanushek, and Kain, 2005), increasingly the K-12 scholars are attempting to identify specific quantifiable elements of quality that result in increased student learning (Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J., 2006; Clotfelter, C., Ladd, H., and J. Vigdor, 2007). It should not be surprising that one element of teacher quality under examination is the training received by the teacher. In other words, can TPPs influence the quality of teaching or is teacher quality independent of pre-service training? While studies about alternative credentials and advanced credentials such as requirements for masters' degrees have not been systematically linked to higher-quality teachers, the new research focus is to assess whether the earlier training in bachelors' programs yields differences in student learning.

Clearly, understanding the role of TPPs is important for policies relating to both the training and licensure of teachers. There are at least three possible answers to the training question: (1) the innate characteristics of teachers determine their success in the classroom, and the training program itself has no independent effects on a teacher's success in the classroom – here observed differences in the effectiveness of students from different TPPs would relate to differential TPP selection practices or applicant pools at those TPPs (TPP as screen); (2) innate characteristics of teachers are completely dominated by the training received in the bachelor's TPP (TPP instructional variation matters); and, finally, (3) innate characteristics are important but TPPs also have an independent and significant effect on student learning (both TPP selection practices and instruction variation matter).

Policy implications differ depending on the answer. If the answer is either the extremes above, drastic changes in current training methods would be in order. For example, if the answer is (1) above, then state efforts should be directed toward identifying those characteristics in individuals and toward encouraging those persons to enter the teaching profession in a less costly way than now exists. On the other hand, (2) implies that weak entrance requirements for TPPs are not a problem, and that states' efforts should be focused on identifying spreading key best practices from effective TPPs to others, or on closing ineffective TPPs altogether. Finally, if the answer is (3), both an identification of the innate characteristics of individuals who select into TPPs and an identification of specific TPP attributes that contribute to student learning are necessary.

Noell (2006) and Boyd, Grossman, Lankford, Loeb, and Wyckoff (2008) represent this emerging evaluation effort of TPPs. Boyd, et al. (2008) provide a high-resolution look at differences in teacher preparation programs in New York State and document that curriculum and training methods do differ across programs. The authors find some evidence that these training differences lead to student learning differences. Noell (2006) also points to evidence of differences in training programs in Louisiana.

As noted in the introduction, the current paper addresses this important issue by describing an evaluation effort for TPPs within a single state, Kentucky. The evaluation focuses on the bachelor degree programs and does not attempt to include alternative teacher pathways or master's degree programs. Many of the lessons learned in this exercise will also apply to evaluation of the alternative pathways to teaching. The steps followed by Kentucky researchers will be useful information for other states attempting to perform similar evaluations.

CONCEPTUAL MODEL AND DATA

The remainder of this paper addresses dual goals. A first goal is to assess the feasibility of introducing a TPP evaluation system in a real-world context. As such, the paper addresses the conceptual question of whether more and less successful teacher preparation programs can be identified. In other words, do teachers from some programs have greater classroom success than teachers from other TPP programs? Success in this case is measured strictly by student scores on state-level standardized test scores. A second goal is to use this study as a guide for other states that also are considering TPP evaluations by highlighting via empirical example the data challenges that arise in these evaluations.

Of particular interest at the outset, the State's teacher licensing agency, the Kentucky Education Professional Standards Board, recognizes 30 institutions of higher education with teacher training programs in a state with approximately 700,000 K-12 students. Among these programs, the vast majority are small, private institutions - each producing a few education graduates per year - and eight are publicly funded institutions with substantial numbers of graduates annually. Even more to the point, 4 public regional comprehensive universities that began as teachers' colleges continue to dominate the market in producing teachers for the entire state. For example, Figure 1 illustrates the market for graduates of a regional institution, Eastern Kentucky University,[1] and the geographical concentration of its graduates. Figure 2 illustrates the dominance of the few institutions across the entire state although as can be seen, individual county markets may be dominated by a small private institution. This regional segmentation of the market served by TPPs has important implications for the evaluation of their impact on student achievement, discussed in our results below.

Like most states, Kentucky's education administrative data serve multiple purposes and the data are not collected for research priorities. Rather, each type of data is collected and coded separately by a variety of divisions within the Kentucky Department of Education (KDE) and the Kentucky Education Professional Standards Board (EPSB). Figure 3 provides an overview of the different sources of administrative data. The data collected by EPSB, for example, deal only with teacher certification and licensure. These data are coded at the individual teacher level but each variable (years of experience, teacher rank, and teacher salary for example) is collected separately each year. The same applies to other data in the various administrative units. This arrangement does not appear to be unique to Kentucky and is the first reality check in statewide evaluations involving K-12 TPPs. States will face options of contracting out or creating research divisions within their organizational structures to collect and manage the databases necessary for these evaluations.

To illustrate the second challenge, consider the appropriate theoretical model that will be applied in a state's TPP evaluation, including in this pilot project. In a typical evaluation to

[1] Terry Hipbshman at the Kentucky Education Professional Standards Board kindly provided these maps.

estimate the effects of any intervention, pre-intervention outcomes data are compared to post-intervention, controlling for other factors that are expected to affect the outcome. The estimated coefficient on the intervention variable then signifies the magnitude of the effects of the program. Conceptually, the intervention model applied to pre-service college training is represented by the following:

$$(1)\quad A_{it} = \beta_0 A_{it-1} + \beta_1 TPP_{jt} + \beta_2 Stu_{it} + \beta_3 Tch_{ijmt} + Sch_{mt} + \lambda_m + u_{it}$$

where A_i and A_{it-1} are standardized student test scores;[2] TPP_{jt} is a vector of indicator variables describing the teacher's preparation program; Stu_{it} is a vector of student-specific characteristics, such as student ability, race, gender, subsidized lunch eligibility, and other family socioeconomic characteristics; Tch_{ijmt} includes teacher-specific characteristics, including gender, race, experience, college entrance scores, and college performance scores; and Sch_{mt} refers to time-varying characteristics of the school such as its resources, size, and student body characteristics. The subscripts denote students (i), teachers (j), schools (m) and time(t), while λ_m is a school fixed effect, and u_{it} is a randomly distributed error term. The fixed effects will capture unobservable time-invariant characteristics of the students, teachers, and schools. Of primary interest is the estimation of TPP, which can be interpreted as the impact of a particular teacher pre-service education on student scores, controlling for all other measurable factors that are expected to influence scores including student prior scores.

Note that this model requires observations for individual students, a match of those students to their individual teachers, observations of these teacher matches with multiple

[2] An alternative specification of this equation considers the dependent variable as the change in achievement between time periods *t* and *t*-1. The gain score is a more restrictive model than the one described above. See Hanuskek (1979) for more discussion.

students, observations of these matches over multiple teachers, and teachers from multiple TPP programs. The same students must be observed over at least two consecutive time periods so that a pre- and post- score can be calculated. If all factors that influence a student's score between periods in time are included in the empirical model, then any single estimated coefficient measures the contribution of a specific variable to the child's level of achievement in the same time period. From this perspective, it initially appears that measuring the value-added of TPP should be no different than the value added by any other policy intervention with the exception of the additional data requirements necessary for estimation of the effects of the TPP.

In Kentucky as 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. Rather, these matches can be made only by extracting classroom rolls from individual schools or districts.[3] 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 Education Professional Standards Board, three anonymous, randomly selected school districts agreed to provide their classroom rolls so that the teacher-student matches could be made.[4] To focus the project more specifically, this evaluation observes only those students in 11th grade math courses in these three districts.[5] The 11th grade math was chosen because it coincides with the Kentucky standardized testing schedule. In particular, all 11th graders participate annually in the Kentucky Core Content

[3] In some of the North Carolina studies, it has been assumed that the classroom teacher that administered the standardized test is the teacher of record for the student. In Kentucky, this is not the case. Teachers can request that certain students sit in their classrooms for test administration and those students are not necessarily current students of the respective teacher.

[4] The decentralized data storage proves particularly problematic in Kentucky where there are a total of 175 school districts within the state.

[5] Choosing a high school year proved more problematic than initially anticipated. Each of the three districts taught different math classes for 11th graders. As a result, the three districts are analyzed separately.

Test (KCCT). The data sample for this paper uses data from students who were in 11th grade in the 2005-2006 school year. Past standardized test results for these students are also available and are included in the analysis.

It is useful to examine the data in more detail. First, the three districts ranged in size from three to five high schools with our sample of 11th grade enrollments in math classes ranging from 771 to 1,699.[6] The aggregate number of math teachers in the high schools of these districts ranged from 23 to 67.[7] At the school level, the 11th grade student totals range from 188 to 440. The number of math teachers whose students took the test ranged from seven to 18 per high school. Table 1 illustrates that these three districts had roughly similar math performance on the KCCT exam in the 2005-2006 school year.

Students in District 1, the smallest of the three districts, look different than those in the other two districts. As illustrated in Table 2, the district has less than one percent each of black, Asian, and Hispanic students. More males than females had scores in our sample from District 1. None of these characteristics pose particular challenges for TPP evaluation.

But now consider the characteristics of the teachers. The average years of teacher experience varies by district. Noell (2006) and others find that TPP effects are greatest during the first three years of the teacher's classroom experience. After the third year, the TPP effects are diluted by the cohort or peer effects of the school in which the teacher is hired. These peer effects are likely to be larger in schools where the teaching staff has been in place for longer

[6] These were the students eligible to be used in further analysis. These totals reflect the omission of some students from the sample for the following reasons: their math teacher taught less than five students in 2005-2006 (73 students); they changed school or district during the year (1); they had either the minimum possible or maximum possible scaled score on the exam (extreme outlying values which are identified in residual diagnostics if included in regressions) (65); or their score on the KCCT 11th grade math test was missing (51). The N's of the Table 7 regressions are in turn lower than these totals due to missing data on at least one independent variable; Table 6 shows descriptive statistics for the students included in the final regression samples.

[7] Includes only teachers who taught more than 5 students in the sample.

periods of time. In the three districts examined, the average years of experience of high school math teachers ranges from 10 to 13 years (Table 3). Restricting the data to those teachers with experience of three years or less or, at most, to those with five years or less would cause a loss of more student observations for those schools with greater numbers of more experienced teachers. The average years of experience itself may be associated with the quality of the school, as teacher turnover is expected to be lower in schools with more amenities, such as a higher-performing student body. Thus one data choice that would make identifying TPP effects more likely – i.e., a focus on recent graduates – would entail (1) a trade-off with sample size, and (2) the risk of a school-level selection bias due to the non-random distribution of new teachers in schools from which those with more seniority tend to transfer.

The variation in time of graduation from a teacher preparation institution presents another complication to the interpretation of estimates of TPP effects, if we do not limit ourselves to recent graduates. Table 4 lists individually the three TPPs from which the largest numbers of teachers received their pre-service training and the years in which the training occurred. It groups all other Kentucky TPPs into a fourth category and then lists a fifth category for those teachers who received their undergraduate degree from a state outside of Kentucky. The data indicate that there are teachers in these districts who received degrees from TPP B as early as 1969 but as recently as 2005. If the policymakers' goal in identifying a TPP effect is to reward institutions that prepare better teachers, or to identify best curricular practices for diffusion to other institutions, one must make the assumption that the TPP effect represents some consistent education or (less optimistically) selection practice on the part of the institution. How realistic is the assumption that a given TPP's curricular or selection approach has remained constant over such a span of years? Yet if this assumption is not met, we are again left with the implication that analysis should be restricted to only recent graduates and the same caveats that result as detailed above.

Table 5 presents the issue from a slightly different perspective, illustrating the clustered distribution of TPP graduates across districts. Rather than grouping the teachers across all districts by institution of teacher training, Table 5 lists separately by district the number of teachers trained in the five categories of TPPs regardless of their graduation year. In District 1, for example, there are at least 21 teachers who teach one or more high school math classes and 16 of these teachers received teacher training from a Kentucky institution. Of these 16 Kentucky graduates, however, 11 were trained in the same TPP (TPP A). No high school math teachers on staff in this district in the study year were trained in TPP C. In District 2, on the other hand, ten of 23 state-trained teachers received credentials from TPP C but no teachers were trained in the dominant training institution for District 1, TPP A. As the data illustrate, District 3 predominantly hires its teachers from yet another program, TPP B, and did not have any high school math teachers who trained in TPP A in the study year.

The data from these three districts starkly illustrate an issue that has been raised in the teacher training literature. Teachers tend to enroll in training programs near their homes and to take jobs in districts near their institutions of training (Boyd, Lankford, and Loeb, 2005). For the districts in this sample, the market segmentation is sufficiently severe that an evaluation of TPPs effectively implies a comparison of only the two most common programs, or a single dominant program vs. all others present. But because these dominant programs with sufficient numbers of graduates are not the same across the three districts, each district must be analyzed separately.[8] It is expected that enlarging the sample to include data across all Kentucky school districts would

[8] This problem diminishes significantly in large urban areas with many teachers who trained at a variety of TPPs. Over 50 percent of school districts in the U.S. are rural and would be expected to share the stratification problems identified above.

expand the set of TPPs that could be evaluated but would not mitigate the cross district problem. The teacher stratification issue appears sufficiently severe to make evaluation of TPPs in rural areas of Kentucky using this type of approach infeasible from a practical perspective.

DISTRICT LEVEL RESULTS

Recognizing the deficiencies of the data as described throughout this paper, the value-added model with school fixed effects (equation 1 above) was estimated separately for each school district in the school year 2005-2006. Student test scores for end of period are their 11th grade math scores on the KCCT exam. All prior high school scores available for the students are included as controls including science and reading scores in addition to math.[9] The student characteristics include gender, race, special abilities status, and free and reduced lunch status. The teacher's gender and experience levels are included. All school characteristics are captured in the fixed effects.

As illustrated in Table 7, the results indicate quite clearly that a student's past scores in both reading and math (and science in the largest district) significantly influence 11th grade math scores.[10] The 10th grade reading and 9th grade math scores are significant and positively related to 11th grade math performance across all districts. In fact, prior student test scores are the only variables that are consistently statistically significant in explaining student outcomes across all three districts. The significance of the school fixed effects observed in two of the districts indicates strong school-level effects, even controlling for student and teacher characteristics.

Finally, and most important for our purposes, we consider the coefficient on the TPP variable. Recognizing that each district compares only its most common TPP to all others

[9] There is an emerging consensus that including as many prior scores as are available is the desired means of controlling for previous achievement. As will be indicated in the results, this analysis supports that consensus.
[10] See McCaffrey, Lockwood, Koretz, & Hamilton (2003) for an excellent discussion of the issues in value-added models.

represented by that district's high school math teachers, we see that there are no significant differences in student performance that can be attributed to the training institution. An alternate model analyzing only scores from students of recently graduated teachers whose ACT scores were available showed the same result.[11]

There is more than one reason that a TPP effect may not have been identified: at the most basic level, perhaps TPP does not matter once the other included student and teacher characteristics are accounted for. We cannot make such a determination with confidence, however, due to unresolved issues in the structure of these data: if a TPP effect truly exists, the inability to identify a significant effect could be due to the fact that all teachers are included, regardless of year in which they graduated. TPPs may have varied in their selection or curricular practices over time, as discussed above. This would make our TPP signal a very noisy one. Furthermore, the approach of comparing a district's most common TPP to all others is a weak one. If the TPPs in the combined reference group have strongly divergent effects on student achievement relative to the most common TPP, it would of course be difficult to show an effect of that TPP in this design. Nonetheless, this admittedly problematic type of comparison is necessary due to the dramatic segmentation of TPPs across districts and the relatively small numbers of high school math teachers from the non-dominant TPP in each district.

CONCLUDING COMMENTS

This paper uses a unique, rich data set from anonymous school districts in Kentucky that matches teacher and school characteristics to individual 11th grade math students for the 2005-

[11] The model could only be run for the largest district in order to have enough observations of recent graduates: ACT scores are available only for a subset of the District 3 teachers because Kentucky has only been saving these data in EPSB records since the 1990s. The District 3 student sample size was reduced from 1,137 to 370 by focusing on students of these teachers alone. In this regression, estimated effect of teacher ACT on student value added math score was positive and significant (p = 0.002), even controlling for all the variables listed in Table 7. The effect of dominant TPP relative to all others remained non-significant in this model.

2006 school year. Despite a robust methodology and rich student-teacher level data, limitations render this study unable to determine whether some teacher training programs are better at training teachers than others in rural areas of the state. The stratification of teachers makes it impossible to distinguish the lasting effects of the TPPs on a statewide basis. Some analysis at the individual school district is possible but biases in the estimates cannot be eliminated. The data requirements that would enable a successful statewide evaluation of TPP's are extensive. One issue we identify here is the possibility that TPP practices have changed over the range of graduation dates of teachers included in a value-added analysis. Restricting analysis to only recent graduates or graduate cohorts in an effort to ameliorate this issue would, in turn, cause a further reduction in sample size and could introduce selection bias due to the likely non-random distribution of seniority levels across schools with different achievement levels. Finally, the most problematic issue we have highlighted – which more data alone cannot solve – is the fact that the distribution of working teachers is not independent of the location of their respective teacher training programs. At least in Kentucky, schools tend to hire disproportionately from a single TPP. Because the factors that influence the hiring may be related to those that influence student achievement and cannot be separated with these limited data, the implications of the empirical work itself are quite limited. The implications of the attempt to perform this evaluation, on the other hand, are enormous. States will have to take a close look at their data management as well as their own stratification of teacher training and hiring if they are to engage in systematic evaluations of teacher training programs. While Title II requires evaluations of TPPs, scientifically rigorous evaluations may be premature in most states.

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. Jacob Fowles provided valuable assistance in manuscript preparation. We are grateful to participants of the 2008 meeting of the American Education Finance Association and the 2008 Kentucky Teacher Quality Conference for their useful comments.

Biographies

Sharon Kukla-Acevedo teaches in the Political Science Department at Central Michigan University. Her research focuses on education economics and policy in K-12 education. Current topics of interest include the evaluation of teacher effectiveness, evaluation of teacher preparation programs, teacher turnover and mobility, and improving educational outcomes for Latino/a students.

Megan Streams is an assistant professor of Public Administration in the College of Public Service and Urban Affairs at Tennessee State University. Her research interests are public finance and budgeting, with a policy interest in education; topics include the interaction of K-12 and higher education policies, impact of demographic trends on education finance, and teacher compensation and career trajectories. She has professional experience in higher education administration and budgeting.

Eugenia F. Toma is the Wendell H. Ford Professor of Public Policy in the Martin School of Public Policy & Administration at the University of Kentucky. She specializes in the economics of schooling. As a Fulbright Fellow to New Zealand in 1992, Toma worked with The Treasury and the Ministry of Education to analyze school reform. In 1998-99, she served as economics program director for the National Science Foundation (NSF) and participated in a multi-agency initiative to enhance scientific inquiry in education research. She currently is serving as PI of an NSF grant to evaluate the effectiveness of teacher professional development programs.

References

Boyd, D., Lankford, H., & Loeb, S. (2005). The Draw of Home: How Teachers' Preferences for Proximity Disadvantage Urban Schools. *Journal of Policy Analysis and Management*, *24,* 1, 113-132.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008) Teacher Preparation and Student Achievement. Working paper presented at the 2008 meeting of the American Education Finance Association, Denver, CO.

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, 176-216.

Clotfelter, C., Ladd, H., & Vigdor, J. (2007). How and why do teacher credentials matter for student achievement? National Bureau of Economic Research. Cambridge, MA: Working Paper #12828. Retrieved on September 21, 2007: http://www.nber.org/papers/w12828

Hanushek, E. (1979) Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, *14*, 351-388.

McCaffrey, D., Lockwood, J. R., Koretz, D., & Hamilton, L. (2003). *Evaluating value-added models for teacher accountability.* Santa Monica, CA: RAND.

Noell, G. (2006). 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*, 417-458.

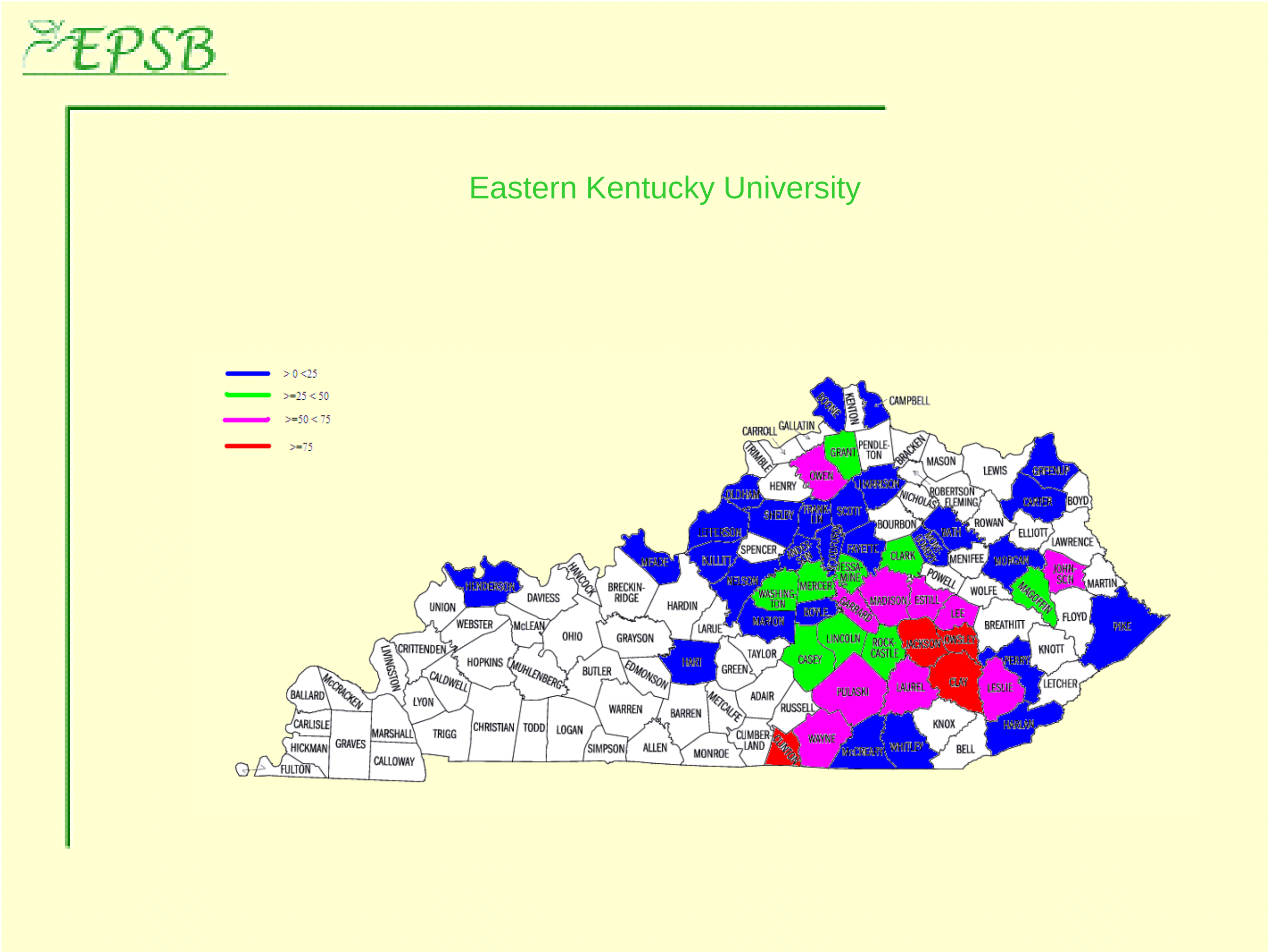

Figure 1. Teacher hiring patterns from a single teacher preparation program located in the southeast region of the state at a single point in time. Note the concentration of graduates in that region, with none at all in western Kentucky and many counties of eastern Kentucky. Figure courtesy of the Kentucky Education Professional Standards Board.

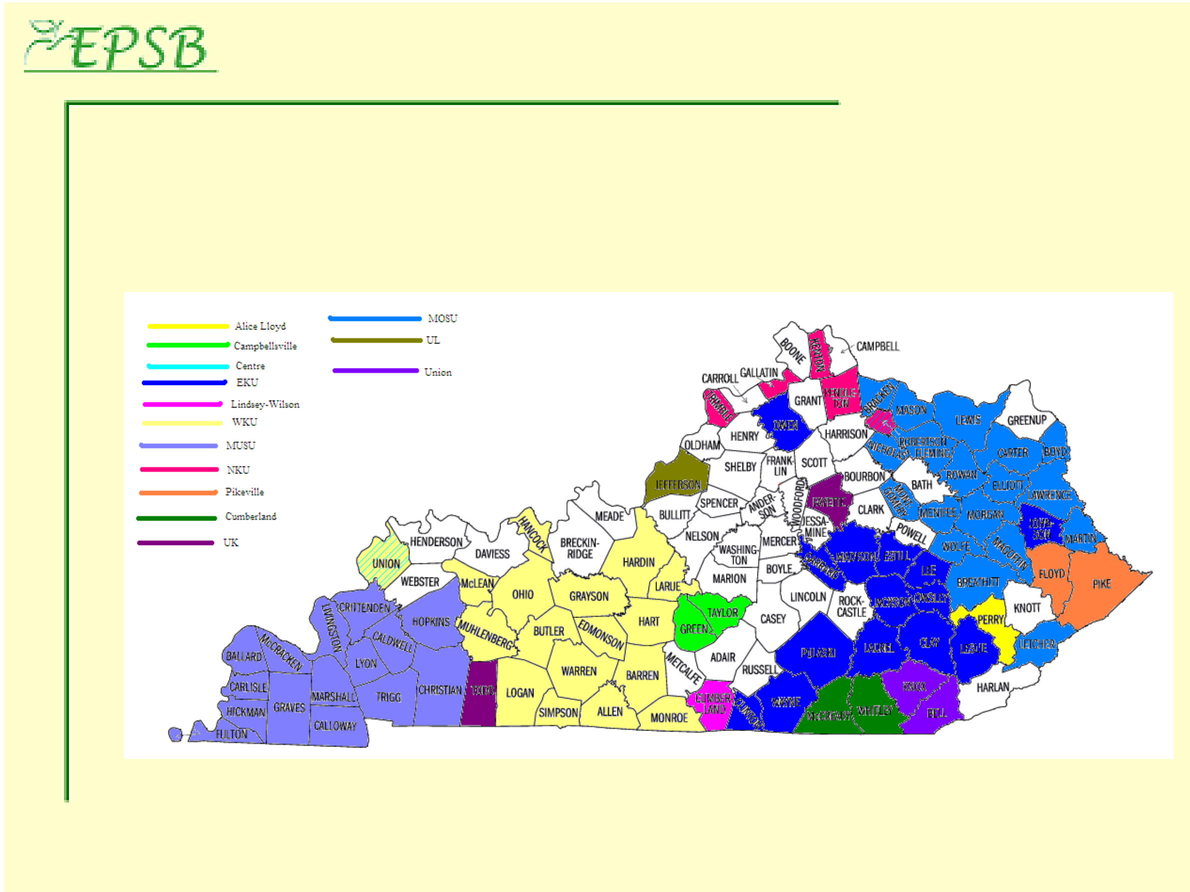

Figure 2: Segmentation of the Teacher Labor Market in Kentucky: Counties with at least 50 percent of teachers from a single teacher preparation program are shaded in. Figure courtesy of the Kentucky Education Professional Standards Board.

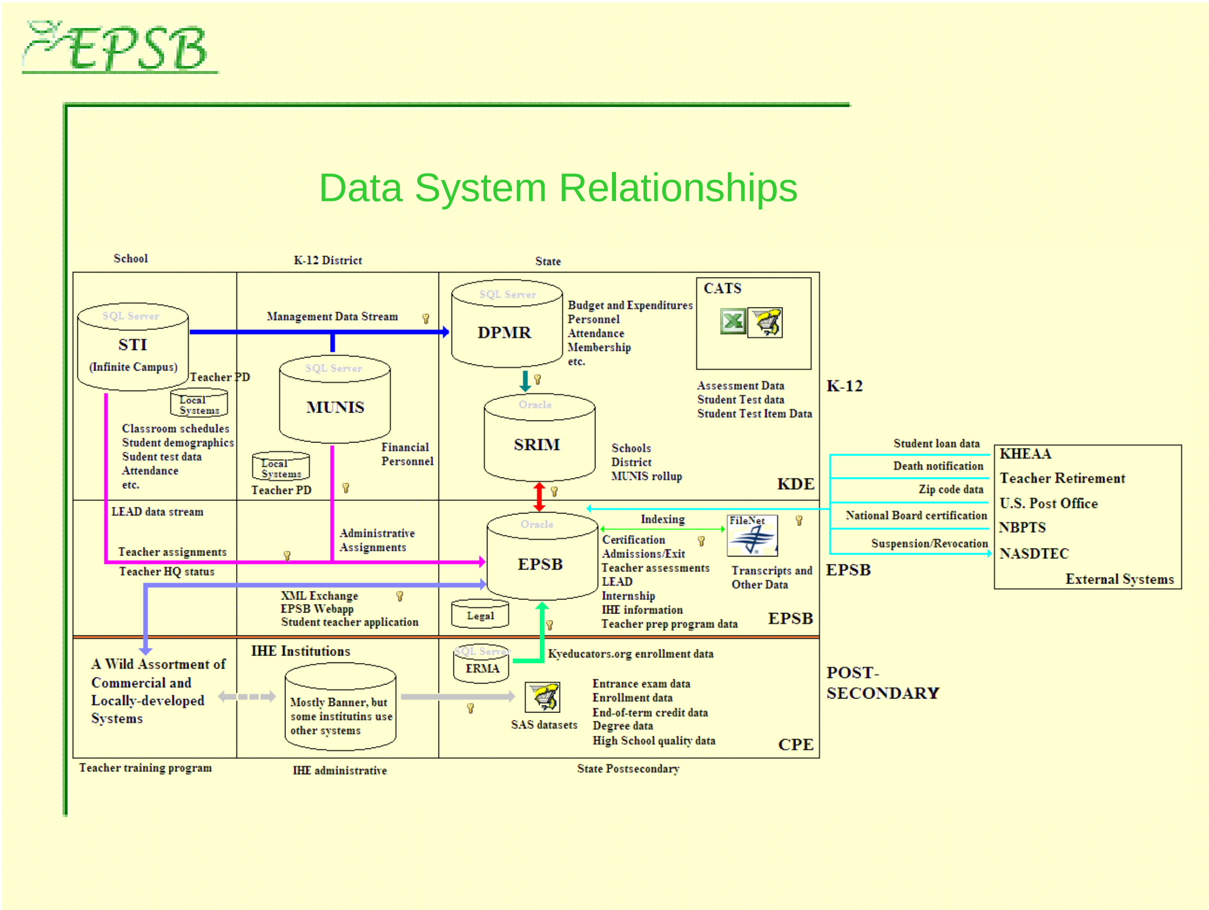

Figure 3: The sources of administrative K-12 educational data in Kentucky. Collection and storage of data in distinct administrative silos creates a challenge for longitudinal studies linking teachers and students that can investigate issues such as teacher preparation program impact on student achievement. Figure courtesy of the Kentucky Education Professional Standards Board.

Table 1

Average Scores of Students Taking the 11th Grade KCCT Math Test in Three Kentucky Districts, 2005-2006.

District	Mean KCCT 11th grade Math Score (standard deviation)	N
1	559 (42)	771
2	546 (43)	892
3	558 (49)	1,699

Table 2

Gender and Ethnicity of Students Taking the 11th Grade KCCT Math Test in Three Kentucky Districts, 2005-2006.

District	Female (%)	Black (%)	Asian (%)	Hispanic (%)	N
1	48.2	0.3	0.5	0.5	771
2	51.5	17.4	3.3	3.5	892
3	52.4	19.4	3.1	2.6	1,699

Table 3

Stratification of High School Math Teachers by Teacher Preparation Program (TPP) in Three Kentucky districts.

	Number of teachers from TPP / *Total number of 11th grade math students taught in sample*					
District	TPP A	TPP B	TPP C	TPP, Other KY	TPP, Other State	Unknown TPP
1	**11** / *347*	**1** / *56*	0	**4** / *111*	**5** / *196*	**2** / *61*
2	**0**	**2** / *143*	**10** / *135*	**11** / *408*	**5** / *206*	**0**
3	**0**	**33** / *871*	**1** / *30*	**17** / *372*	**12** / *327*	**4** / *93*

Note: TPPs A, B, and C are teacher preparation programs of three different Kentucky institutions of higher education, and are singled out because they were the most common TPP of high school math teachers in their respective districts in 2005-2006. "TPP, Other KY" indicates the number of graduates of all other Kentucky colleges and universities teaching 11th grade math in the district. "TPP, Other State" indicates the number of graduates from programs in other states. Teachers who taught five or fewer students with test scores in the sample are not shown. The number of students with 11th grade math test scores in the sample taught by the teachers from each type of TPP are shown in italics.

Table 4

Gender and Experience Level of High School Math Teachers Across Three Kentucky Districts, 2005-2006.

District	Female (%)	Mean Experience (years)	N (Teachers)	N (Schools)
1	60.9	13.1	23	3
2	75.0	9.9	28	3
3	67.2	13.1	67	5

Table 5

Range of Graduation Years of High School Math Teachers from Various TPPs Across Three Kentucky Districts, 2005-2006.

TPP	Year Bachelor's Degree Received	N (year of degree known)
A	1981-2004	7
B	1969-2005	16
C	1994-1999	3
Other KY	1974-2004	16
Other States	1967-2005	17

Note: The wide range of graduation years makes estimating the effect of a given teacher preparation program on student achievement challenging, if the assumption that preparation program quality and content remained constant over the period is not realistic.

Table 6

Variable Means for Final Sample Used in the Table 7 Regressions.

	District 1	District 2	District 3
Student Variables			
KCCT Grade 11 Math Score (scaled)	565 (43)	547 (44)	562 (48)
% Female	46.3	47.9	49.7
% Black [a]	0.4	16.1	16.5
% Gifted	33.3	22.9	23.4
% Disability	5.9	6.9	4.8
% Free/Reduced Lunch	14.9	25.9	18.5
CTBS Grade 9 Math Score (NCE) [b]	63 (19)	58 (19)	64 (21)
CTBS Grade 9 Science Score (scaled)	702 (31)	698 (30)	707 (34)
KCCT Grade 10 Reading Score (scaled)	530 (43)	531 (45)	544 (51)
Teacher Variables			
Years Experience	14.5 (11.5)	12.3 (8.5)	13.0 (10.0)
% Female	61.0	67.2	73.4
N (final regression sample)	21	28	63
% Teachers who graduated from most common TPP in district	52.4 %	35.7 %	52.4 %
% Students in regression sample from:			
School 1	30.4	27.7	27.9
School 2	28.7	33.0	8.3
School 3	40.9	39.4	22.5
School 4	--	--	19.7
School 5	--	--	21.6
N (number of student observations)	477	564	1,137

[a] Black student percentage was too low to justify use the use of this variable as a predictor in the regression.
[b] Scaled scores were unavailable for this test, so the normal curve equivalent values were used as the control variable instead. For this reason magnitude of the coefficients should not be compared between the 9th grade math score and the other two scores used as independent variables (9th grade science and 10th grade reading).
[c] Reference category for District 1.
[d] Reference category for District 2.
[e] Reference category for District 3.
Note: Sample size is reduced from the values in Table 1 due to observations missing data for independent variables included in the model.

Table 7

Individual Student Achievement Outcomes in Three Districts as a Function of Student Characteristics, 11th Grade Math Teacher Characteristics, TPP, and Test Score History.

	District 1	District 2	District 3
Student Variables			
% Female	-2.99	-4.21	-4.78 **
	(-1.23)	(-1.82)	(-2.97)
% Black	--	-5.62	-2.13
	--	(-1.77)	(-0.94)
% Gifted	9.83 ***	1.99	7.50 ***
	(3.52)	(0.63)	(3.72)
% Disability	0.23	-12.76 **	-8.36 *
	(0.04)	(-2.73)	(-2.20)
% Free/Reduced Lunch	1.46	0.39	-0.58
	(0.45)	(0.15)	(-0.26)
CTBS Grade 9 Math Score (NCE)	1.29 ***	1.18 ***	1.32 ***
	(14.86)	(14.70)	(24.33)
CTBS Grade 9 Science Score (scaled)	0.02	0.08	0.13 ***
	(0.35)	(1.52)	(4.10)
KCCT Grade 10 Reading Score (scaled)	0.23 ***	0.27 ***	0.18 ***
	(5.67)	(7.44)	(8.07)
Teacher Variables			
Years Experience	-0.05	0.07	0.04
	(-0.43)	(0.46)	(0.46)
% Female	3.58	5.83 *	3.84 *
	(1.28)	(2.22)	(2.22)
Most common TPP in district	-4.54	4.06	1.67
	(-1.63)	(1.15)	(1.07)
School 1	-10.79 ***	0.23	-5.56 *
	(-3.82)	(0.08)	(-2.55)
School 2	-1.58	2.54	-8.61 **
	(-0.49)	(0.84)	(-2.75)
School 3	--	--	-7.48 **
	--	--	(-3.27)
School 4	--	--	-0.97
	--	--	(-0.40)
Constant	352.37 ***	277.53 ***	288.40 ***
	(10.86)	(8.90)	(14.17)
N (number of student observations)	477	564	1,137
adjusted R^2	0.66	0.66	0.72

Note: *** p-value < 0.001; ** p-value < 0.01; * p-value < 0.05
T scores are shown in parentheses below coefficient estimates. Ordinary least squares regression is used with school fixed effects (shown; one school is the omitted category in each district).

The TPP indicator variable takes a 1 for students of teachers that graduated from the most common TPP of teachers in the district, and 0 otherwise. The coefficient indicates the effect of that TPP on student achievement relative to the effect of graduates of any other TPP present in the district, holding all else constant. Each model was significant in explaining variation in the dependent variable (District 1 model $F_{(12, 464)} = 77.98$, $p < 0.0001$; District 2 model $F_{(13, 550)} = 86.54$, $p < 0.0001$; District 3 model $F_{(15, 1,121)} = 199.68$, $p < 0.0001$).

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.

No. 2009-07: "State and Local Government Finance in the Current Crisis: Time for Emergency Federal Relief?" David E. Wildasin.

No. 2009-08: "State Corporation Income Taxation; An Economic Perspective on Nexus," David E. Wildasin.

No. 2009-09: "Evaluation of Teacher Preparation Programs: A Reality Show in Kentucky," Sharon Kukla-Acevedo, Megan Streams and Eugenia F. Toma.