We evaluate TPPs [teacher preparation programs] in Missouri using value-added models similar to those used in previous research studies (Boyd et al., 2009; Goldhaber et al., 2012) and at least two ongoing statewide evaluations (in Louisiana and Tennessee). The work we perform here is along the lines of what has been encouraged by the United States Department of Education (2011) and scholars from the Center for American Progress (Crowe, 2010) and Education Sector (Aldeman et al., 2011), among others. Moreover, all twelve Race-to-the-Top winners have committed to using achievement data for public disclosure of the effectiveness of TPP graduates, and five winners have committed to using teacher effects on student achievement for program accountability (Crowe, 2011). A key finding from our study, and one that we feel has not been properly highlighted in previous studies and reports, is that the measureable differences in effectiveness across teachers from different preparation programs are very small. The overwhelming majority of the variation in teacher quality occurs within programs. We encourage policymakers to think carefully about our findings as achievement-based evaluation systems, and associated accountability consequences, are being developed for TPPs.

Our study also adds to the body of evidence showing that it is difficult to identify which teachers will be most effective based on pre-entry characteristics, including TPP of attendance. That said, work by Boyd et al. (2009) suggests that variation within programs in terms of preparation experiences may be large. For example, Boyd et al. find that better oversight of student teaching for prospective teachers is positively associated with success in the classroom later on. The current research literature is too thin to fully understand how differences in within-program experiences affect teaching performance, but it would not be unreasonable, for example, to expect that within-program variability in the quality of training exceeds across-program variability (in the student-teaching oversight example, this would be the case if the attentiveness of prospective teachers’ mentors is unrelated to TPP of attendance).

Another area of inquiry that we investigated in some detail is with regard to initial selection into TPPs – we show that differential selection across teachers from different TPPs, based on ACT scores, is much smaller than what is implied by comparing the umbrella institutions that house the TPPs. That is, teachers who are trained at different TPPs are more similar to each other than the typical, non-teaching students who attend the universities where the TPPs are located. This may help to explain our substantive finding that there are not large differences in the effectiveness of graduates from different TPPs – the selection dimension does not appear to be as important among would-be teachers as would be expected based on institution-level differences in student selection across all fields of study.

We conclude by noting that our findings need not be interpreted to suggest that formal, outcome-based evaluations of TPPs should be abandoned. In fact, the lack of variability in TPP effects could partly reflect a general lack of innovation at TPPs, which is facilitated by the absence of a formal evaluation mechanism. The mere presence of an evaluation system, even if it is not immediately fruitful, may induce improvements
in teacher preparation that could improve students' short-term and long-term outcomes in meaningful ways (Chetty et al., 2011; Hanushek, 2011; Hanushek and Rivkin, 2010).

Still, we caution researchers and policymakers against overstating the present differences in TPP effects as statewide rankings become increasingly available. If administrators do not understand how small the differences in TPP effects really are, they could make poor hiring decisions by over weighting TPP rankings in their decisions. (22-23)

**Where You Come From or Where You Go? Distinguishing Between School Quality and the Effectiveness of Teacher Preparation Program Graduates**

Authors: Kata Mihaly, Daniel McCaffery, Tim R. Sass, and J. R. Lockwood

National Center for Analysis of Longitudinal Data in Education Research, Working Paper No. 63, January 2012


States like Florida that won the RTTT competition must provide measures of the performance of degree-granting teacher preparation programs in their states. One of the major concerns with such analyses is that program graduates may be teaching in very different contexts and those differences could be confounded with measures of the programs. This concern is exacerbated by the strong tendency for preparation program graduates to take jobs geographically close to the programs where they trained, potentially creating regional clusters of graduates. Models with school fixed effects would typically be seen as the best approach to removing potential confounding of context differences, because program estimates would rely on differences among student outcomes within the same schools to identify the program effects. However, such estimates may not be feasible if the training programs are not connected to each other. (18-19)

We found that the rankings of preparation programs based on relative effectiveness were significantly different when school fixed effects were included in the models. Regardless of the sample we used in the analysis (all teacher or only inexperienced teachers), we found that at least one preparation program switched rankings from the bottom quartile to the top quartile when school fixed effects were used. We observed that the rankings were more stable across specifications at the bottom of the ranking distribution than at the top, indicating that the use of student growth models may be more effective at capturing low performing programs than top tier programs. (20)

**Certification Requirements and Teacher Quality**

Author: Tim R. Sass

National Center for Analysis of Longitudinal Data in Education Research, Working Paper No. 64, December 2011


Abstract: Traditionally, states have required individuals complete a program of study in a university based teacher preparation program in order to be licensed to teach. In recent years, however, various “alternative certification” programs have been developed and the number of teachers obtaining teaching certificates through routes other than completing a traditional teacher preparation program has skyrocketed. In this paper I use a rich longitudinal data base from Florida to compare the characteristics of alternatively certified teachers with their traditionally prepared colleagues. I then analyze the relative effectiveness of teachers who enter the profession through different pathways by estimating “value-added” models of student achievement. In general, alternatively certified teachers have stronger pre-service qualifications than do traditionally prepared teachers, with the least restrictive alternative attracting the most qualified perspective teachers. These differences are less pronounced when controlling
for the grade level of teachers, however. On average, alternatively certified science teachers have also had much more coursework in science while in college than traditionally prepared science teachers. The same is not true for math teachers, where the hours of college coursework are approximately equal across pathways. Of the three alternative certification pathways studied, teachers who enter through the path requiring no coursework have substantially greater effects on student achievement than do either traditionally prepared teachers or alternative programs that require some formal coursework in education. These results suggest that the additional education coursework required in traditional teacher preparation programs either does little to boost the human capital of teachers or that whatever gains accrue from traditional teacher education training are offset by greater innate ability of individuals who enter teaching through routes requiring little formal training in education.

**Teacher Preparation and Student Achievement**

Authors: Donald Boyd, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff

National Center for Analysis of Longitudinal Data in Education Research, Working Paper No. 20, August 2008


This study of teacher education observes programs that prepare teachers for New York City (NYC) schools from what we might term an aerial perspective (c.f. Boyd, et al., 2006). Such a vantage point has its obvious disadvantages, particularly when it comes to portraying nuances of individual programs. Our goal, however, is to develop a broader picture of the terrain of teacher education in a single, large district, portraying, in general, how teachers are prepared to teach in NYC public schools and how variation in this preparation affects student learning. (3)

The difference between the average of the institutions and the highest value-added institution is approximately 0.05 standard deviations in math and 0.04 standard deviations in ELA. This magnitude is about the same size as difference in average learning between students eligible for free- or reduced-price lunch and those who are not. It is also about the same size as the difference in effectiveness between first-year and second-year teachers. Second, the variation in average teacher effectiveness across institutions is approximately the same in math and ELA. Finally, on average, institutions that produce teachers who are more effective at increasing student learning in math are also more effective in ELA (correlation of 0.60). (20-21)

The results also suggest that features of teacher preparation can make a difference in outcomes for students. One factor stands out. Teacher preparation that focuses more on the work of the classroom and provides opportunities for teachers to study what they will be doing produces teachers who are more effective during their first year of teaching. This finding holds up across various model specifications and both for measures created from data on the requirements of programs and for measures created from surveys of teachers. Thus, similar measures created from two independent data collection efforts reach a shared conclusion.

As an example, programs that provide more oversight of student teaching experiences or require a capstone project supply significantly more effective first-year teachers to New York City schools. Teachers who have had the opportunity in their preparation to engage in the actual practices involved in teaching (e.g., listening to a child read aloud for the purpose of assessment, planning a guided reading lesson, or analyzing student math work) also show greater student gains during their first year of teaching. Similarly, teachers who have had the opportunity to review curriculum used in New York City perform better in terms of student test score gains in both math and ELA. Student teaching and the congruence of the student
teaching placement are also positively associated with student learning in ELA and math, for first-year teachers. (26-27)

We also find some support for the hypothesis that math content preparation improves the outcomes of students of second-year teachers, but not first-year teachers. This result is supported by statistically significant and meaningful estimates across the measures created from the program requirements and from the teacher surveys, but the effects in some specifications are estimated imprecisely. Taken with the findings on the actual work of teachers, these estimates suggest that inexperienced teachers may make use of their preparation sequentially. Teachers with stronger preparation in day-to-day issues are relatively more effective in their first year, while those with stronger content knowledge are able to make use of that knowledge by their second year.

Finally, we fail to find consistent support for any of our other teacher preparation hypotheses. For example, our results do not support the hypothesis that greater opportunities to learn how students learn influences student achievement among first-year or second-year teachers. (27)

Everyone’s Doing It, But What Does Teacher Testing Tell Us About Teacher Effectiveness?
Author: Dan Goldhaber
National Center for Analysis of Longitudinal Data in Education Research, Working Paper No. 9, April 2007
http://www.caldercenter.org/PDF/1001072_everyones_doing.PDF

I find a positive relationship between some teacher licensure tests and student achievement; however, the point estimates are generally small in the specifications that account for the nonrandom sorting of teachers across students. This suggests that states face significant tradeoffs when using these tests as a screening device: despite the testing, many teachers whom we might wish were not in the teacher workforce based on their contribution toward student achievement are nevertheless eligible because they scored well on their test. Conversely, many individuals who would be effective teachers are ineligible due to their poor test performance. However, this does not necessarily suggest a diminished value of these tests: it is a value judgment as to whether the tradeoffs are worthwhile, and individual teacher test performance may provide an important signal about teacher quality that local hiring authorities could weigh against other teacher attributes when making hiring decisions. (2-3)

While it is interesting to look at the average effects of teacher testing as a signal, it is also informative to determine the distributional consequences of using these tests to determine employment eligibility. It is here that it becomes clear that teacher testing is not without its costs. Specifically, because the point estimates do not provide evidence of a terribly strong relationship between teacher test performance and student achievement, there are likely to be a significant number of false negatives (individuals who fail to achieve a minimum requirement on the licensure test but who would have been high-quality teachers) and false positives (individuals who do well on the licensure test but who are not very effective teachers). (29)

The research presented here suggests that licensure test performance is clearly not a ‘silver bullet’ credential that can be used to predict teacher effectiveness. If anything, the findings speak to the need for districts to be selective when hiring teachers. A large body of empirical evidence suggests that credentials like teacher licensure provide only a weak signal of teacher quality. To the extent that hiring officials cannot a priori discern more subtle teacher attributes that predict effectiveness, they must consider policies designed to shape their labor forces once they have had the chance to observe the effects of teachers in the classroom. (31)