



Carnegie Foundation for the Advancement of Teaching

WHAT DO VALUE-ADDED MEASURES OF TEACHER PREPARATION PROGRAMS TELL US?

DAN GOLDHABER

CENTER FOR EDUCATION DATA & RESEARCH
UNIVERSITY OF WASHINGTON BOTHELL

CARNEGIE KNOWLEDGE NETWORK
What We Know Series:
Value-Added Methods and Applications

ATIL

ADVANCING TEACHING - IMPROVING LEARNING

HIGHLIGHTS

- Policymakers are increasingly adopting the use of student growth measures to measure the performance of teacher preparation programs.
- Value-added measures of teacher preparation programs may be able to tell us something about the effectiveness of a program's graduates, but they cannot readily distinguish between the pre-training talents of those who enter a program from the value of the training they receive.
- Research varies on the extent to which prep programs explain meaningful variation in teacher effectiveness. This may be explained by differences in methodologies or by differences in the programs.
- Research is only just beginning to assess the extent to which different features of teacher training, such as student selection and clinical experience, influence teacher effectiveness and career paths.
- We know almost nothing about how teacher preparation programs will respond to new accountability pressures.
- Value-added based assessments of teacher preparation programs may encourage deeper discussions about additional ways to rigorously assess these programs.

INTRODUCTION

Teacher training programs are increasingly being held under the microscope. Perhaps the most notable of recent calls to reform was the 2009 declaration by U.S. Education Secretary Arne Duncan that “by almost any standard, many if not most of the nation's 1,450 schools, colleges, and departments of education are doing a mediocre job of preparing teachers for the realities of the 21st century classroom.”¹ Duncan's indictment comes despite the fact that these programs require state approval and, often, professional accreditation. The problem is that the scrutiny generally takes the form of input measures, such as minimal requirements for the length of student teaching and assessments of a program's curriculum. Now the clear shift is toward measuring outcomes.

The federal Race to the Top initiative, for instance, considered whether states had a plan for linking student growth to teachers (and principals) and to link this information to the schools in that state that trained them. The new Council for the Accreditation of Educator Preparation (CAEP) just endorsed the use of student outcome measures to judge preparation programs.² And there is a new push to use changes in student test scores—a teachers' value-added—to assess teacher preparation providers (TPPs). Several states are already doing so or plan to soon.³

Much of the existing research on teacher preparation has focused on comparing teachers who enter the profession through different routes—traditional versus alternative certification—but more recently, researchers have turned their attention to assessing individual TPPs.⁴ And in doing so, researchers face many of the same statistical challenges that arise when value-added is used to evaluate individual teachers. The measures of effectiveness might be sensitive to the student test that is used, for instance, and statistical models might not fully distinguish a teacher's contributions to student learning from other factors.⁵ Other issues are distinct, such as how to account for the possibility that the effects of teacher training fade the longer a teacher is in the workforce.

WHAT DO VALUE-ADDED MEASURES OF TEACHER PREPARATION PROGRAMS TELL US?

Before detailing what we know about value-added assessments of TPPs, it is important to be explicit about what value-added methods can and cannot say about the quality of TPPs. Value-added methods *may* be able to tell us something about the effectiveness of a program's graduates, but this information is a function both of graduates' experiences in a program and of who they were when they entered. The point is worth emphasizing because policy discussions often treat TPP value-added as a reflection of training alone. Yet the students admitted to one program may be very different from those admitted to another. Stanford University's program, for instance, enrolls students who generally have stronger academic backgrounds than those of Fresno State University. So if a teacher who graduated from Stanford turns out to be more effective than a teacher from Fresno State, we should not necessarily assume that the training at Stanford is better.

Because we cannot disentangle the value of a candidate's selection to a program from her experience there, we need to think carefully about what we hope to learn from comparing TPPs. Some stakeholders may care about the value of the training itself, while others may care about the combined effects of selection and training. Many also are likely to be interested in outcomes other than those measured by value-added, such as the number of teachers, or types of teachers that graduate from different institutions. They may also want to know how likely it is that a graduate from a certain institution will actually find a job or stay in teaching. I elaborate on these issues below.

WHAT IS KNOWN ABOUT VALUE-ADDED MEASURES OF TEACHER PREPARATION PROGRAMS?

There are only a handful of studies that link the effectiveness of teachers, based on value-added, to the program in which they were trained.⁶ This body of research assesses the degree to which training programs explain the variation in teacher effectiveness, whether there are specific features of training that appear to be related to effectiveness, and what important statistical issues arise when we try to tie these measures of effectiveness back to a training program.

Empirical research reaches somewhat divergent conclusions about the extent to which training programs explain meaningful variation in teacher effectiveness. A study of teachers in New York City, for instance, concludes that the difference between teachers from programs that graduate teachers of average effectiveness and those whose teachers are the most effective is roughly comparable to the (regression-adjusted) achievement difference between students who are and are not eligible for subsidized lunch.⁷ Research on TPPs in Missouri, by contrast, finds only very small—and statistically insignificant—differences among the graduates of different programs.⁸ There is also similar work on training programs in other states: Florida,⁹ Louisiana,¹⁰ North Carolina,¹¹ and Washington.¹² The findings from these studies fall somewhere between those of New York City and Missouri in terms of the extent to which the colleges from which teachers graduate provide meaningful information about how effective their program graduates are as teachers.

One explanation for the different findings among states is the extent to which graduates from TPPs—specifically those who actually end up teaching—differ from each other.¹³ As noted, regulation of training programs is a state function, and states have different admission requirements. The more that institutions within a state differ in their policies for candidate selection and training, the more likely we are to see differences in the effectiveness of their graduates. There is relatively little quantitative research on the features of TPPs that are associated with student achievement, but what does exist offers *suggestive* evidence that some features may matter.¹⁴ There is some evidence, for instance, that

WHAT DO VALUE-ADDED MEASURES OF TEACHER PREPARATION PROGRAMS TELL US?

licensure tests, which are sometimes used to determine admission, are associated with teacher effectiveness.¹⁵

More recently, some education programs have begun to administer an exit assessment (the “edTPA”) designed to ensure that TPP graduates meet minimum standards. But because the edTPA is new, there is not yet any large-scale quantitative research that links performance on it with eventual teacher effectiveness. There is also some evidence that certain aspects of training are associated with a teacher’s later success (as measured by value-added) in the classroom. For example, the New York City research shows that teachers tend to be more effective when their student teaching has been well-supervised and aligned with methods coursework, and when the training program required a capstone project that related their clinical experience to training.¹⁶ Other work finds that trainees who student-teach in higher functioning schools (as measured by low attrition) turn out to be more effective once they enter their own classrooms.¹⁷ These studies are intriguing in that they suggest ways in which teacher training may be improved, but, as discussed below, it is difficult to tell definitively whether it is the training experiences themselves that influence effectiveness.

The differences in findings across states may also relate to the methodologies used to determine teacher-training effects. A number of statistical issues arise when we try to estimate these effects based on student achievement. One issue is that we are less sure, in a statistical sense, about the impacts of graduates from smaller programs than we are from graduates of larger ones. This is because the smaller sample of graduates gives us estimates of their effects that are less precise; it is virtually guaranteed that findings for smaller programs will not be statistically significant.¹⁸ Another issue is the extent to which programs should be judged in a particular year based on graduates from prior years. The effectiveness of teachers in service may not correspond with their effectiveness at the time they were trained, and it is certainly plausible that the impact of training a year after graduation would be different than it is five or 10 years later, when a teacher has acculturated to his particular school or district. Most of the studies cited above handle this issue by including only novice teachers—those with three or fewer years of experience—in their samples (exacerbating the issue of small sample sizes).¹⁹ Another reason states may limit their analysis to recent cohorts is that it is politically impractical to hold programs accountable for the effectiveness of teachers who graduated in the distant past, regardless of the statistical implications.

Limiting the sample of teachers used to draw inferences about training programs to novices is not the only option, however. Goldhaber et al., for instance, estimate statistical models that allow training program effects to diminish with the amount of workforce experience that teachers have.²⁰ These models allow more experienced teachers to contribute toward training program estimates, but they also allow for the possibility that the effect of pre-service training is not constant over a teacher’s career. As illustrated by Figure 1, regardless of whether the statistical model accounts for the selectivity of college (dashed line) or not (solid line), this research finds that the effects of training programs do decay. They estimate that their “half-life”—the point at which half the effect of training can no longer be detected—is about 13–16 years, depending on whether teachers are judged on students’ math or reading tests and on model specification.

A final statistical issue is how a model accounts for factors of school context that might be related both to student achievement and the districts and schools in which TPP graduates are employed. This is a challenging problem; one would not, for instance, want to misattribute district-led induction programs or principal-influenced school environments to TPPs.²¹ This issue of context is not fundamentally different from the one that arises when we try to evaluate the effectiveness of individual teachers. We

WHAT DO VALUE-ADDED MEASURES OF TEACHER PREPARATION PROGRAMS TELL US?

worry, for instance, that between-school comparisons of teachers may conflate the impact of teachers with that of, for instance, principals,²² but it is particularly problematic in the case of TPPs, when there is little mixing of TPP graduates within a school or school district. This situation may arise when TPPs serve a particular geographic area or espouse a mission aligned to the needs of a particular school district.

WHAT MORE NEEDS TO BE KNOWN ON THIS ISSUE?

As noted, there are a few studies that connect the features of teacher training to the effectiveness of teachers in the field, but this research is in its infancy. I would argue that we need to learn more about (1) the effectiveness of graduates who typically go on to teach at the high school level; (2) what features of TPPs seem to contribute to the differences we see between in-service graduates of different programs; (3) how TPPs respond to accountability pressures; and (4) how TPP graduates compare in outcomes other than value-added, such as the length of time that they stay in the profession or their willingness to teach in more challenging settings.

First, most of the evidence from the studies cited above is based on teaching at the elementary and middle school levels; we know very little about how graduates from different preparation programs compare at the high school level. And while much of the research and policy discussion treats each TPP as a single institution, the reality is much more complex. Programs produce teachers at both the undergraduate and graduate levels, and they train students to teach different subjects, grades, and specialties. It is conceivable that training effects in one area—such as undergraduate training for elementary teachers—correspond with those in other areas, such as graduate education for secondary teachers. But aside from research that shows a correlation between value-added and training effects across subjects, we do not know how much estimates of training effects from programs within an institution correspond with one another.²³

Second, we know even less about what goes on inside training programs, the criteria for recruitment and selection of candidates, and the features of training itself. The absence of this research is significant, given the argument that radical improvements in the teacher workforce are likely to be achieved only through a better understanding of the impacts of different kinds of training. There is certainly evidence that these programs differ from one another in their requirements for admission, in the timing and nature of student teaching, and in the courses in pedagogy and academic subjects that they require. But while program features may be assessed during the accreditation and program approval process, there is little data that can readily link these features to the outcomes of graduates who actually become teachers. Teacher prep programs are being judged on some of these criteria (e.g., NCTQ, 2013),²⁴ but we clearly need more research on the aspects of training that may matter. As I suggest above, it is challenging to disentangle the effects of program selection from training effects; there are certainly experimental designs that could separate the two (e.g., random assignment of prospective teachers to different types of training), but these are likely to be difficult to implement given political or institutional constraints.

Third, we will want to assess how TPPs respond to increased pressure for greater accountability.²⁵ Post-secondary institutions are notoriously resistant to change from within, but there is evidence that they do respond to outside pressure in the form of public rankings.²⁶ Rankings for training programs have just been published by the National Council on Teacher Quality (NCTQ) and *U.S. News & World Report*. Do programs change their candidate selection and training processes because of new accountability pressures? If so, can we see any the impact on the teachers that graduate from them? These are questions that future research could address.²⁷

WHAT DO VALUE-ADDED MEASURES OF TEACHER PREPARATION PROGRAMS TELL US?

Lastly, I would be remiss in not emphasizing the need to understand more about ways, aside from value-added estimates, that training programs influence teachers. Research, for instance, has just begun to assess the degree to which training programs or their particular features relate to outcomes as fundamental as the probability of a graduate's getting a teaching job²⁸ and of staying in the profession.²⁹ This line of research is important, given that policymakers care not only about the effectiveness of teachers but of their paths in and out of teaching careers.

WHAT CAN'T BE RESOLVED BY EMPIRICAL EVIDENCE ON THIS ISSUE?

Value-added based evaluation of TPPs can tell us a great deal about the degree to which differences in measureable effects on student test scores may be associated with the program from which teachers graduated. But there are a number of policy issues that empirical evidence will not resolve because they require us to make value judgments. First and foremost is whether or to what extent value-added should be used at all to evaluate TPPs.³⁰ This is both an empirical and philosophical issue, but it boils down to whether value-added is a true reflection of teacher performance,³¹ and whether student test scores should be used as a measure of teacher performance.³²

Even policymakers who believe that value-added should be used for making inferences about TPPs face several issues that require judgment. The first is what statistical approach to use to estimate TPP effects. For instance, as outlined above, estimating the statistical models requires us to make decisions about whether and how to separate the impact of TPP graduates from school and district environments. And how policymakers handle these statistical issues will influence estimates of how much we can statistically distinguish the value-added based TPP estimates from one program to another (i.e., the standard errors of the estimates and related confidence intervals). And the statistical judgments have potentially important implications for accountability. For instance, we see the 95 percent confidence level used in academic publications as a measure of statistical significance. But, as noted above, this standard likely results in small programs being judged as statistically indistinguishable from the average, an outcome that can create unintended consequences. Take, for instance, the case in which graduates from small TPP "A" are estimated to be far less effective than graduates from large TPP "B," who are themselves judged to be somewhat less effective than graduates from the average TPP in a state. Graduates from program "A" may not be statistically different from the mean level of teacher effectiveness, at the 95 percent confidence interval, but graduates from program "B" might be. If, as a consequence of meeting this 95 percent confidence threshold, accountability systems single out "B" but not "A," a dual message is sent. One message might be that a program ("B") needs improvement. The second message, to program "A," is "don't get larger unless you can improve," and to program "B," is "unless you can improve, you might want to graduate fewer prospective teachers." In other words, program "B" can rightly point out that it is being penalized when program "A" is not, precisely because it is a large producer of teachers.

Is this the right set of incentives? Research can show how the statistical approach affects the likelihood that programs are identified for some kind of action, which creates tradeoffs; some programs will be rightly identified while others will not. For more on the implications of such tradeoffs in the case of individual teacher value-added.³³ Research, however, cannot determine what confidence levels ought to be used for taking action, or what actions ought to be taken. Likewise, research can reveal more about whether TPPs with multiple programs graduate teachers of similar effectiveness, but it cannot speak to how, or whether, estimated effects of graduates from different programs within a single TPP should be aggregated to provide a summative measures of TPP performance. Combining different measures may

WHAT DO VALUE-ADDED MEASURES OF TEACHER PREPARATION PROGRAMS TELL US?

be useful, but the combination itself could mask important information about institutions. Ultimately, whether and how disparate measures of TPPs are combined is inherently a value judgment.

Lastly, it is conceivable that changes to teacher preparation programs—changes that affect the effectiveness of their graduates—could conflict with other objectives. A case in point is the diversity of the teacher workforce, a topic that clearly concerns policymakers and teacher prep programs themselves.³⁴ It is conceivable that changes to TPP admissions or graduation policies could affect the diversity of students who enter or graduate from TPPs, ultimately affecting the diversity of the teacher workforce.³⁵ Research could reveal potential tradeoffs inherent with different policy objectives, but not the right balance between them.

EVALUATING TPPS: PRACTICAL IMPLICATIONS AND CONCLUSIONS

It is surprising how little we know about the impact of TPPs on student outcomes given the important role these programs could play in determining who is selected into them and the nature of the training they receive. That's largely because, in some states, due to new longitudinal data systems, we only now have the capacity to systematically connect teacher trainees from their training programs to the workforce, and then to student outcomes. But we clearly need better data on the admissions and training policies if we are to better understand these links. That said, the existing empirical research does provide some specific guidance for thinking about value-added based TPP accountability. In particular, while it cannot definitively identify the right model to determine TPP effects, it does show how different models change the estimates of these effects and the estimated confidence in them.

To a large extent, how we use estimates of the effectiveness of graduates from TPPs depends on the role played by those who are using them. Individuals considering a training program likely want to know about the quality of training itself.³⁶ Administrators and professors probably also want information about the value of training to make programmatic improvements. State regulators might want to know more about the connections between TPP features and student outcomes to help shape policy decisions about training programs, such as whether to require a minimum GPA or test scores for admission or whether to mandate the hours of student teaching. But, as stressed above, researchers need more information than is typically available to disentangle the effects of candidate selection from training itself. Given the inherent difficulty of disentangling, one should be cautious about inferring too much about training based on value-added itself.

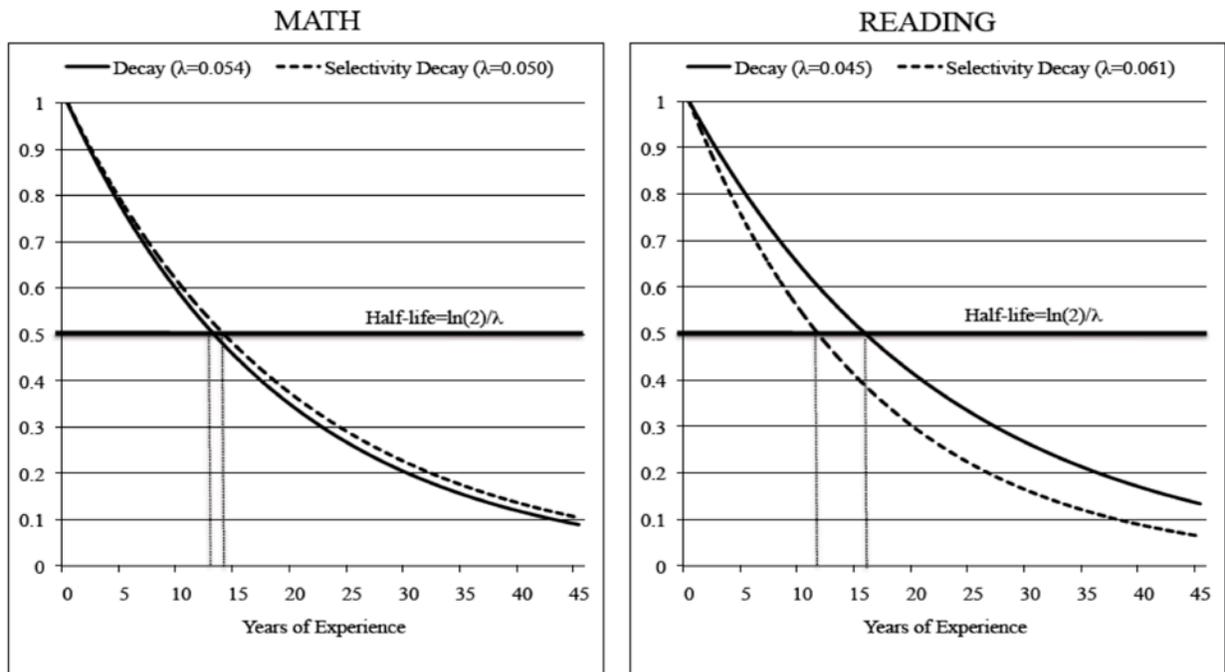
For other policy or practical purposes, however, it may be sufficient to know only the combined impact of selection and training. Principals or other district officials, for instance, likely wonder whether estimated TPP effects should be considered in hiring, but they may not care much about precisely what features of the training programs lead to differences in teacher effectiveness. Likewise, state regulators are charged with ensuring that prep program graduates meet a minimal quality standard; it is not their job to determine whether that standard is achieved by how the program selects its students or how it teaches them.

The bottom line is that while our knowledge about teacher training is now quite limited, there is a widespread belief that it could be much better. As Arthur Levine, former president of Teachers College, Columbia University, says, "Under the existing system of quality control, too many weak programs have achieved state approval and been granted accreditation."³⁷ The hope is that new value-added based assessments of teacher preparation programs will not only provide information about the value of

WHAT DO VALUE-ADDED MEASURES OF TEACHER PREPARATION PROGRAMS TELL US?

different training approaches, but also will encourage a much-needed and far deeper policy discussion about how to rigorously assess programs for training teachers.

Figure 1: Decay of program effect estimates in math and reading³⁸



ENDNOTES

¹ Duncan is hardly alone in criticizing education programs. A number of researchers and stakeholders have explicitly questioned the quality control of teacher training institutions, and, in some cases, the value of teacher training.

See, for instance: Ballou, D., & Podgursky, M. (2000). Reforming teacher preparation and licensing: What is the evidence? *The Teachers College Record*, 102(1), 5-27.

Cochran-Smith, M., & Zeichner, K. M. (Eds.). (2005). *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education*. Washington, D.C.: The American Educational Research Association.

Crowe, E. (2010). *Measuring what matters: A stronger accountability model for teacher education*. Washington, DC: Center for American Progress.

Levine, A. (2006). *Educating School Teachers*. Washington, D.C.: The Education Schools Project.

Vergari, S., & Hess, F. M. (2002). The Accreditation Game. *Education Next*, 2(3), 48-57.

Walsh, K., & Podgursky, M. (2001). Teacher certification reconsidered: Stumbling for quality. A rejoinder. *The Abell Foundation*.

² CAEP was formed by the merger of Teacher Education Accreditation Council (TEAC) and the National Council for the Accreditation of Teacher Education (NCATE) and is the professional accrediting body for over 900 educator preparation providers or TPPs. The CAEP June 11, 2013 update of standards says, “surmounting all others, insist that preparation be judged by outcomes and impact on P-12 student learning and development.”

³ Put another way, this value-added metric ties the performance of students back to the programs at which their teachers were trained. States using value-added include Colorado, Delaware, Georgia, Louisiana, Massachusetts, North Carolina, Ohio, Tennessee, and Texas. For more detail on the methodologies states are using, or plan to use. See: Henry, G. T., Thompson, C. L., Bastian, K. C., Fortner, C. K., Kershaw, D. C., Marcus, J. V., & Zulli, R. A. (2011). *UNC teacher preparation program effectiveness report*.

⁴ This includes comparisons between and among traditional and alternative providers. Despite the growth in alternative providers, about 75 percent of teachers are still being prepared in colleges and universities (National Research Council, 2010). These numbers imply that much of what we stand to learn about the value of a particular kind of training will likely come from studying differences in traditional providers, and that substantial improvements in the skills of new teachers will likely be achieved only through improvement to these traditional programs.

⁵ For more on the context of assessing individual teachers, see:

Raudenbush, S., & Jean, M. (October 2012). How Should Educators Interpret Value-Added Scores? Carnegie Knowledge Network. <http://carnegieknowledge.org/briefs/value-added/interpreting-value-added>

McCaffrey, D. (June 2013). Do Value-Added Methods Level the Playing Field for Teachers? Carnegie Knowledge Network. <http://carnegieknowledge.org/briefs/value-added/different-schools>

Loeb, S., & Candelaria, C. (July 2013). How Stable are Value-Added Estimates across Years, Subjects, and Student Groups? Carnegie Knowledge Network. <http://carnegieknowledge.org/briefs/value-added/value-added-stability>

⁶ Far more studies assess the relative effectiveness of teachers trained in traditional college and university programs versus those who entered the profession through alternative routes, often having received some level of training though usually far less than what is typical from traditional teacher providers.

See: Constantine, J., Player, D., Silva, T., Hallgren, K., Grider, M., & Deke, J. (2009). An evaluation of teachers trained through different routes to certification. *Final Report for National Center for Education and Regional Assistance*, 142.

Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative routes to teaching: The impacts of Teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management*, 25(1), 75-96.

Goldhaber, D. 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.

⁷ This is also roughly equivalent to the typical difference in effectiveness between first- and second-year teachers.

See: Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, 31(4), 416-440.

- ⁸ Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2012). Teacher Preparation Programs and Teacher Quality: Are There Real Differences Across Programs? University of Missouri Department of Economics Working Paper Series. http://economics.missouri.edu/working-papers/2012/WP1204_koedel_et_al.pdf
- ⁹ Mihaly, K., McCaffrey, D., Sass, T., & Lockwood, J.R. (Forthcoming). Where You Come From or Where You Go? Distinguishing Between School Quality and the Effectiveness of Teacher Preparation Program Graduates. Forthcoming in *Education Finance and Policy*, 8:4 (Fall 2013).
- ¹⁰ Gansle, Kristin A., Noell, George H., R. Maria Knox, Michael J. Schafer. (2010). Value Added Assessment of Teacher Preparation in Louisiana: 2005-2006 to 2008-2009. Unpublished report.
- Noell, George H., Bethany A. Porter, R. Maria Patt, Amanda Dahir. 2008. Value Added Assessment of Teacher Preparation in Louisiana: 2004-2005 to 2006-2007. Unpublished report.
- ¹¹ Henry et al., 2011, *ibid*.
- ¹² Goldhaber, D., & Liddle, S. (2011). The Gateway to the Profession: Assessing Teacher Preparation Programs Based on Student Achievement. CEDR. University of Washington, Seattle, WA. [http://www.cedr.us/papers/working/CEDR%20WP%202011-2%20Teacher%20Training%20\(9-26\).pdf](http://www.cedr.us/papers/working/CEDR%20WP%202011-2%20Teacher%20Training%20(9-26).pdf)
- Goldhaber, D., Liddle, S., & Theobald, R. (2013a). The gateway to the profession: Assessing teacher preparation programs based on student achievement. *Economics of Education Review*.
- ¹³ Note that all the comparisons between TPP graduates are based on within-state differences in the measured effectiveness of in-service teachers. A significant share of TPP graduates never end up teaching, or they move from the state in which they graduate to another state.
- See, for instance: Goldhaber, D., Krieg, J., and Theobald, R. (2013b). Knocking on the Door to the Teaching Profession? Modeling the entry of Prospective Teachers into the Workforce. CEDR Working Paper 2013-2. University of Washington, Seattle, WA.
- ¹⁴ Goldhaber et al. (2013a) find there are significant differences between teacher trainees at different institutions in terms of measures of academic proficiency such as college entrance exam scores. Boyd et al. (2009) find that training experiences differ, at least to some degree, across institutions.
- ¹⁵ Goldhaber, D. (2007). Everyone's Doing It, but What Does Teacher Testing Tell Us about Teacher Effectiveness? *Journal of Human Resources*, 42(4), 765-794.
- ¹⁶ Boyd et al., 2009, *ibid*.
- ¹⁷ Ronfeldt, M. (2012). Where should student teachers learn to teach? Effects of field placement school characteristics on teacher retention and effectiveness. *Educational Evaluation and Policy Analysis*, 34(1), 3-26.
- ¹⁸ Related to this, some researchers (Koedel et al., 2012) argue that earlier studies (e.g. Gansle et al., 2010) of the variation in effectiveness of teachers from different training programs overstate what we know about the true differences between graduates of different programs. This is because the studies fail to properly account for the "clustering" of students within teachers. To put this another way, when estimating the effectiveness of TPP graduates it is important to make sure that the statistical model recognizes that students with a particular teacher, for instance one who graduated from program "A" and who has 25 students, are not treated as if they *each* have a different teacher from program "A" since these students likely have shared experiences with the other students in the classroom.
- ¹⁹ In part because institutions themselves may change over time.
- ²⁰ Goldhaber et al., 2013a, *ibid*.
- ²¹ For more detail on the data and labor market conditions necessary to support models designed to disentangle school and district factors from the effectiveness of TPP graduates, see Mihaly et al. (forthcoming) and Goldhaber et al. (2013).
- ²² Raudenbush S. (April 2013). What Do We Know About Using Value-Added to Compare Teachers Who Work in Different Schools? Carnegie Knowledge Network. <http://carnegieknowledge.org/briefs/value-added/different-schools>
- ²³ See Boyd et al. (2009) and Goldhaber and Cowan. (2013). Both find the correlation of the effectiveness of TPP graduates in math and reading/English language arts to be in the range 0.50-0.60.
- Goldhaber, D., and Cowan J. (2013). Excavating the Teacher Pipeline: Teacher Training Programs and Teacher Attrition. CEDR Working Paper 2013-3. University of Washington, Seattle, WA.
- ²⁴ Moreover, this kind of research could go beyond training generally to explore questions of fit, such as: Is the value of different training contingent on the type of school in which teachers find themselves employed?

²⁵ For instance, the edTPA, which was developed jointly by Stanford University and the American Association of Colleges for Teacher Education, is a new multi-measure assessment of teacher trainee skills that is designed to be aligned to state and national standards and used to assess teacher trainees before graduation. In theory, the edTPA may guarantee a minimum level of competency, and it may be able to distinguish good teacher preparation programs from bad ones (Overview: edTPA, 2013). It remains unclear whether it will serve any of these purposes (Greenberg and Walsh, 2012). That's partly because it was only in 2012 that selected programs and states used it to link trainee performance on the assessment to measures of effectiveness of in-service teachers. Overview: edTPA. (2013). Retrieved July 15, 2013, from <http://edtpa.aacte.org/about-edtpa>.

Greenberg, J., & Walsh, K. (2012). What Teacher Preparation Programs Teach about K-12 Assessment: A Review. *National Council on Teacher Quality*.

²⁶ For more on the challenges of changing postsecondary institutions, see:

McPherson, M. S., & Schapiro, M. O. (1999). Tenure issues in higher education. *The Journal of Economic Perspectives*, 13(1), 85-98.

McCormick, R. E., & Meiners, R. E. (1988). University governance: A property rights perspective. *Journal of Law and Economics*, 31, 423.

For evidence that institutions respond to public rankings, see:

Griffith, A., & Rask, K. (2007). The influence of the US News and World Report collegiate rankings on the matriculation decision of high-ability students: 1995–2004. *Economics of Education Review*, 26(2), 244-255.

Hossler, D., & Foley, E. M. (1995). Reducing the noise in the college choice process: The use of college guidebooks and ratings. *New Directions for Institutional Research*, 1995(88), 21-30.

Meredith, M. (2004). Why do universities compete in the ratings game? An empirical analysis of the effects of the US News and World Report college rankings. *Research in Higher Education*, 45(5), 443-461.

²⁷ These pressures could lead programs to distort their programs (often referred to as “Campbell’s Law”), but if the ranking criteria are aligned with good practices, they may well result in program improvements.

²⁸ Constantine et al., 2000, *ibid*.

Goldhaber, et al., 2013b, *ibid*.

²⁹ Moreover, some of this work (Goldhaber and Cowan, 2013) suggests there may be tradeoffs between the effectiveness of graduates from different programs and the length of time they spend in the teacher workforce. See: Ronfeldt, 2012, *ibid*.

Ronfeldt, M., Reininger, M., & Kwok, A. (2013). Recruitment or Preparation? Investigating the Effects of Teacher Characteristics and Student Teaching. *Journal of Teacher Education*.

³⁰ For more discussion on this see: Plecki, M. L., Elfers, A. M., & Nakamura, Y. (2012). Using Evidence for Teacher Education Program Improvement and Accountability An Illustrative Case of the Role of Value-Added Measures. *Journal of Teacher Education*, 63(5), 318-334.

³¹ McCaffrey, Daniel, June 2013, *ibid*.

³² Harris, D. N. (May 2013) How Do Value-Added Indicators Compare to Other Measures of Teacher Effectiveness? Carnegie Knowledge Network. <http://carnegieknowledge.org/briefs/value-added/value-added-other-measures>

³³ Goldhaber, D., & Loeb, S. (April 2013) What are the Tradeoffs Associated with Teacher Misclassification in High Stakes Personnel Decisions? Carnegie Knowledge Network. <http://carnegieknowledge.org/briefs/value-added/teacher-misclassifications>

³⁴ See, for instance: Council for the Accreditation of Educator Preparation (CAEP). (2013). Commission on Standards and Performance Reporting. CAEP Accreditation Standards and Evidence: Aspirations for Education Preparation. June 11, 2013.

³⁵ There is, for instance, significant empirical evidence that minority teachers perform less well on licensure tests, some of which are used for admission into TPPs.

See: Eubanks, S. C., & Weaver, R. (1999). Excellence through diversity: Connecting the teacher quality and teacher diversity agendas. *Journal of Negro Education*, 451-459.

Goldhaber, 2007, *ibid*.

Goldhaber, D. & Hansen, M. (2010). Race, Gender, and Teacher Testing: How Objective a Tool is Teacher Licensure Testing? *American Educational Research Journal*, 47(1), 218-251.

WHAT DO VALUE-ADDED MEASURES OF TEACHER PREPARATION PROGRAMS TELL US?

Gitomer, D. H., & Latham, A. S. (2000). Generalizations in teacher education: Seductive and misleading. *Journal of Teacher Education*, 51(3), 215-220.

³⁶ They also likely would want to know about a non-value-added outcome: whether graduating from a particular TPP might change the likelihood that they find a job.

³⁷ Levine, 2006, *ibid*.

³⁸ For more information on the distinction between the “decay” and “selectivity decay” estimates, see Goldhaber et al. (2013a).

AUTHOR



Dan Goldhaber is the Director of the Center for Education Data & Research and a Professor in Interdisciplinary Arts and Sciences at the University of Washington Bothell. He is also the co-editor of *Education Finance and Policy*, and a member of the Washington State Advisory Committee to the U.S. Commission on Civil Rights. Dan previously served as an elected member of the Alexandria City School Board from 1997-2002, and as an Associate Editor of *Economics of Education Review*. Dan's work focuses on issues of educational productivity and reform at the K-12 level, the broad array of human capital policies that influence the composition, distribution, and quality of teachers in the workforce, and connections between students' K-12 experiences and postsecondary outcomes. Topics of published work in this area include studies of the stability of value-added measures of teachers, the effects of teacher qualifications and quality on student achievement, and the impact of teacher pay structure and licensure on the teacher labor market. Previous work has covered topics such as the relative efficiency of public and private schools, and the effects of accountability systems and market competition on K-12 schooling. Dan's research has been regularly published in leading peer-reviewed economic and education journals such as: *American Economic Review*, *Review of Economics and Statistics*, *Journal of Human Resources*, *Journal of Policy and Management*, *Journal of Urban Economics*, *Economics of Education Review*, *Education Finance and Policy*, *Industrial and Labor Relations Review*, and *Educational Evaluation and Policy Analysis*. The findings from these articles have been covered in more widely accessible media outlets such as National Public Radio, the *New York Times*, the *Washington Post*, *USA Today*, and *Education Week*. Dr. Goldhaber holds degrees from the University of Vermont (B.A., Economics) and Cornell University (M.S. and Ph.D., Labor Economics).

ABOUT THE CARNEGIE KNOWLEDGE NETWORK

The Carnegie Foundation for the Advancement of Teaching has launched the Carnegie Knowledge Network, a resource that will provide impartial, authoritative, relevant, digestible, and current syntheses of the technical literature on value-added for K-12 teacher evaluation system designers. The Carnegie Knowledge Network integrates both technical knowledge and operational knowledge of teacher evaluation systems. The Foundation has brought together a distinguished group of researchers to form the *Carnegie Panel on Assessing Teaching to Improve Learning* to identify what is and is not known on the critical technical issues involved in measuring teaching effectiveness. Daniel Goldhaber, Douglas Harris, Susanna Loeb, Daniel McCaffrey, and Stephen Raudenbush have been selected to join the Carnegie Panel based on their demonstrated technical expertise in this area, their thoughtful stance toward the use of value-added methodologies, and their impartiality toward particular modeling strategies. The Carnegie Panel engaged a User Panel composed of K-12 field leaders directly involved in developing and implementing teacher evaluation systems, to assure relevance to their needs and accessibility for their use. This is the first set of knowledge briefs in a series of Carnegie Knowledge Network releases. Learn more at carnegieknowledgenetwork.org.

CITATION

Dan Goldhaber. Carnegie Knowledge Network, "What Do Value-Added Measures of Teacher Preparation Programs Tell Us?" Last modified November 2013.
http://www.carnegieknowledgenetwork.org/briefs/teacher_prep/



Carnegie Foundation for the Advancement of Teaching

Carnegie Foundation for the Advancement of Teaching
51 Vista Lane
Stanford, California 94305
650-566-5100

Carnegie Foundation for the Advancement of Teaching seeks to vitalize more productive research and development in education. We bring scholars, practitioners, innovators, designers, and developers together to solve practical problems of schooling that diminish our nation's ability to educate all students well. We are committed to developing networks of ideas, expertise, and action aimed at improving teaching and learning and strengthening the institutions in which this occurs. Our core belief is that much more can be accomplished together than even the best of us can accomplish alone.

www.carnegiefoundation.org

We invite you to explore our website, where you will find resources relevant to our programs and publications as well as current information about our Board of Directors, funders, and staff.



This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 Unported license. To view the details of this license, go to:
<http://creativecommons.org/licenses/by-nc/3.0/>.

Knowledge Brief 12
November 2013
carnegieknowledge.org

Funded through a cooperative agreement with the Institute of Education Sciences. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.