

## **What do surveys of program completers tell us about teacher preparation quality?**

Kevin C. Bastian  
Senior Research Associate and Associate Director  
Education Policy Initiative at Carolina  
Department of Public Policy  
University of North Carolina at Chapel Hill  
314 Cloister Court  
Chapel Hill, NC 27514  
919.698.6658  
[kbastian@email.unc.edu](mailto:kbastian@email.unc.edu)

Min Sun  
Assistant Professor in Quantitative Policy Research  
College of Education  
University of Washington  
2012 Skagit Lane  
M205 Miller Hall  
Seattle, WA 98195  
206.221.1625  
[misun@uw.edu](mailto:misun@uw.edu)

Heather Lynn  
Assistant Professor of the Practice  
East Carolina University  
Department of Special Education, Foundations, and Research  
504 Speight Hall  
Greenville, NC 27858  
336.337.7533  
[lynnh16@ecu.edu](mailto:lynnh16@ecu.edu)

### **Acknowledgements**

We wish to thank the University of North Carolina General Administration for its continued promotion and support of the Teacher Quality Research Initiative. We also recognize the North Carolina Department of Public Instruction for aiding in the development and annual administration of the New Teacher Preparation Survey. We acknowledge funding for this project from the Teacher Quality Research Initiative.

Paper presented at the 43<sup>rd</sup> Annual AEFP Conference in Portland, Oregon  
March 15, 2018

## **What do surveys of program completers tell us about teacher preparation quality?**

### **Abstract**

Surveys of teacher preparation program (TPP) completers have become one widely used measure for program accountability and improvement yet there is little evidence as to whether perceptions of preparation experiences predict the workforce outcomes of early-career teachers. In the present study we use statewide completer survey data from North Carolina to assess whether perceptions of preparation quality and opportunities to learn during teacher preparation predict completers' value-added estimates, evaluation ratings, and retention. Results indicate that completers' perceptions of their TPP are modestly associated with these early-career teacher outcomes. In particular, a construct for preparation experiences focused on creating supportive learning environments predicts beginning teacher effectiveness and retention. These results suggest that completer surveys can be a valuable source of data for evidence-based TPP improvement.

## **Introduction**

In recent years teacher preparation programs (TPPs) have been under increasing pressure to demonstrate the quality of program components, track the effectiveness of program completers, and use data and evidence for program improvement. This pressure is exemplified by TPP accreditation standards and state-level evaluation systems that emphasize outcomes for program completers and continuous improvement cycles (Council for the Accreditation of Educator Preparation, 2013). While measures of teacher effectiveness (e.g. value-added estimates or evaluation ratings) often receive the most attention, these accreditation standards and evaluation systems also mandate that TPPs assess completers' perceptions of program quality through surveys administered at program completion or during the first year of teaching. For example, CAEP's Program Impact standard requires TPPs to collect valid and reliable data as to whether completers' preparation was effective and relevant to their job responsibilities (CAEP, 2013). Likewise, 33 states report assessing traditional TPPs with surveys of completers' satisfaction with the preparation they received (Government Accountability Office, 2015).

Relative to measures of teacher effectiveness, these completer surveys may offer several advantages to states and TPPs in their program accountability and improvement efforts. First, since completer surveys are typically available for analysis prior to measures of completer effectiveness, they can provide more timely feedback on program performance. For example, completer surveys are available upon program completion or during the first-year of teaching; measures of teacher effectiveness are not generally available until more than a year after program completion. Second, since completer surveys typically include items on program quality across many areas (e.g. classroom management, instructional strategies, content knowledge), they provide states and TPPs more granular and actionable feedback than a single value-added

estimate or evaluation rating. Finally, since the survey data focus on completer satisfaction, these surveys offer programs and states more proximal and direct ratings of TPP quality than measures of completer effectiveness.

Given these potential strengths, surveys of TPP completers may be an important contributor to data-driven program accountability and improvement. However, there is little evidence as to whether completers' perceptions of TPP quality or their experiences within the TPP go on to predict their effectiveness and retention as early-career teachers. This evidence is crucial to the utility of completer surveys: if completers' perceptions of TPP quality do not predict their outcomes as early-career teachers, then states and TPPs should critically examine whether and how they act on the evidence provided by the survey responses.

Therefore, to inform the accountability and improvement efforts of states and TPPs, the present study addresses the following questions: (1) Do completers' perceptions of their TPP predict their effectiveness as early-career teachers? (2) Do completers' perceptions of their TPP predict their retention in teaching? Specifically, we assess whether completers' responses to the North Carolina New Teacher Preparation Survey (NTPS) predict their value-added estimates, evaluation ratings, and whether they remain teaching in North Carolina public schools (NCPS). These analyses are based on a straightforward hypothesis: those who feel better prepared to teach or who are offered more opportunities to learn during their preparation program will be more effective and more likely to persist in teaching. By testing this hypothesis, we make two inter-related contributions to teacher education. First, we assess whether states and TPPs can use completer survey responses for evidence-based accountability and improvement efforts. Second, we examine whether there are specific elements or features of TPPs that significantly predict the

outcomes of early-career teachers. These results may help TPPs focus their reform efforts on program elements related to completers effectiveness and retention.

In the remainder of the present study we describe the literature connecting teacher preparation to workforce outcomes and provide more information on the NTPS. Next, we detail our research sample, outcome variables, and the survey measures we create from the NTPS responses. We follow this with a description of our analytical methods and findings. Finally, we highlight the implications of this study for program accountability and improvement efforts.

## **Background**

### ***Relationships Between Preparation Programs and Teacher Workforce Outcomes***

Estimating the associations between completers' perceptions of their preparation programs and teacher effectiveness and retention is conceptually and empirically challenging. It is difficult to isolate the link between teachers' pre-service preparation and their on-the-job outcomes because many in-service factors influence teachers' effectiveness and decisions to stay in teaching. These in-service factors may include teacher pay, schools' geographical location and community characteristics, districts' personnel management practices, peer collaboration, and school culture (Boyd, Grossman, Ing, Lankford, Loeb, & Wyckoff, 2011; Boyd, Lankford, Loeb, & Wyckoff, 2005; Clotfelter, Glennie, Ladd, & Vigdor, 2008; Guarino, Brown, & Wyse, 2011; Papay & Kraft, 2016; Ladd, 2011; Sun, Loeb, & Grissom, 2017).

Moreover, these in-service factors may influence how teachers perceive the quality of their preparation. This adds another layer of difficulty in linking teachers' perceptions of TPPs to their effectiveness and retention. Teachers who secure employment in advantaged schools with supportive colleagues and administrators may be more likely to succeed in their early years of teaching, while the transition to teaching is often harder for those working in disadvantaged

schools or less supportive working environments (Hansen, 2016; Kilgore, Ross, & Zbikowski, 1990). Beginning teachers in more advantaged schools may be more likely to attribute their initial success to the perceived quality of their preparation programs; conversely, peers who are struggling in challenging work environments may underrate their preparation program quality. This problem of using perception and outcome measures from the same individual is called common source bias (Boyd et al., 2011). In our Sample and Measures section we discuss an approach to alleviate this concern.

With regard to completer outcomes, this study focuses on multiple measures, including value-added, evaluation ratings, and retention in teaching. We refer to the first two outcomes as teacher effectiveness. Value-added provides an estimate of the degree to which teachers contribute to the learning of students in a given subject and year in comparison to other teachers who taught similar students. Evaluation ratings focus on teachers' instructional practices and broader contributions to the school, while retention measures teachers' persistence in the profession. Collectively, these three measures capture some of the multi-dimensional contributions of TPPs to K-12 schools. This multi-outcome view recognizes that preparation programs that are effective in one area (e.g. the value-added of program completers) are not necessarily effective in other domains (e.g. teacher retention; Goldhaber & Cowan, 2014).

*Associations between TPPs and Value-Added:* In light of evidence questioning the reliability and validity of value-added estimates, there are concerns about holding individual teachers or TPPs accountable for student achievement (AERA, 2015). Despite this concern, research indicates that teacher effectiveness predicts students' long-term success: students assigned to higher value-added teachers are more likely to attend college, have higher salaries in their professions, and save more for retirement (Chetty, Friedman, & Rockoff, 2014). As such,

teacher value-added is one policy-relevant measure—as of 2015, 15 states report evaluating traditional TPPs with student achievement data—with which to assess teacher preparation quality (Government Accountability Office, 2015).

Research on the effectiveness of TPP completers has returned mixed findings. Some studies find meaningful effectiveness differences based on the preparation program attended (Boyd, Lankford, Loeb, & Wyckoff, 2009; Gansle, Noell, & Burns, 2012; Goldhaber, Liddle, & Theobald, 2013). For example, Boyd and colleagues (2009) find that the effectiveness difference between an average TPP and the highest value-added TPP is equivalent in magnitude to the achievement differences between students who do and do not qualify for subsidized school meals. Conversely, other studies have found that value-added differences across TPPs are small and generally statistically insignificant (Koedel, Parsons, Podgursky, & Ehlert, 2015; Mihaly, McCaffrey, Sass, & Lockwood, 2013).

A concern with many of these studies is that they regard teacher preparation as a “black box” without attempting to understand the components of these programs that may explain completer effectiveness. However, there are a few exceptions. Harris and Sass (2011) and Preston (2017) examine the extent to which teachers’ pre-service course-taking is associated with student achievement gains. They find little evidence that course-taking and programmatic course requirements predict teacher value-added. In contrast, Boyd and colleagues (2009) find that first-year teachers are more effective when they have more opportunities to learn and practice classroom teaching. Both Boyd and colleagues (2009) and Henry and colleagues (2013) find that more content preparation in mathematics predicts value-added in that subject-area. Our study will extend this body of work on teacher preparation components by assessing whether features

of programs, as measured by completer surveys, predict teachers' contributions to student achievement.

*Associations between TPPs and Evaluation Ratings:* The utility of value-added estimates to assess TPP quality has been critiqued for (1) only providing effectiveness data for teachers in tested-grades and subject-areas; (2) capturing a narrow set of teaching practices (Floden, 2012); and (3) failing to distinguish teaching quality across different domains to aid program improvement (Bastian, Patterson, & Pan, 2017; Lincove, Osborne, Dillon, & Mills, 2014; Ronfeldt & Campbell, 2016). As such, the development of state-level teacher evaluation systems offers an opportunity to address these concerns and assess whether TPPs are associated with the evaluation ratings of program completers.

To date, two studies have examined evaluation ratings for TPP completers: Ronfeldt and Campbell (2016) with ratings from Tennessee and Bastian and colleagues (2017) with ratings from North Carolina. Both studies return very similar findings. Specifically, the studies found meaningful and statistically significant differences across TPPs in the evaluation ratings of program completers. Furthermore, results indicated that (1) it is important to control for school context when linking TPPs to their completers' ratings—this echoes findings from Steinberg and Garrett (2016) and Whitehurst and colleagues (2014)—and (2) evaluation ratings provide evidence on the performance of TPPs that is distinct from value-added estimates. Like previous research on the value-added of TPP completers, neither of these studies examined relationships between components or features of TPPs and the evaluation ratings of program completers.

*Associations between TPPs and Completer Retention:* Besides developing teacher effectiveness, states, districts, and schools are also concerned about TPPs' ability to prepare teachers who persist in teaching. Teacher turnover is costly, not only in terms of the financial

burden of recruiting and inducting new teachers, but also, in terms of the loss of stability of practice and culture in the school (Alliance for Excellent Education, 2014; Ronfeldt, Loeb, & Wyckoff, 2013). Teacher preparation programs may influence retention through the quality of teachers they produce—on average, more effective teachers remain in teaching (Goldhaber, Gross, & Player, 2011)—and by selecting on and inculcating traits associated with persistence. As such, retention is another policy relevant outcome with which to assess TPPs.

Using data from Washington State, Goldhaber and Cowan (2014) find large differences across TPPs in the rate at which teachers exit their schools and the teaching profession. These findings are robust to within-school retention comparisons and indicate that considerations of TPP quality should also weigh the retention of TPP completers. Most relevant to the present study, additional work has linked completer perceptions of TPPs to their planned and actual retention decisions. Analyses by Darling-Hammond and colleagues (2002) show that the extent to which teachers felt well-prepared upon entering the profession was significantly correlated with their plans to remain in teaching. Later analyses by DeAngelis and colleagues (2013) indicate that perceptions of preparation quality predict planned *and* actual retention decisions. While DeAngelis and colleagues' work provides initial evidence on the relationship between TPP quality and teacher retention, their measure of perceived TPP quality stemmed from a single survey item. Our analyses measure perceptions of TPP quality and opportunities to learn with a much broader set of survey items.

### ***The New Teacher Preparation Survey***

During the 2013-14 academic year a team of education researchers, teacher education faculty, and state education officials designed the North Carolina NTPS to meet two objectives: (1) as a source of data for state-level TPP accountability reporting and (2) as evidence for

program improvement efforts. To design the instrument, the survey team gathered beginning teacher surveys administered by other researchers and states and identified survey domains and items to include in the NTPS. The survey design team refined these items throughout the fall and winter of 2013-14 and piloted the items with a set of beginning teachers in spring 2014. Using an online survey platform, North Carolina administered the NTPS to first-year teachers in late spring 2014 and has subsequently administered the online survey to first-year teachers in the spring of the 2014-15, 2015-16, and 2016-17 school years.

The NTPS asks teachers to report on a range of topics, including their academic background, route and timing of preparation, perceptions of their preparation program, confidence in their ability to implement key instructional tasks, and job satisfaction and future career plans. The present study focuses on two sections of the survey—*Perceptions of Teacher Preparation Quality* and *Opportunities to Learn*. The *Perceptions of Teacher Preparation Quality* section includes a series of items detailing first-year teachers' perceptions of how well their TPP prepared them to carry out key teaching tasks. The *Opportunities to Learn* section is a series of items capturing first-year teachers' perceptions of how much opportunity they had to acquire/develop key knowledge and skills during their preparation.

These two survey sections align with North Carolina's Professional Teaching Standards that are used to guide school administrators' classroom observations and evaluations of teachers' performance. The NTPS items are also aligned with several dimensions of classroom observation rubrics designed to capture effective teaching, including Danielson's Framework for Teaching (2007) and Pianta and colleagues' Classroom Assessment Scoring System (2006). Specifically, the NTPS items capture the degree to which teachers perceive that their TPP prepared them well or provided sufficient learning opportunities to plan and implement explicit

teaching strategies in their content area and to use assessment data to reflect and adjust their teaching. Survey items also capture the perceived contribution of the TPP to teachers' establishment of a supportive learning environment through classroom management and building relationships with students and families. Lastly, survey items focus on teachers' engagement with diverse learners, particularly English language learners and special education students.

## **Sample and Measures**

### ***Research Sample***

We focus our analyses on survey respondents (first-year teachers) in the 2013-14, 2014-15, and 2015-16 school years who were traditionally prepared at in-state colleges or universities. Here, traditionally prepared includes those with an undergraduate education degree, those with a graduate level education degree resulting in an initial teaching license, or those completing their initial licensure through a license/certificate only program. We target traditionally prepared teachers, rather than those in alternative or lateral entry programs, since they have already completed their teacher education coursework and can more fully reflect on the quality of their preparation and opportunities to learn. Furthermore, we focus on completers of in-state preparation programs because North Carolina, like many other states, provides financial support to these institutions and evaluates them on a range of program quality and completer outcomes. Following Boyd and colleagues (2009), we estimate whether perceptions of TPPs predict the effectiveness and persistence of first-year teachers. For our first two survey cohorts—2013-14 and 2014-15—we also estimate whether perceptions of TPPs (taken in the spring of the first-year of teaching) predict outcomes in the second-year of teaching.<sup>1</sup> These results allow us to assess whether any significant relationships between perceptions of TPPs and teacher outcomes persist.

Additionally, these second-year teacher models allow us to assess whether new relationships emerge as teachers develop on the job.

During the 2013-14 through 2015-16 school years, North Carolina distributed the NTPS to nearly 8,900 first-year teachers who had been traditionally prepared at an in-state college or university. Overall, 3,937 of these first-year teachers completed the NTPS—a response rate of nearly 45 percent. These survey respondents are our analytical sample for value-added, evaluation rating, and retention analyses. One potential concern with the use of completer survey data for program accountability and improvement is differential response rates across TPPs. Appendix Table 1 shows that response rates varied across North Carolina TPPs: from a low of 20 percent to a high of 65 percent. While these response rate differences do not necessarily effect our statewide estimates, low response rates may adversely impact the utility of completer survey data for TPPs.

A second potential concern is whether our analytical sample generalizes to the full population of in-state prepared beginning teachers. That is, would our results remain comparable if all eligible teachers had completed the NTPS? While we cannot directly test this, Table 1 displays descriptive data for respondents and non-respondents. When comparing in-state prepared respondents to non-respondents, we find that respondents were more likely to be female, less likely to be a minority, and were more likely to return for a second year of teaching. Respondents and non-respondents had comparable value-added estimates and evaluation ratings and worked in classrooms and schools with similar characteristics. These descriptive statistics suggest that observable differences should not seriously compromise the generalizability of findings. Nonetheless, results need to be interpreted carefully given the potential for unobserved differences between respondents and non-respondents.

[Insert Table 1 about here]

### ***Outcome Measures***

*EVAAS Estimates:* To examine teachers' contributions to student achievement we use Education Value-Added Assessment System (EVAAS) scores from North Carolina's early grades reading exam; End-of-Grade (EOG) exams in mathematics, reading, and science; and End-of-Course (EOC) and final exams in secondary grades. These EVAAS estimates are produced by the SAS Institute and serve as teachers' official value-added effectiveness measure in North Carolina. For the early-grades reading exam and EOG exams in mathematics and reading, these value-added estimates come from the EVAAS multi-variate response model (MRM), a random effects model that accounts for students clustering within teachers and for students and their peers clustering within different teachers in different years. For the EOG exams in science and for EOC and final exams in secondary grades, these value-added estimates come from the EVAAS univariate response model (URM), a random effects model that adjusts for the clustering of students within teachers and controls for at least two years of prior student test scores (Rose, Henry, & Lauen, 2012; Wright, White, Sanders, & Rivers, 2010). Research with simulated and actual student achievement data shows that EVAAS estimates, in comparison to teacher value-added generated by alternative approaches (e.g. student fixed effects, teacher fixed effects, hierarchical models) accurately rank and classify teacher effectiveness and have high levels of year-to-year reliability (Rose, Henry, & Lauen, 2012).

To ease interpretability of EVAAS data, which are expressed in normal curve equivalency units or scale score points, we standardized the EVAAS scores within test and year (e.g. 5<sup>th</sup> grade science, algebra II) across all NCPS teachers. With these standardized estimates we can interpret results as the association between perceptions of TPP quality or opportunities to

learn and a percentage of a standard deviation in teacher effectiveness. Our primary models consider these value-added estimates across all subject-areas and school levels; we also estimate models by subject-area or school level to determine whether the associations between TPP perceptions and teacher value-added differ for teachers prepared in certain licensure areas.<sup>2</sup>

*Evaluation Ratings:* To assess teachers' instructional practices we use their evaluation ratings from the North Carolina Educator Evaluation System (NCEES). NCEES is an evaluation rubric in place across NCPS in which principals rate teachers on up to five professional teaching standards: Teachers Demonstrate Leadership (Standard 1); Teachers Establish a Respectful Classroom Environment for a Diverse Group of Students (Standard 2); Teachers Know the Content They Teach (Standard 3); Teachers Facilitate Learning for Their Students (Standard 4); and Teachers Reflect on Their Practice (Standard 5). To rate teachers, principals use teacher submitted artifacts and conduct a series of formal and/or informal classroom observations throughout the school year. Principals do not generate ratings after each observation session, rather, at the end of the school year, principals and teachers have an evaluation conference and teachers receive summative ratings of either not demonstrated, developing, proficient, accomplished, or distinguished on each standard. Although these ratings occur at the end of the school year, it is important to note that informal principal feedback throughout the year may inflate the associations between perceptions of TPPs and evaluation ratings. This can be one type of common source bias in our completer-level analyses.

Our analyses of the NCEES ratings—using correlations, Cronbach's alpha, and exploratory factor analysis—indicate that the five evaluation standards measure a single construct. Therefore, for our primary analyses, we create an evaluation rating composite by summing teachers' ratings across all five teaching standards. For first-year teachers, the mean

composite rating is 15.5 with a standard deviation of 2.1; for second-year teachers, the mean composite rating is 16.6 with a standard deviation of 2.3. Our preferred models assess whether perceptions of TPPs predict this evaluation composite; supplemental models examine whether perceptions of TPPs predict ratings on specific teaching standards.

*Teacher Retention:* To consider teacher retention, we use certified salary files from the North Carolina Department of Public Instruction to code whether individuals return to teach in NCPS. Specifically, we create a dichotomous indicator for whether first-year teachers—in 2013-14, 2014-15, or 2015-16—return for a second-year of teaching in the state. We also create an indicator for returning for a third-year of teaching in North Carolina. This outcome tracks whether first-year teachers in 2013-14 or 2014-15 are third-year teachers in 2015-16 and 2016-17, respectively.

### ***Measures from the NTPS***

Our study focuses on teachers' responses to two sections of the NTPS: *Perceptions of Teacher Preparation Quality* and *Opportunities to Learn*. The *Perceptions of Teacher Preparation Quality* section includes a set of 24 items that start with the survey stem “how well did your teacher preparation program prepare you to...” and ask about specific teaching tasks (e.g. relate classroom teaching to the real world). The response categories for these items are not addressed, not well, somewhat well, well, and very well. The *Opportunities to Learn* section includes a set of 14 items that start with the survey stem “in your teacher preparation program how much experience did you have with the following...” and ask about specific teacher preparation elements (e.g. create formative and summative assessments). The response categories for these items are no experience, few experiences, some experiences, many experiences, and extensive experiences. We created a summative measure of program quality by

averaging responses across all 24 items; likewise, we created a summative measure of opportunities to learn by averaging completers' responses across all 14 items. Overall, in-state prepared teachers perceived their programs to be relatively high in quality, with an average summative rating of 4 on a 1-5 scale ('not addressed' to 'very well'). Regarding opportunities to learn, the average summative rating for in-state prepared teachers was approximately 3.50 (between 'some experiences' and 'many experiences').

To complement these summative measures, we performed exploratory factor analysis with an orthogonal rotation method on each of these NTPS sections. As detailed in Table 2, these analyses revealed a three factor structure for both the *Perceptions of Teacher Preparation Quality* and the *Opportunities to Learn* sections. Factor 1, which we label *Instruction*, is comprised of items on planning, implementing a variety of classroom teaching practices, and developing and using assessment to inform instruction. Factor 2, which we label *Supportive Learning Environments*, includes items on establishing orderly and positive classroom learning environments, communicating high expectations for students, and forming supportive relationships with all students. Finally, factor 3, which we label *Teaching Diverse Learners*, focuses on teaching students with diverse backgrounds (e.g. race/ethnicity, socioeconomic, and lingual) and levels of achievement.<sup>3</sup> Relative to the summative measures we created, these constructs allow states and TPPs to identify whether programs have strengths or weaknesses in particular preparation areas (as judged by program completers) and to determine whether different elements of preparation are associated with teacher effectiveness or retention.

[Insert Table 2 about here]

To assess whether perceptions of TPPs are associated with outcomes for early-career teachers we analyze the NTPS measures at two levels. Our first approach estimates the

associations between completers' own values for the summative and construct measures and their effectiveness and retention as beginning teachers. Using completers' own responses as predictors provides more variation for analyses and may more accurately reflect the program quality and opportunities to learn experienced by the completer. However, using individuals' self-reporting creates a common source bias. That is, those who feel more (less) successful as beginning teachers may then rate their programs higher (lower) irrespective of the actual quality of the program or opportunities to learn in the program. Therefore, as a second approach, we adapt a strategy from Boyd and colleagues (2011) and aggregate completers' responses up to the TPP-level. With this approach we assess how the average response for each institution predicts outcomes for completers. Although these aggregated values mitigate concerns about common source bias, they substantially reduce the available variation with which to predict teacher outcomes. In this study we present analysis results from both of these approaches: (1) summative measures of program quality and opportunities to learn at the individual completer level and aggregated to the university level and (2) three constructs of program quality and opportunities to learn at the individual completer level and aggregated to the university level. All of these measures are standardized to have a mean of zero and standard deviation of one.

To support our teacher effectiveness and retention analyses, we performed two descriptive checks with the NTPS measures. First, we estimated correlations between the *Perceptions of Teacher Preparation Quality* measures and the corresponding *Opportunities to Learn* measures (e.g. correlations between the *Instruction* construct for program quality and the *Instruction* construct for opportunities to learn). Large positive correlations between these measures may lend credence to the survey responses, as more opportunities to learn a particular teaching practice or strategy may be associated with teachers feeling better prepared to perform

such a strategy. At the individual teacher level, the summative measures of program quality and opportunities to learn are correlated at 0.75; aggregated to the university level, these summative measures are correlated at 0.88. At the individual teacher level, the values for the *Instruction*, *Supportive Learning Environments*, and *Teaching Diverse Learners* constructs are correlated at 0.60, 0.34, and 0.60; aggregated to the university-level, these constructs are correlated at 0.70, 0.58, and 0.66, respectively. These correlations indicate that perceptions of the quality of teacher preparation are related to perceived opportunities to learn and practice.

Second, we examined correlations between the focal survey measures (at the individual completer level) and select K-12 school characteristics. This helps us assess whether first-year teachers who rate their TPPs higher are working in more advantaged schools. This sorting could threaten the validity of our analyses, since teachers' perceptions may be related to their work environment and not the quality of their TPP. Appendix Table 2 shows that correlations between perceptions of TPPs and school characteristics are weak. Correlations range from -0.049 to 0.044 and less than 10 percent of the correlations (3 out of 32) are statistically significant. This indicates that teachers sorting into schools is not a significant concern for our analyses of TPP perceptions.

### **Analysis Methods**

To assess teacher effectiveness, as measured by EVAAS estimates and the composite evaluation rating, we estimate linear regression models. Although our retention outcome is dichotomous, which often suggests logistic regression approaches, we assess associations between perceptions of TPPs and teacher retention with a linear probability model.<sup>4</sup> This eases the interpretation of retention findings. To better isolate the associations between perceptions of TPPs and teacher value-added, evaluation ratings, and retention, we control for teacher

demographics, select classroom characteristics, and school characteristics in all models. We control for these measures since (1) EVAAS models do not explicitly adjust for classroom or school context when estimating teacher value-added; and (2) prior work suggests the need to control for teaching contexts with evaluation ratings (Bastian, Patterson, & Pan, 2017; Steinberg & Garrett, 2016; Whitehurst, Chingos, & Lindquist, 2014) and teacher retention (Borman & Dowling, 2008). The teacher demographics include age, gender, and an indicator for being a racial/ethnic minority. The classroom characteristics include the percentage of students taught by the teacher who are a racial/ethnic minority and the average days absent of those students. Finally, the school characteristics include percent economically disadvantaged, percent racial/ethnic minority, school size, per-pupil expenditures, average teacher salary supplements, short-term suspension rates, and violent acts rates. All models include year indicators; since our value-added analyses combine EVAAS estimates across subject-areas, we also insert a set of subject-area indicators into those models.

From these analyses we assess how a one standard deviation increase in perceptions of teacher preparation quality or opportunities to learn is associated with a percentage of a standard deviation change in teacher value-added, the change in composite evaluation rating points, or the probability of returning to teach in NCPS. In our analyses we cluster standard errors at the level of the NTPS data. This means we cluster standard errors at the teacher-level when we assess completer-level survey responses and at the university-level when we aggregate survey responses up to the program level. Standard error values are generally comparable when we cluster all standard errors at the university level. Our basic analysis model is as follows:

$$Y_{ist} = \alpha + \beta(TPP \text{ quality or } OTL)_i + (Demo)_{ist}\Gamma + (Class)_{ist}\Phi + (School)_{st}\Psi + \varepsilon_{ist}$$

where  $Y_{ist}$  is a teacher value-added estimate, composite evaluation rating, or retention decision for teacher  $i$  in school  $s$  at time  $t$ ;

$(TPP\ quality\ or\ OTL)_i$  represents the summative and construct survey measures (at the completer or university-level) for teacher  $i$  at the time of completing the survey;

$Demo_{it}$ ,  $Class_{ist}$ , and  $School_{ist}$  represent a set of teacher demographic, classroom, and school characteristics;

$\beta$ ,  $\Gamma$ ,  $\Phi$ , and  $\Psi$  represent the associations between our focal and control variables and teacher outcomes;

and  $\varepsilon_{ist}$  is an error term for unexplained variation in teacher effectiveness and retention.

There are two potential concerns with our analyses. First, it is possible that completers from different TPPs have different standards about what constitutes program quality or sufficient opportunities to learn. This reference bias may lead completers from certain TPPs to rate their TPP higher or lower (Duckworth & Yeager, 2015). Therefore, to supplement our analyses, we estimate models (displayed in Appendix Table 3) with a university fixed effect. This approach accounts for time-invariant university characteristics—such as common expectations of completers—to assess how within-university variation in perceptions of quality or opportunities to learn predict completer effectiveness and retention. Second, it is possible that teachers sort into schools such that completers who rate their program high (low) generally work in more (less) advantaged environments. Although we could include a school fixed effect to address this type of selection issue, the descriptive checks in Appendix Table 2 mitigate this concern. In our sample there is not a strong relationship between school characteristics and perceptions of TPPs. Furthermore, a school fixed effect limits our analytical sample to schools employing multiple beginning teachers and places greater reliance on the homogeneity of effects assumption (Mihaly, McCaffrey, Sass, & Lockwood, 2013). As such, we do not estimate school fixed effect models.

## Results

### *Do Perceptions of TPPs Predict Teacher Effectiveness?*

Table 3 displays associations between perceptions of TPP quality and opportunities to learn and teacher value-added estimates. For completer-level survey measures (top panel of Table 3), results indicate that teachers' perceptions of program quality and opportunities to learn about creating supportive and well-managed classroom environments predict first and second-year teachers' value-added estimates. For example, a one standard deviation increase in perceived TPP quality for *Supportive Learning Environments* is associated with an increase in teacher value-added of five percent of a standard deviation. Likewise, a one standard deviation increase in perceived opportunities to learn about *Supportive Learning Environments* is associated with a seven percent of a standard deviation increase in teachers' EVAAS estimates. To put the magnitude of these value-added estimates into perspective, we note that during our study period the average difference in EVAAS estimates between first and second-year teachers was 12 percent of a standard deviation. As such, the significant *Supportive Learning Environments* results represent approximately 40 to 60 percent of the average effectiveness difference between first and second-year teachers. These results are similar to those of Boyd and colleagues (2009) who find that more opportunities to develop strategies for handling student misbehavior is associated with the value-added estimates of second-year teachers in New York City. Estimates from models with university fixed effects (shown in Appendix Table 3) return results that are very similar to those displayed in Table 3.

Beyond *Supportive Learning Environments* there are no statistically significant associations between completer-level responses—neither the summative nor the other construct measures—and teacher value-added. For the aggregated, university-level survey measures, we

find that the summative program quality measure predicts higher value-added estimates for first-year teachers. There are no other significant EVAAS results for the university-level survey measures.

[Insert Table 3 about here]

Turning to teacher evaluation ratings, we find that completers' perceptions of their TPP are significantly associated with their composite evaluation rating. These findings are very consistent across our preferred and university fixed effects (Appendix Table 3) approaches. Specifically, the top panel of Table 4 shows that the summative ratings of program quality and opportunities to learn predict significantly higher evaluation ratings for first and second-year teachers. For example, a one standard deviation increase in teachers' summative measure of program quality is associated with an increase of 0.18 points in the composite evaluation rating for first-year teachers and of 0.20 points for second-year teachers. Similarly, the summative measure of opportunities to learn is associated with an increase of 0.16 composite evaluation rating points for first-year teachers and 0.15 evaluation rating points for second-year teachers. To put the magnitude of these coefficients into perspective, we note that the average difference between first and second-year teachers on this composite evaluation rating is 0.50 points. As such, the estimates for these summative measures represent 30 to 40 percent of the average rating difference between first and second-year teachers.

[Insert Table 4 about here]

Like the value-added results, we find that the *Supportive Learning Environments* construct predicts higher scores for first and second-year teachers on the composite evaluation rating. For example, a one standard deviation increase in the *Supportive Learning Environments* construct for TPP quality is associated with an increase of 0.18 composite evaluation rating

points for first-year teachers. Further analyses (available from the authors) show that the *Supportive Learning Environments* construct also predicts significantly higher ratings on the closely-aligned Classroom Environment evaluation standard (standard 2 of the NCEES). Across teacher value-added and evaluation ratings, the consistency in positive results for *Supportive Learning Environments* suggests that in-depth preparation focused on classroom management, building supportive relationships with students, and creating high expectations for all students contributes to early-career teachers' effectiveness.

Findings for the other completer-level survey constructs indicate that perceptions of the quality of preparation in *Instruction* predict completers' composite evaluation rating<sup>5</sup>; for first-year teachers, perceptions of TPP quality and opportunities to learn about *Teaching Diverse Learners* predict higher composite evaluation ratings. Finally, there are few significant results when considering the aggregate, university-level survey measures. In these models (displayed at the bottom of Table 4), the quality of preparation for *Supportive Learning Environments* (first-year teachers) and opportunities to learn about *Supportive Learning Environments* (second-year teachers) predict higher composite ratings.

### ***Do Perceptions of TPPs Predict Teacher Retention?***

Beyond teacher effectiveness, TPPs want to produce teachers who persist in teaching. This is particularly important in the early-career period when many teachers exit the profession and teachers are learning rapidly on the job (Borman & Dowling, 2008; Papay & Kraft, 2015). Therefore, Table 5 displays associations between perceptions of TPP quality and opportunities to learn and whether beginning teachers return for a second and third year of teaching.

[Insert Table 5 about here]

Regarding completers' own perspectives, the top panel of Table 5 reveals several significant findings. The summative measure of program quality predicts returning for a second and third year of teaching. These results are consistent with those of DeAngelis and colleagues (2013) who find that perceptions of program quality are related to teacher retention. Likewise, the summative measure of opportunities to learn predicts returning for a second year of teaching. For example, a one standard deviation increase in the summative values for preparation quality and opportunities to learn is associated with a one percentage point increase in retention for first-year teachers. As a context for these results, we note that nearly 93 percent of the beginning teachers in our sample returned for a second-year of teaching in NCPS; thus, a one percentage point increase in retention related to TPP quality and opportunities to learn would be meaningful in practice. For the subdomains of TPP quality and opportunities to learn, we find that perceptions of program quality around *Teaching Diverse Learners* predicts returning for a third year of teaching. Similar to the results for teacher effectiveness, opportunities to learn about *Supportive Learning Environments* is also significantly associated with returning for a second and third year of teaching. Estimates from retention models with university fixed effects (shown in the bottom panel of Appendix Table 3) yield results that are comparable to those displayed in Table 5.

Finally, when aggregated to the university-level, we find that perceptions of TPP quality and opportunities to learn about *Instruction* predict returning for a second and third-year of teaching (bottom panel of Table 5). These results echo those of Ingersoll and colleagues (2014) who find that first-year teachers are more likely to return if their preparation program offered more opportunities to practice and observe classroom teaching. The summative, university-level measure for opportunities to learn also predicts returning for a second-year of teaching. None of

the results for the *Supportive Learning Environments* and *Teaching Diverse Learners* constructs are statistically significant.

## **Discussion**

As states and accreditation agencies emphasize the role of completer surveys in TPP evaluation systems, it is vital that policymakers and teacher educators better understand the evidence that survey responses provide for program accountability and improvement. With this motivation, we assessed whether completers' perceptions of their program quality and their opportunities to learn predicted their effectiveness and retention in NCPS. Our findings contribute to the literature on teacher preparation and have several practical implications.

We provide the first, large-scale study to assess how perceptions of preparation quality and opportunities to learn are associated with the workforce outcomes of early-career teachers. Importantly, we created summative survey measures and analyzed the survey responses to identify subdomains of TPP quality and opportunities to learn. These more granular data allow researchers to assess whether particular features of programs predict beginning teacher outcomes and help TPPs target their evidence-based reforms. In addition, our study uses three outcome measures to provide a more comprehensive assessment of completers' perceptions and the contributions of TPPs to teachers and schools.

Overall, we find that perceptions of preparation programs are modestly associated with the effectiveness and retention of first and second-year teachers. This suggests that, on average, those who feel better prepared to teach are more effective and more likely to remain in teaching. In particular, we find that perceptions of *Supportive Learning Environments* predict the effectiveness and retention of beginning teachers. These *Supportive Learning Environments* results align with those of Boyd and colleagues (2009), who find that more opportunities to

develop strategies for handling student misbehavior predict second-year teacher value-added. Furthermore, recent analyses by Gill and colleagues (2016) support the importance of classroom management and relationships with students. In a study examining teacher ratings across five commonly used observation protocols (e.g. CLASS, Framework for Teaching), they find that classroom management dimensions/domains have the strongest and most consistent associations with teacher value-added estimates. The importance of *Supportive Learning Environments* suggests that TPPs should consider ways, through coursework, early field experiences, and clinical placements to enhance the quality of preparation and candidates' opportunities to learn about classroom management, building relationships with students, and creating high expectations for student success.

Our findings are more robust at the completer-level rather than aggregated to the university-level. Since completers take the NTPS late in their first-year of teaching, this may reflect common source bias—relationships between completer-level responses and outcomes could be inflated due to perceptions of teaching success. Conversely, these differences in results may also be attributable to completer-level responses more accurately depicting the preparation experiences of candidates (as so much of what candidates learn in their TPP depends on the interactions between individuals and the programmatic elements). When it comes to the use of survey data, completer-level responses are often better-suited for the programmatic decisions of TPPs while aggregate-level data are better-aligned with the needs of policymakers. As such, a takeaway from our results is that survey responses may have more utility for teacher educators and their program improvement efforts than for policymakers and their accountability systems. Completer surveys for accountability purposes may also be adversely impacted by Campbell's

Law: the more states or accreditation agencies use surveys to hold TPPs accountable, the more programs or teachers may corrupt the processes surveys are intended to monitor.

While there are potential benefits to the use of completer surveys, there are also several limitations to consider. First, completer surveys are only as valuable as the content of the items. States need to ensure that a diverse set of researchers, teacher educators, and school officials participate in the creation of a completer survey and that the survey items align with state standards for teacher preparation and in-service teaching. Second, low response rates may threaten the representativeness of survey results for program accountability and improvement purposes. These low response rates may particularly complicate efforts to collect data for small TPPs. As a way to improve response rates, states can consider embedding completer surveys into the official job responsibilities of beginning teachers (e.g. as part of the teacher evaluation process) and/or consider mechanisms to incentivize responses. Finally, there are legitimate questions about when program completers should answer survey items. Completer surveys administered prior to/concurrent with beginning teaching are less likely to suffer from common source bias; however, this timing does not allow respondents to reflect on their TPP with the benefit of actual teaching experience. As research indicates that teacher self-efficacy declines from the student teaching period through the end of the first-year of teaching, completer surveys may provide a more accurate portrayal of TPPs after completers have acquired some teaching experience (Hoy & Spero, 2005).

Given the interest of accreditation bodies, state agencies, and TPPs in using completer surveys for accountability and program improvement, it is important for researchers to regularly assess whether survey responses predict the effectiveness and retention of beginning teachers. To

the extent that completer surveys are associated with these workforce outcomes, they may represent a valuable source of data for evidence-based decision making.

## Notes

<sup>1</sup> We cannot yet assess the performance and retention outcomes of NTPS completers in 2015-16 who were second-year teachers in the 2016-17 school year.

<sup>2</sup> These subject-area and school-level results are available upon request.

<sup>3</sup> For *Perceptions of Teacher Preparation Quality*, Cronbach's alpha values were 0.95, 0.91, and 0.86 for the survey items loading on the Instruction, Supportive Learning Environments, and Teaching Diverse Learners factors. For *Opportunities to Learn*, Cronbach's alpha values were 0.89, 0.74, and 0.83 for the survey items loading on the Instruction, Supportive Learning Environments, and Teaching Diverse Learners factors.

<sup>4</sup> Results from logistic regressions and linear probability models are similar. Logistic regression results are available upon request.

<sup>5</sup> While *Supportive Learning Environments* predicts higher evaluation ratings on standard 2 of the NCEES (Classroom Environment), the *Instruction* construct, either for preparation quality or opportunities to learn, is rarely associated with the closely-aligned Facilitating Student Learning evaluation standard.

## References

- American Education Research Association. (2015). AERA statement on the use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, 44(8), 448-452.
- Alliance for Excellent Education. (2014). On the path to equity: Improving the effectiveness of beginning teachers. Available from: <http://all4ed.org/wp-content/uploads/2014/07/PathToEquity.pdf>
- Bastian, K.C., Patterson, K.M., & Pan, Y. (2017). Evaluating teacher preparation programs with teacher evaluation ratings: Implications for program accountability and improvement. In press, *Journal of Teacher Education*.
- Borman, G.D. & Dowling, N.M. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research*, 78(3), 367-409.
- Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (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. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, 31(4), 416-440.
- Boyd, D., Grossman, P., Ing, M., Lankford, H., & Loeb, S. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal*, 48(2), 303-333.
- Chetty, R., Friedman, J.N., & Rockoff, J.E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *The American Economic Review*, 104(9), 2633-2679.
- Clotfelter, C., Glennie, E., Ladd, H., & Vigdor, J. (2008). Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. *Journal of Public Economics*, 92(5-6), 1352-1370
- Council for the Accreditation of Educator Preparation. (2013). 2013 CAEP Standards. Available from: <http://caepnet.org/~media/Files/caep/standards/caep-standards-one-pager-061716.pdf?la=en>
- Danielson, C. (2007). *Enhancing Professional Practice: A Framework for Teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L., Chung, R., & Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education*, 53(4), 286-302.

- DeAngelis, K.J., Wall, A.F., & Che, J. (2013). The impact of preservice preparation and early career support on novice teachers' career intentions and decisions. *Journal of Teacher Education*, 64(4), 338-355.
- Duckworth, A.L. & Yeager, D.S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237-251.
- Floden, R.E. (2012). Teacher value-added as a measure of program quality: Interpret with caution. *Journal of Teacher Education*, 63(5), 356-360.
- Gansle, K.A., Noell, G.H., & Burns, J.M. (2012). Do student achievement outcomes differ across teacher preparation programs? An analysis of teacher education in Louisiana. *Journal of Teacher Education*, 63(5), 304-317.
- Gill, B., Shoji, M., Coen, T., & Place, K. (2016). The content, predictive power, and potential bias in five widely used teacher observation instruments. Available from: [https://ies.ed.gov/ncee/edlabs/regions/midatlantic/pdf/REL\\_2017191.pdf](https://ies.ed.gov/ncee/edlabs/regions/midatlantic/pdf/REL_2017191.pdf)
- Goldhaber, D., Gross, B., & Player, D. (2011). Teacher career paths, teacher quality, and persistence in the classroom: Are public schools keeping their best? *Journal of Policy Analysis and Management*, 30(1), 57-87.
- Goldhaber, D., Liddle, S., & Theobald, R. (2013). The Gateway to the Profession: Assessing Teacher Preparation Programs Based on Student Achievement. *Economics of Education Review*, 34(2), 29-44.
- Goldhaber, D. & Cowan, J. (2014). Excavating the teacher pipeline: Teacher preparation programs and teacher attrition. *Journal of Teacher Education*, 65(5), 449-462.
- Government Accountability Office. (2015). Teacher preparation programs: Education should ensure states identify low-performing programs and improve information-sharing. Available from: <http://www.gao.gov/assets/680/671603.pdf>
- Guarino, G.M., Brown, A.B., & Wyse, A.E. (2011). Can districts keep good teachers in the schools that most need them? *Economics of Education Review*, 30(5), 962-979.
- Hansen, B. (2016). Teachers at low-income schools deserve respect. Education Weeks. Available from: <http://www.edweek.org/ew/articles/2016/02/10/teachers-at-low-income-schools-deserve-respect.html>
- Harris, D.N. & Sass, T.R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7-8), 798-812.

- Henry, G.T., Campbell, S.L., Thompson, C.L., Patriarca, L.A., Luterbach, K.J., Lys, D.B., & Covington, V.M. (2013). The predictive validity of measures of teacher candidate programs and performance: Toward an evidence-based approach to teacher preparation. *Journal of Teacher Education*, 64(5), 439-453.
- Hoy, A.W. & Spero, R.B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, 21(4), 343-356.
- Ingersoll, R., Merrill, L., & May, H. (2014). What are the effects of teacher education and preparation on beginning teacher attrition? CPRE Research Report. Available from: [http://repository.upenn.edu/cpre\\_researchreports/78/?utm\\_source=repository.upenn.edu%2Fpre\\_researchreports%2F78&utm\\_medium=PDF&utm\\_campaign=PDFCoverPages](http://repository.upenn.edu/cpre_researchreports/78/?utm_source=repository.upenn.edu%2Fpre_researchreports%2F78&utm_medium=PDF&utm_campaign=PDFCoverPages)
- Kilgore, K., Ross, D., & Zbikowski, J. (1990). Understanding the teaching perspectives of first-year teachers. *Journal of Teacher Education*. 41(1), 28-38.
- Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2015). Teacher preparation programs and teacher quality: Are there real differences across programs? *Education Finance and Policy*, 10(2), 508-534.
- Ladd, H. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis*, 33(2), 235-261.
- Lincove, J.A., Osborne, C., Dillon, A., & Mills, N. (2014). The politics and statistics of value-added modeling for accountability of teacher preparation programs. *Journal of Teacher Education*, 65(1), 24-38.
- Mihaly, K., McCaffrey, D., Sass, T.R., & Lockwood, J.R. (2013). Where you come from or where you go? Distinguishing between school quality and the effectiveness of teacher preparation program graduates. *Education Finance and Policy*, 8(4), 459-493.
- Papay, J.P. & Kraft, M.A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, 130, 105-119.
- Papay, J.P. & Kraft, M.A. (2016). The productivity costs of inefficient hiring practices: Evidence from late teacher hiring. *Journal of Policy Analysis and Management*, 35(4), 791-817.
- Pianta, R.C., Hamre, B.K., Haynes, N.J., Mintz, S., & La Paro, K.M. (2006). *Classroom Assessment Scoring System (CLASS) Manual: Middle/Secondary Version Pilot*. Charlottesville: Curry School of Education, University of Virginia.
- Preston, C. (2017). University-based teacher preparation and middle grades teacher effectiveness. *Journal of Teacher Education*, 68(1), 102-116.

- Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, 50(1), 4-36.
- Ronfeldt, M. & Campbell, S.L. (2016). Evaluating teacher preparation using graduates' observational ratings. *Educational Evaluation and Policy Analysis*, 38(4), 603-625.
- Rose, R., Henry, G. T., & Lauen, D. L. (2012). Comparing Value Added Models for Estimating Teacher Effectiveness: Technical Briefing. Consortium for Educational Research and Evaluation – North Carolina. Available from:  
<https://publicpolicy.unc.edu/files/2015/07/Comparing-Value-Added-Models-for-Estimating-Teacher-Effectiveness1.pdf>
- Steinberg, M.P., & Garrett, R. (2016). Classroom composition and measured teacher performance: What do teacher observation scores really measure? *Educational Evaluation and Policy Analysis*, 38(2), 293-317.
- Sun, M., Loeb, S., & Grissom, J. (2017). Building teacher teams: Evidence of positive spillovers from more effective colleagues. *Educational Evaluation and Policy Analysis*, 39(1), 104-125.
- Whitehurst, G.J., Chingos, M.M., & Lindquist, K.M. (2014). Evaluating teachers with classroom observations: Lessons learned in four districts. *The Brookings Institute*. Available from:  
<http://www.brookings.edu/~media/research/files/reports/2014/05/13-teacher-evaluation/evaluating-teachers-with-classroom-observations.pdf>
- Wright, S.P., White, J.T., Sanders, W.L., & Rivers, J.C. (2010). SAS EVAAS Statistical Models. Available from:  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.437.6615&rep=rep1&type=pdf>

**Table 1: Characteristics of NTPS Respondents and Non-Respondents**

	<b>NTPS Respondents</b>	<b>NTPS Non-Respondents</b>
<b><i>Teacher Characteristics</i></b>		
Female	83.67%	79.82%
Minority	15.55%	17.45%
Age	27.99	27.35
Standardized Value-Added	-0.247	-0.189
Evaluation Rating Composite	15.44	15.36
Returns for a 2 <sup>nd</sup> Year of Teaching	92.93%	90.64%
<b><i>Classroom Characteristics</i></b>		
Percent Minority	54.42%	55.37%
Average Number of Student Absences	6.73	6.60
<b><i>School Characteristics</i></b>		
Elementary School	56.79%	57.13%
Middle School	18.39%	16.89%
High School	24.82%	25.97%
Percent Minority	53.12%	54.73%
Percent Economically-Disadvantaged	59.98%	59.95%
Percent of State Assessments Passed	53.51%	53.48%
Unique Teachers	3,937	4,949

Note: This table displays teacher, classroom, and school characteristics for teachers traditionally prepared at in-state institutions who responded and did not respond to the NTPS.

**Table 2: Factor Loadings for TPP Quality and Opportunities to Learn Survey Items**

NTPS Items	Factor 1	Factor 2	Factor 3
<b><i>Perceptions of Teacher Preparation Quality</i></b>			
Collaborate with colleagues to improve student learning	<b>0.42</b>	<b>0.56</b>	0.24
Set challenging and appropriate goals for student learning and performance	<b>0.44</b>	<b>0.62</b>	0.29
Empower students to become self-directed and productive learners	0.35	<b>0.67</b>	0.34
Maintain discipline and an orderly, purposeful learning environment	0.22	<b>0.64</b>	0.42
Work with parents and families to better understand students and to support their learning	0.23	<b>0.47</b>	<b>0.55</b>
Develop positive and supportive relationships with students	0.30	<b>0.76</b>	0.24
Create an environment of high expectations for all students	0.39	<b>0.72</b>	0.25
Teach in ways that support English language learners	0.22	0.24	<b>0.75</b>
Teach in ways that support students with diverse ethnic, racial, cultural, and socioeconomic backgrounds	0.30	0.34	<b>0.66</b>
Teach in ways that support special education students	0.18	0.24	<b>0.76</b>
Teach in ways that support academically gifted students	0.32	0.25	<b>0.65</b>
Develop a classroom environment that promotes respect and group responsibility	0.35	<b>0.70</b>	0.31
Teach the concepts, knowledge, and skills of your discipline	<b>0.70</b>	0.34	0.07
Align instruction with state standards	<b>0.73</b>	0.29	0.03
Relate classroom teaching to the real world	<b>0.67</b>	0.39	0.24
Use knowledge of student learning and curriculum to plan instruction	<b>0.69</b>	0.38	0.23
Develop a variety of assessments	<b>0.70</b>	0.26	0.33
Provide purposeful feedback to students to guide their learning	<b>0.68</b>	0.29	0.36
Differentiate instruction	<b>0.57</b>	0.29	<b>0.51</b>
Use technology in the classroom to improve learning outcomes	<b>0.59</b>	0.21	0.38
Help students think critically to solve problems	<b>0.65</b>	0.31	0.37
Develop students' questioning and discussion skills	<b>0.64</b>	0.33	<b>0.41</b>
Analyze student performance data	<b>0.68</b>	0.22	<b>0.40</b>
Adapt practice based on research and student performance data	<b>0.65</b>	0.24	<b>0.45</b>
<b><i>Opportunities to Learn</i></b>			
Study stages of child development and learning	0.14	<b>0.47</b>	0.36
Develop strategies for managing student behavior	0.18	<b>0.83</b>	0.30
Develop strategies for establishing classroom procedures	0.30	<b>0.81</b>	0.23
Develop strategies for teaching English language learners	0.11	0.25	<b>0.67</b>
Develop strategies for teaching students from diverse racial, ethnic, cultural, and socioeconomic backgrounds	0.26	0.26	<b>0.66</b>
Develop strategies for teaching students with special needs	0.08	0.13	<b>0.78</b>
Develop strategies for teaching students who are academically gifted	0.27	0.22	<b>0.65</b>
Develop strategies for teaching students of varying ability	0.35	0.22	<b>0.70</b>
Apply state or national standards to instruction	<b>0.77</b>	0.15	0.07
Plan units and lessons	<b>0.80</b>	0.19	0.03
Create formative and summative student assessments	<b>0.81</b>	0.12	0.22
Analyze student assessment data/work to adjust instruction	<b>0.71</b>	0.16	0.38
Provide meaningful and specific academic feedback to students	<b>0.66</b>	0.22	<b>0.43</b>
Develop instructional strategies to promote students' critical thinking skills	<b>0.68</b>	0.21	<b>0.41</b>

Note: This table displays factor loadings from exploratory factor analysis on the Perceptions of Teacher Preparation Quality and Opportunities to Learn sections of the NTPS. All factor loadings of 0.40 or above are bolded. We label the three factors as *Instruction* (factor 1), *Supportive Learning Environments* (factor 2), and *Teaching Diverse Learners* (factor 3).

**Table 3: Predicted Associations from the Perceptions of TPPs to Teacher EVAAS**

	Quality of Teacher Preparation		Opportunities to Learn	
	1 <sup>st</sup> Year Teachers	2 <sup>nd</sup> Year Teachers	1 <sup>st</sup> Year Teachers	2 <sup>nd</sup> Year Teachers
<i>Completer-Level Measures</i>				
Summative	0.026 (0.019)	0.018 (0.022)	0.032 (0.023)	0.009 (0.023)
Instruction	0.015 (0.019)	-0.001 (0.021)	0.033 (0.024)	0.006 (0.022)
Supporting Learning Environment	<b>0.048*</b> (0.019)	<b>0.051*</b> (0.023)	<b>0.073**</b> (0.023)	<b>0.062**</b> (0.022)
Teaching Diverse Learners	-0.011 (0.019)	-0.015 (0.021)	-0.029 (0.023)	-0.032 (0.022)
Observation Count (EVAAS estimates)	3,177	2,571	2,246	2,501
<i>University-Level Measures</i>				
Summative	<b>0.041<sup>+</sup></b> (0.023)	0.007 (0.023)	0.034 (0.033)	0.029 (0.028)
Instruction	0.022 (0.025)	0.002 (0.022)	0.035 (0.025)	0.027 (0.021)
Supporting Learning Environment	0.047 (0.028)	0.016 (0.022)	0.038 (0.026)	0.021 (0.022)
Teaching Diverse Learners	-0.021 (0.026)	-0.010 (0.026)	-0.026 (0.025)	-0.000 (0.023)
Observation Count (EVAAS estimates)	3,177	2,571	2,246	2,501

Note: This table displays results from models assessing whether perceptions of TPPs predict the EVAAS estimates of first-year teachers and second-year teachers. +, \*, and \*\* indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. Summative refers to the average of all items for teacher preparation quality or opportunities to learn; Instruction, Supportive Learning Environments, and Teaching Diverse Learners refer to construct measures identified through factor analysis.

**Table 4: Predicted Associations from the Perceptions of TPPs to Evaluation Ratings**

	Quality of Teacher Preparation		Opportunities to Learn	
	1 <sup>st</sup> Year Teachers	2 <sup>nd</sup> Year Teachers	1 <sup>st</sup> Year Teachers	2 <sup>nd</sup> Year Teachers
<i>Completer-Level Measures</i>				
Summative	<b>0.178<sup>**</sup></b> (0.037)	<b>0.197<sup>**</sup></b> (0.045)	<b>0.159<sup>**</sup></b> (0.046)	<b>0.149<sup>**</sup></b> (0.044)
Instruction	<b>0.061<sup>+</sup></b> (0.037)	<b>0.076<sup>+</sup></b> (0.044)	0.047 (0.045)	0.043 (0.045)
Supporting Learning Environment	<b>0.177<sup>**</sup></b> (0.037)	<b>0.217<sup>**</sup></b> (0.045)	<b>0.115<sup>**</sup></b> (0.043)	<b>0.207<sup>**</sup></b> (0.047)
Teaching Diverse Learners	<b>0.082<sup>*</sup></b> (0.036)	0.055 (0.047)	<b>0.128<sup>**</sup></b> (0.044)	0.055 (0.047)
Observation Count	3,264	2,406	2,332	2,337
<i>University-Level Measures</i>				
Summative	0.051 (0.053)	0.058 (0.055)	0.075 (0.081)	0.074 (0.058)
Instruction	0.009 (0.054)	0.014 (0.061)	0.079 (0.072)	0.027 (0.054)
Supporting Learning Environment	<b>0.095<sup>*</sup></b> (0.046)	0.060 (0.063)	0.012 (0.051)	<b>0.097<sup>*</sup></b> (0.041)
Teaching Diverse Learners	-0.050 (0.069)	-0.023 (0.072)	0.021 (0.082)	-0.008 (0.054)
Observation Count	3,264	2,406	2,332	2,337

Note: This table displays results from models assessing whether perceptions of TPPs predict the NCEES evaluation ratings (composite rating) of first-year teachers and second-year teachers. +, \*, and \*\* indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. Summative refers to the average of all items for teacher preparation quality or opportunities to learn; Instruction, Supportive Learning Environments, and Teaching Diverse Learners refer to construct measures identified through factor analysis.

**Table 5: Predicted Associations from the Perceptions of TPPs to Teacher Retention**

	Quality of Teacher Preparation		Opportunities to Learn	
	Returns for a 2 <sup>nd</sup> Year of Teaching	Returns for a 3 <sup>rd</sup> Year of Teaching	Returns for a 2 <sup>nd</sup> Year of Teaching	Returns for a 3 <sup>rd</sup> Year of Teaching
<i>Completer-Level Measures</i>				
Summative	<b>0.010*</b> (0.005)	<b>0.014*</b> (0.007)	<b>0.010+</b> (0.005)	0.009 (0.007)
Instruction	0.006 (0.005)	0.008 (0.007)	0.005 (0.005)	0.001 (0.007)
Supporting Learning Environment	0.006 (0.005)	0.005 (0.007)	<b>0.011*</b> (0.005)	<b>0.017*</b> (0.007)
Teaching Diverse Learners	0.005 (0.004)	<b>0.012+</b> (0.007)	0.002 (0.005)	0.001 (0.007)
Observation Count	3,482	2,720	2,519	2,643
<i>University-Level Measures</i>				
Summative	0.000 (0.005)	0.006 (0.007)	<b>0.006+</b> (0.003)	0.004 (0.007)
Instruction	0.000 (0.006)	<b>0.012+</b> (0.007)	<b>0.009*</b> (0.004)	<b>0.012+</b> (0.006)
Supporting Learning Environment	-0.002 (0.006)	-0.007 (0.006)	-0.002 (0.005)	-0.008 (0.007)
Teaching Diverse Learners	0.002 (0.005)	0.003 (0.006)	0.001 (0.004)	-0.002 (0.007)
Observation Count	3,482	2,720	2,519	2,643

Note: This table displays results from models assessing whether perceptions of TPPs predict the retention of first-year teachers and second-year teachers. +, \*, and \*\* indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. Summative refers to the average of all items for teacher preparation quality or opportunities to learn; Instruction, Supportive Learning Environments, and Teaching Diverse Learners refer to construct measures identified through factor analysis.

**Appendix Table 1: Response Rates by University on the NTPS**

<b>NC University</b>	<b>Response Rate</b>
NC University 1	37.41
NC University 2	46.67
NC University 3	52.82
NC University 4	56.52
NC University 5	64.86
NC University 6	50.00
NC University 7	49.38
NC University 8	53.03
NC University 9	57.14
NC University 10	27.27
NC University 11	46.56
NC University 12	62.07
NC University 13	50.55
NC University 14	46.83
NC University 15	46.51
NC University 16	40.00
NC University 17	37.14
NC University 18	44.33
NC University 19	38.14
NC University 20	50.60
NC University 21	43.33
NC University 22	53.25
NC University 23	65.51
NC University 24	47.46
NC University 25	39.34
NC University 26	20.00
NC University 27	55.17
NC University 28	44.69
NC University 29	52.54
NC University 30	44.09
NC University 31	42.68
NC University 32	50.00
NC University 33	45.13
NC University 34	45.28
NC University 35	43.33
NC University 36	41.10
NC University 37	43.90
NC University 38	37.96
NC University 39	39.39
NC University 40	41.46
NC University 41	33.74

**Appendix Table 2: Correlations between NTPS Measures and School Characteristics**

School Characteristics	Summative Measures		Perceptions of Teacher Preparation Quality			Opportunities to Learn		
	Quality	Opportunities to Learn	Instruction	Supportive Learning Environments	Teaching Diverse Learners	Instruction	Supportive Learning Environments	Teaching Diverse Learners
Economically-Disadvantaged	-0.005	-0.027	-0.023	-0.004	0.018	-0.029	-0.007	-0.014
Racial/Ethnic Minority	-0.005	-0.011	-0.001	-0.006	-0.003	-0.024	0.013	-0.006
Performance Composite	0.006	0.026	-0.011	<b>0.044**</b>	-0.012	0.022	0.023	0.007
Novice Teacher Percentage	-0.018	<b>-0.049**</b>	-0.003	-0.025	-0.009	<b>-0.048*</b>	-0.012	-0.023

Note: This table displays correlations between focal NTPS measures (at the individual teacher level) and school characteristics for first-year teachers. +, \*, and \*\* indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

**Appendix Table 3: Predicted Associations from the Perceptions of TPPs to Teacher Outcomes with University Fixed Effects**

	Quality of Teacher Preparation		Opportunities to Learn	
	1 <sup>st</sup> Year Teachers	2 <sup>nd</sup> Year Teachers	1 <sup>st</sup> Year Teachers	2 <sup>nd</sup> Year Teachers
<i>Completer-Level Measures: Teacher Value-Added</i>				
Summative	0.015 (0.020)	0.011 (0.023)	0.022 (0.025)	-0.001 (0.023)
Instruction	0.010 (0.019)	-0.008 (0.022)	0.021 (0.025)	-0.004 (0.024)
Supporting Learning Environment	<b>0.035<sup>+</sup></b> (0.019)	<b>0.046<sup>+</sup></b> (0.024)	<b>0.065<sup>**</sup></b> (0.024)	<b>0.061<sup>*</sup></b> (0.024)
Teaching Diverse Learners	-0.012 (0.019)	-0.016 (0.022)	-0.025 (0.024)	<b>-0.038<sup>+</sup></b> (0.022)
Observation Count	3,177	2,571	2,246	2,501
<i>Completer-Level Measures: Teacher Evaluation Ratings</i>				
Summative	<b>0.177<sup>**</sup></b> (0.036)	<b>0.189<sup>**</sup></b> (0.046)	<b>0.150<sup>**</sup></b> (0.045)	<b>0.138<sup>**</sup></b> (0.047)
Instruction	<b>0.063<sup>+</sup></b> (0.038)	<b>0.081<sup>+</sup></b> (0.046)	0.034 (0.045)	0.036 (0.048)
Supporting Learning Environment	<b>0.173<sup>**</sup></b> (0.039)	<b>0.211<sup>**</sup></b> (0.045)	<b>0.122<sup>**</sup></b> (0.043)	<b>0.196<sup>**</sup></b> (0.049)
Teaching Diverse Learners	<b>0.083<sup>*</sup></b> (0.036)	0.051 (0.047)	<b>0.122<sup>**</sup></b> (0.044)	0.054 (0.049)
Observation Count	3,264	2,406	2,332	2,337
<i>Completer-Level Measures: Teacher Retention</i>				
Summative	<b>0.010<sup>*</sup></b> (0.005)	<b>0.013<sup>+</sup></b> (0.007)	<b>0.009<sup>+</sup></b> (0.005)	0.008 (0.007)
Instruction	0.006 (0.005)	0.006 (0.007)	0.004 (0.005)	-0.002 (0.007)
Supporting Learning Environment	0.006 (0.005)	0.005 (0.007)	<b>0.011<sup>*</sup></b> (0.005)	<b>0.020<sup>*</sup></b> (0.008)
Teaching Diverse Learners	0.006 (0.005)	<b>0.013<sup>+</sup></b> (0.007)	0.002 (0.005)	0.001 (0.007)
Observation Count	3,482	2,720	2,519	2,643

Note: This table displays results from models assessing whether perceptions of TPPs predict the EVAAS estimates, NCEES evaluation ratings, and retention of first-year teachers and second-year teachers. All models include a university fixed effect. +, \*, and \*\* indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. Summative refers to the average of all items for teacher preparation quality or opportunities to learn; Instruction, Supportive Learning Environments, and Teaching Diverse Learners refer to construct measures identified through factor analysis.