DATA SHARING TO DRIVE THE IMPROVEMENT OF TEACHER PREPARATION PROGRAMS

Kevin C. Bastian, UNC–Chapel Hill

C. Kevin Fortner, Georgia State University

Alisa Chapman, UNC–General Administration

Jayne Fleener, North Carolina State University

Ellen McIntyre, UNC–Charlotte

Linda Patriarca, East Carolina University

EPIC

EDUCATION POLICY
INITIATIVE *at* CAROLINA

Data Sharing to Drive the Improvement of Teacher Preparation Programs

Kevin C. Bastian
University of North Carolina at Chapel Hill
Education Policy Initiate at Carolina (EPIC)

C. Kevin Fortner
Georgia State University
Education Policy Studies

Alisa Chapman
University of North Carolina General Administration
Academic and University Programs

Jayne Fleener
North Carolina State University
College of Education

Ellen McIntyre
University of North Carolina at Charlotte
College of Education

Linda Patriarca
East Carolina University
College of Education

Contents

Acknowledgements

Data Sharing to Drive the Improvement of Teacher Preparation Programs

Kevin C. Bastian, UNC-Chapel Hill
C. Kevin Fortner, Georgia State University
Alisa Chapman, UNC-General Administration
Jayne Fleener, North Carolina State University
Ellen McIntyre, UNC-Charlotte
Linda Patriarca, East Carolina University

January 2015

Abstract

Teacher preparation programs (TPP) face increasing pressure to improve the quality of their practices and graduates yet do not often possess the level of data necessary to make evidence-based reforms. Therefore, we call for the establishment of partnerships between TPP and researchers or state education agencies and the sharing of *individual-level* data with TPP. Individual-level data sharing would allow TPP to develop systems of continuous improvement by examining whether their preparation practices align with the types of environments in which their graduates teach and how graduates' preparation experiences predict their characteristics and performance as teachers-of-record. To illustrate the potential of individual-level data sharing, we describe the data being shared with TPP in the University of North Carolina (UNC) system, how these data can lead to program improvement, the challenges TPP must overcome to effectively use individual-level data, and how UNC TPP are using these data for program improvement.

Introduction

In recent years accreditation agencies and policymakers have initiated efforts to both hold teacher preparation programs (TPP) accountable for the performance of their graduates and push TPP to make evidence-based reforms. For example, the newly formed Council for the Accreditation of Educator Preparation (CAEP) requires TPP to demonstrate the impact of their graduates on student learning, classroom instruction, and employer satisfaction and to institute a system of data analysis and continuous improvement (CAEP, 2013). Likewise, the U.S. Department of Education recently announced plans to rate TPP based on their graduates' job placement rates, retention rates, and effectiveness and surveys of their graduates and their employers (Rich, 2014). While these efforts correctly recognize teachers' significant effects on student outcomes (Bill and Melinda Gates Foundation, 2013; Nye, Konstantopoulos, & Hedges, 2004) and the importance of teacher preparation to teacher performance (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Darling-Hammond, Chung, & Frelow, 2002; Goldhaber, Liddle, & Theobald, 2013; Henry, Purtell, Bastian, Fortner, Thompson, Campbell, & Patterson, 2014), initiatives to hold TPP accountable for the performance of their graduates often leave an important question unanswered: With what data can TPP best make evidence-based reforms?

As detailed in a recent National Academy of Education report, evaluations of TPP typically serve a primary purpose—either accountability, providing information to consumers, or program improvement—and the evaluation data required for one purpose may not be well-aligned with the evaluation data required for another purpose (Feuer, Floden, Chudowsky, & Ahn, 2013). Many current TPP evaluations, such as estimating the average value-added of a TPP's graduates (Gansle, Noell, & Burns, 2012; Goldhaber, Liddle, & Theobald, 2013; Henry, Patterson, Campbell, & Pan, 2013) or rating the quality of a TPP's inputs (National Council on Teacher Quality, 2014), fall into the accountability and/or consumer information categories. When performed well, these evaluation efforts benchmark the performance of a TPP against a reference category or a set of standards and may direct TPP to look towards high-performing or highly-rated TPP for program improvement ideas.[1] However, even with these types of *aggregate* evaluation data, TPP are often driving blind, operating without the level of data necessary to guide evidence-based program improvement (Peck, Singer-Gabella, Sloan, & Lin, 2014).

Instead, to initiate systems of continuous improvement, TPP and researchers or state-level education agencies need to establish partnerships so that TPP receive *individual-level* data on the characteristics, work environments, and performance of their graduates. Such data could include teachers' credentials (e.g. National Board Certification status and licensure exam scores), measures of their employment/teaching context (e.g. school free and reduced-price lunch percentage, students' average prior scores, and the percentage of English language learners taught), and their outcomes (e.g. value-added estimates, evaluation ratings, and retention). With

---

[1] For example, TPP performing at average or below average levels, based on the value-added of their graduates, can look to TPP with highly effective graduates to try to identify and replicate promising preparation practices.

such individual-level data TPP can better achieve evidence-based program improvement by examining whether their preparation practices are aligned with the types of school and classroom environments in which their graduates teach and by exploring how variation in graduates' preparation experiences explain variation in the characteristics and performance of those graduates when they become teachers. Given the research showing significant within-program heterogeneity in graduates' value-added effectiveness, these types of analyses represent a promising way to better understand that variability (Koedel, Parsons, Podgursky, & Ehlert, 2012). Furthermore, such data sharing partnerships can help TPP develop the internal capacity for rigorous data analysis, determine what additional data measures they should collect to advance program improvement, and create a coordinated and systemic view of teacher education reform (Cochran-Smith & BCET, 2009; Peck & McDonald, 2014).

To illustrate the need for individual-level data sharing with TPP, we begin by detailing the strengths and shortcomings of accountability-based TPP evaluation systems. Responding to the shortcomings of these evaluation efforts, we then discuss the creation of a data sharing initiative in North Carolina. Specifically, we focus on the individual-level data being shared with TPP in the University of North Carolina (UNC) system, how this data can lead to program improvement, and the obstacles data sharing must overcome to achieve its potential. Finally, to better understand how TPP can use individual-level data on their graduates to drive evidence-based decision-making, deans from three UNC system TPP share how they are using the data sharing initiative to guide program improvement.

Overall, TPP face strong incentives to improve the quality of their preparation practices, and subsequently, the quality of their graduates. Doing so, however, will require more than accountability-based evaluations of TPP; as a first step, it will require providing TPP with the resources—the data—to make evidence-based decisions.

Strengths and Shortcomings of Accountability-Based TPP Evaluation Systems

Over the last decade school districts and states, such as New York City, Louisiana, North Carolina, Tennessee, and Washington, have initiated efforts to estimate teachers' value-added to student achievement and link teachers' value-added scores to their preparation (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006, 2009; Gansle, Noell, & Burns, 2012; Goldhaber, Liddle, & Theobald, 2013; Henry, Purtell, Bastian, Fortner, Thompson, Campbell, & Patterson, 2014; TSBE, 2012, 2013). At a high level, these efforts have asked whether teachers entering the profession through different routes are more or less effective than their peers entering with other forms of preparation or certification. For example, work by Boyd and colleagues in New York City compared the effectiveness of college recommended teachers with that of teachers entering New York schools through five additional categories (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006); comparable work in North Carolina assessed the effectiveness of teachers prepared at UNC system institutions with that of teachers entering the profession through 10 other portals (Henry, Purtell, Bastian, Fortner, Thompson, Campbell, & Patterson, 2014). More

narrowly, these efforts have focused on graduates of individual TPP and have asked whether they are more or less effective than graduates from other TPP. For example, research in Louisiana and Washington indicates that there is a substantial degree of overlap in the value-added effectiveness of TPP graduates but that some programs' graduates significantly outperform their peers from other programs (Gansle, Noell, & Burns, 2012; Goldhaber, Liddle, & Theobald, 2013).

Overall, these accountability-based research efforts provide a broad perspective on the effectiveness of teachers with different forms of preparation and allow individual TPP to both see the effectiveness of their graduates, in aggregate, and identify particular grade-levels or subject-areas in which their graduates are high (low) performing. Further, these accountability-based evaluations document the significant heterogeneity in the effectiveness of novice teachers with the same type of preparation (route or program), suggesting the need for continued research to help explain that variability (Kane, Rockoff, & Staiger, 2008; Koedel, Parsons, Podgursky, & Ehlert, 2012). The benefits of these accountability-based research efforts to TPP are two-fold. First, these studies show TPP how they fare on outcomes that are of interest to policymakers and the general public; this accountability may encourage or force TPP to focus on program improvement. Second, these studies may help TPP become more aware of and ready to use research evidence to inform program decisions and may make TPP leadership and faculty/staff better consumers of research findings. These benefits, in turn, highlight the key weakness of accountability-based TPP evaluations: the inability of such evaluations to formatively drive TPP reforms. For example, current analyses of TPP effectiveness only identify which programs' graduates are performing well or poorly; they do not pinpoint why or suggest changes programs can make to improve performance (Henry, Patterson, Campbell, & Pan, 2013). Therefore, while accountability-based TPP evaluations serve an important role, they are not sufficient to inform program improvement efforts. Instead, TPP need access to individual-level data on program graduates to establish systems of continuous improvement and make evidence-based reforms.

Data Sharing Initiative

Accountability pressures from policymakers and practitioners have pushed theories of evidence-based reform into a wide variety of fields and professions in recent years (Achenbach, 2005; Estabrooks, 2007; National Research Council, 2002). In teacher education, this effort is exemplified by Cochran-Smith's (2005) challenge to build "chains of evidence" linking teacher education pedagogy and program design with meaningful candidate learning and the Teachers for a New Era (TNE) initiative, which sought to achieve significant program reform through a respect for evidence (Fallon, 2006). Building from such initiatives and our own TPP evaluation work within the UNC system, data sharing represents an important next step in evidence-based reform by providing TPP the individual-level data they need to connect measures of candidates' preparation experiences to their characteristics and performance as teachers. Below, we detail the individual-level data being shared with UNC system TPP, the theory of change linking data sharing to program improvement, and the obstacles that may prevent data sharing from improving teacher education.

Data Shared with UNC System TPP

Data sharing is an initiative designed to stimulate a culture of evidence and program reforms by providing TPP with individual-level data on their graduates. We are sharing individual-level data with UNC system TPP in five broad categories: (1) teacher employment; (2) teacher characteristics; (3) classroom characteristics; (4) school characteristics; and (5) teacher outcomes. Specifically, we are providing TPP with separate data files, per academic year (currently 2005-06 through 2012-13), with each file containing data on all the individuals who were initially prepared to teach by a given TPP and employed as teachers in North Carolina public schools in that academic year. Furthermore, because teachers can work at more than one school in an academic year, files contain observations for each unique teacher-school combination. Below, we detail the data provided in each of these five categories and briefly consider questions TPP can ask with such data. Table 1 provides a list of the variables we are providing to TPP; Appendix Table 1 includes a description of the variables.

Table 1: Individual-Level Data Shared with the UNC System Teacher Preparation Programs

Employment Status	Teacher Characteristics	Classroom Characteristics	School Characteristics	Teacher Outcomes
- District and school - Number of pay periods - First pay period - Last pay period - Amount of time worked (full-time equivalency status)	- Teaching experience - Graduate degree status - National Board Certification status - Licensure areas - Licensure basis - Exams taken - Exam scores - Teaching a tested-grade/subject-area	- Number of classes taught - Average class size - Grade level(s) taught - Subject-area(s) taught - Race/ethnicity proportions - Free and reduced-price lunch proportions - Gifted proportion - Disabled proportion - Limited English Proficient proportion - Average days absent - Average prior achievement scores - Average prior achievement level	- Urbanicity - School size - Percentage free and reduced-price lunch - Short-term suspension rate - Violent acts rate - Race/ethnicity percentages - Total per-pupil expenditures - Per-pupil expenditures in spending categories (e.g. regular instruction) - AYP percentage - NC accountability status and growth - Performance composite - Teacher credentials—percentage fully-licensed, novice, holding an advanced degree or NBC - Pupil to teacher ratio - Teacher stay ratio	- Returns to NC public schools - Returns to the same school - Teacher value-added estimate (across 10 separate subject-areas) - Quintile for value-added estimate

Note: We are providing TPP with separate data files per academic year (beginning in 2005-06), with each file containing data on all the individuals who were initially prepared to teach by a given TPP and employed as teachers in that academic year. Files contain observations for each unique teacher-school combination.

Teacher Employment Data: The variables in this category include the district and school in which a teacher was employed, the number of pay periods a teacher worked in a given school, the first and last pay periods a teacher worked in a given school, and how much—the full-time equivalency (FTE)—a teacher worked in a given school and across all schools. With such data TPP can know: (1) whether and how quickly their graduates secure teaching jobs in a state's

public schools; (2) whether their graduates were hired after the start of the school year or exited teaching during the middle of the year; (3) the nature of the employment as full or part-time; and (4) which districts and schools hire their graduates and whether their graduates work in close proximity to the TPP.

Teacher Characteristics: The variables in this category include a teacher's level of experience, whether a teacher holds a graduate degree or National Board Certification, the licensure areas a teacher holds and the basis for those teaching licenses (e.g. from an in-state preparation program or a reciprocal license from out-of-state), the tests (e.g. Praxis II Middle School Mathematics, SAT) a teacher has taken and a teacher's score on those exams, and whether a teacher teaches in a tested-grade/subject. With such data TPP can know: (1) whether and in what areas their graduates have earned additional teaching licenses after graduation; (2) how well their graduates scored on licensure exams linked to their teacher preparation; (3) whether their graduates have secured additional credentials—graduate degrees or National Board Certification—after graduation; and (4) which of their graduates face accountability pressure as a tested-grades/subject teacher.

Classroom Characteristics: The variables in this category include the number of classes taught by a teacher in an academic year, the average size of those classes, the subjects and grades taught by a teacher in an academic year, the average prior performance of a teacher's students on End-of-Grade and/or End-of-Course exams, the average number of days absent for a teacher's students, and the proportion of a teacher's students who are white, black, Hispanic, American Indian, qualify for free or reduced-price lunches, currently are or were Limited English Proficient, and receive gifted or exceptional children services. With such data TPP can know: (1) the teaching load of their graduates; (2) whether their graduates are teaching in-field or out-of-field and whether the TPP prepared them to teach in their current subject/grade area(s); (3) whether their graduates instruct low, average, or high performing students; and (4) whether their graduates teach classes with high percentages of students who are minority, economically-disadvantaged, non-native English speakers, or exceptional.

School Characteristics: The variables in this category include the number of students enrolled at the school, the urbanicity status of a school, measures of a school's orderliness (the number of suspensions and violent acts), the racial/ethnicity percentages of a school's students, the percentage of a school's students qualifying for free or reduced-price lunch, measures of a school's academic performance (accountability status and growth and the percentage of students passing standardized exams), total per-pupil expenditures and per-pupil expenditures in key spending categories (e.g. regular instruction, special instruction, school leadership), and measures of teachers' persistence and credentials at a school (the proportion of teachers who returned from the previous year and who are fully licensed, novice, Nationally Board Certified, or holding a graduate degree). With such data TPP can know: (1) whether their graduates teach in safe and orderly environments; (2) whether their graduates teach in schools with high percentages of minority or economically-disadvantaged students; (3) whether their graduates teach in low,

average, or high-performing schools; (4) the financial resources available in the schools where their graduates teach; and (5) whether their graduates teach in schools with high amounts of turnover and with better or less well-credentialed peers.
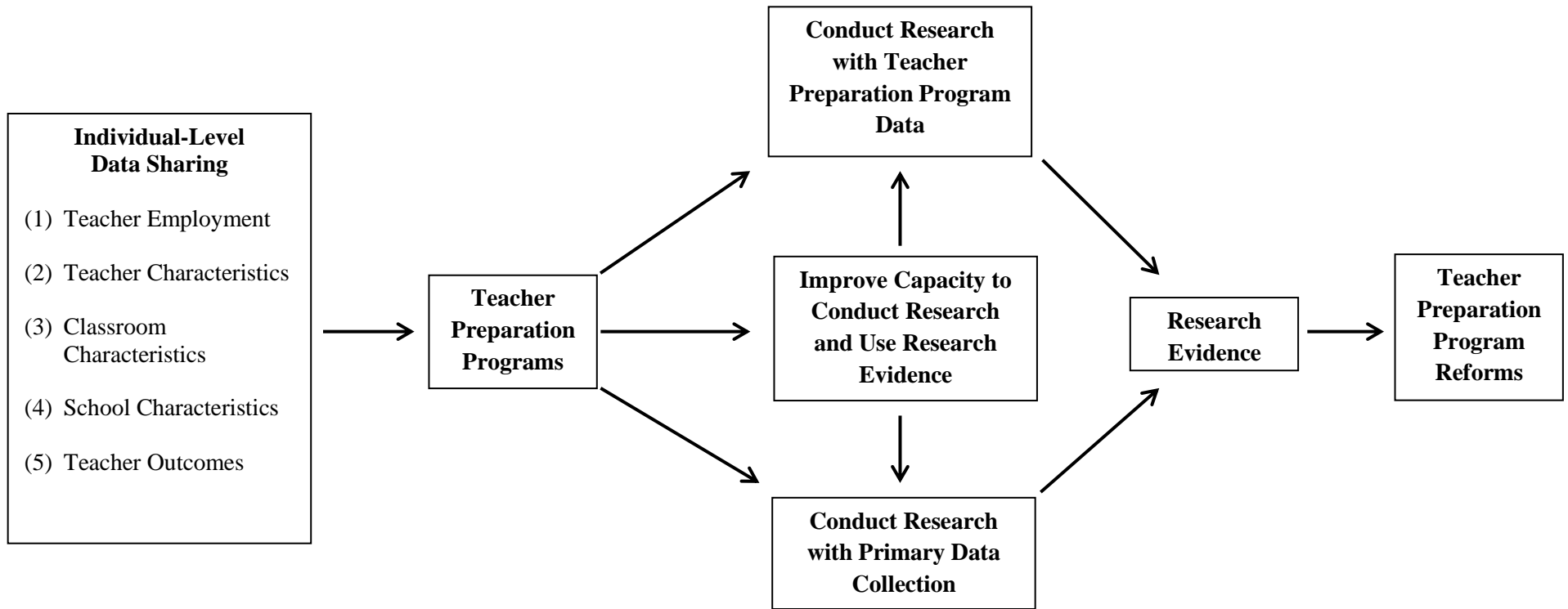
Teacher Outcomes: The variables in this category include indicators for whether a teacher returns to North Carolina public schools in the following year and the same school in the following year, estimates of individual teacher value-added across ten different subjects/grade-levels—elementary mathematics, reading, and science, middle grades mathematics, reading, and science, and high school mathematics, English, science, and social studies[2]—and the quintile for each value-added estimate. With such data TPP can know: (1) the persistence of the teachers they prepare; (2) how effective their graduates are at promoting student achievement gains; and (3) the relative effectiveness of their graduates compared to peers teaching the same level/subject-area.

How Data Sharing Can Lead to TPP Improvement

Teacher preparation programs can use individual-level data on program graduates to leverage program improvement in three ways: (1) conducting research with shared data and indicators of TPP progress and performance; (2) conducting research with shared data and primary data collected by TPP faculty and staff; and (3) improving the capacity of TPP faculty and staff to conduct research and think strategically about data use. Importantly, TPP can tailor the use of this individual-level data to the particular needs, elements, and unique questions facing their program. Below, we describe the processes connecting individual-level data to program improvement; Figure 1 provides a visual depiction of this theory of change.

---

[2] We estimate individual teacher value-added using a three level (student, teacher, school) hierarchical linear model with a rich set of student, teacher/classroom, and school covariates. In this model the teacher effectiveness estimate is the random effect from the second (teacher) level of analysis.

Figure 1: How Individual-Level Data Sharing Can Lead to Teacher Preparation Program Improvement

```
                                        ┌──────────────────────┐
                                        │   Conduct Research   │
                                        │    with Teacher      │
                                        │ Preparation Program  │
                                        │        Data          │
                                        └──────────────────────┘

┌────────────────────────┐                                                              ┌──────────────┐
│   Individual-Level     │                   ┌──────────────────────┐                   │   Teacher    │
│    Data Sharing        │    ┌────────────┐ │ Improve Capacity to  │  ┌────────────┐   │ Preparation  │
│                        │    │  Teacher   │ │  Conduct Research    │  │  Research  │   │   Program    │
│ (1) Teacher Employment │    │Preparation │ │ and Use Research     │  │  Evidence  │   │   Reforms    │
│                        │    │ Programs   │ │     Evidence         │  └────────────┘   └──────────────┘
│ (2) Teacher Characteristics│ └────────────┘ └──────────────────────┘
│                        │
│ (3) Classroom          │                   ┌──────────────────────┐
│     Characteristics    │                   │   Conduct Research   │
│                        │                   │  with Primary Data   │
│ (4) School Characteristics│                │     Collection       │
│                        │                   └──────────────────────┘
│ (5) Teacher Outcomes   │
└────────────────────────┘
```

Research Studies with TPP Data: Teacher preparation programs collect and store a wide range of data on their teacher candidates. For instance, TPP typically measure candidates': high school academic performance, SAT/ACT and Praxis I scores, collegiate GPA, courses taken and the sequence of courses (e.g. number of content courses, number of pedagogy courses), university personnel serving in instructor or advisor roles, ratings across dispositional, portfolio, and student teaching instruments, and responses to program exit surveys. Many TPP also use teacher performance assessments (e.g. edTPA) to measure their candidates' readiness to enter the profession. Collectively, this is a wealth of internal (TPP) data capturing the experiences and performance of teacher candidates. To drive program improvement, TPP can combine this internal data with externally-provided, individual-level data on their program graduates to examine (1) whether their graduates' preparation experiences are aligned with the types of schools and classrooms in which they work and (2) how variation in graduates' programmatic components or performance predict variation in their outcomes (e.g. entry into or exit from the workforce, earning advanced credentials, teacher value-added).

For example, through analysis of the shared individual-level data, a TPP may discover that their recent elementary grades graduates are teaching in classrooms with many English language learners (ELLs). In response, the TPP could design and require additional learning segments or courses to provide candidates the knowledge and skills to succeed with ELLs. Likewise, after examining the relationships between program data and individual teacher value-added, a TPP may find that, on average, those graduates who received instructional coaching during student teaching are more effective than others without such an experience. As a result, the TPP could provide instructional coaching to all student teachers. Importantly, the TPP could also use their internal data to examine competing hypotheses for this result—e.g. the higher value-added was due to these graduates' higher GPAs.

Research Studies with Primary Data Collection: In addition to their extant administrative data, TPP can use the individual-level shared data as the impetus for primary data collection initiatives to better understand the performance and perspectives of their graduates. These primary data collection efforts could focus on classroom observations (e.g. general protocols, such as the Framework for Teaching (Danielson, 2013), or content-specific protocols, such as Mathematical Quality of Instruction), interviews/focus groups, or teacher surveys. For instance, if a TPP wanted to assess why some of their middle grades mathematics graduates generated significantly larger student achievement gains than other middle grades graduates, the TPP could: (1) use the individual-level shared data to identify their graduates in the top and bottom quintiles of effectiveness;[3] (2) observe those teachers with a classroom observation protocol (in which the observer is blind to the teacher's prior effectiveness); and (3) administer surveys to examine these graduates' perceptions of preparation quality (Hill, Kapitula, & Umland, 2011). Analyses with such data may reveal that a TPP's highly-effective middle grades mathematics

---

[3] Due to the potential for bias and measurement error in individual teacher value-added estimates, TPP should use multiple years of student test score data to identify graduates in the top and bottom quintiles of effectiveness.

graduates better engage their students in meaning-making and mathematical reasoning and more clearly articulate mathematical ideas. With this evidence, TPP faculty and staff may design and require an additional course in pedagogical content knowledge to improve teacher performance.

Improving TPP Capacity: Beyond the direct support of research, individual-level data sharing can also lead to program improvement by improving the capacity of TPP faculty and staff to conduct research and strategically use research evidence. Quite simply, the provision of individual-level data should give TPP faculty and staff opportunities to: (1) develop research questions; (2) determine the required data, sample, and analytical methods to answer those questions; (3) interpret results; and (4) consider beneficial programmatic changes in response to research findings. As a result of this capacity building, TPP can create/strengthen a culture of evidence and a coordinated, systemic view of TPP reform (Peck & McDonald, 2014).

Challenges to Program Improvement with Data Sharing

While individual-level data sharing has the potential to drive evidence-based program improvement, TPP must overcome research-based and organizational challenges for data sharing to fulfill its promise. Below, we detail some of these challenges and introduce ways that TPP can meet these obstacles.

Research-Based Challenges: Unless TPP possess the internal capacity to conduct rigorous research analyses, they cannot fully leverage individual-level data to make evidence-based reforms. Here, internal capacity starts with TPP collecting measures of teacher candidate progress and performance that have predictive validity—associated with teachers' performance after beginning teaching (Henry, Campbell, Thompson, Patriarca, Luterbach, Lys, & Covington, 2013). This requirement may be problematic for TPP because many of the measures they currently collect are meant to determine whether teacher candidates meet a competency threshold, rather than distinguish between the performance of teacher candidates, and therefore, do not have the sufficient variation needed for analyses. Without such measures, it may be challenging for TPP to identify programmatic components in need of reform. For TPP that do not have these measures, however, individual-level data sharing can help determine that current data instruments are not predictive and push these programs to develop/begin using additional measures (Henry, Campbell, Thompson, Patriarca, Luterbach, Lys, & Covington, 2013).

Beyond data measures, individual-level data sharing requires that TPP have a robust data management system that allows them to connect the program-level data they collect on teacher candidates to the characteristics and outcomes data shared by researchers or state education officials. This means that TPP need a data management platform that stores measures of teacher candidate progress and performance over a number of years, a unique identification number for candidates/graduates—to connect separate elements of program-level data to externally provided data—and protocols established to securely handle sensitive information.

With such data structures in place, the next capacity concern is the extent of faculty expertise at TPP to conduct rigorous research analyses. At many TPP the primary focus of

faculty is preparing teacher candidates and only recently has a stronger research focus developed. This means that there may only be a small number of faculty members with the ability and interest required to effectively analyze the shared data, and as a result, the timeliness and breadth of research may be limited. To address these capacity concerns, TPP can collaborate with researchers in other schools/departments of their respective institutions and across institutions. If TPP received governmental or philanthropic financial support for evidence-based program improvement efforts, they could hire research coaches who would work with TPP to create and improve data systems, develop a research agenda, analyze data, and most importantly, build the internal capacity of TPP to independently conduct analyses.

The last research-based challenge concerns the small size of many TPP and whether there is sufficient statistical power to predict significant differences in outcomes for program graduates. Quite simply, insufficient statistical power may limit the ability of TPP to make evidence-based reforms because the evidence does not meet a threshold—statistical significance—for taking action. In response to this concern, TPP can increase sample size by pooling individual-level data from multiple graduating cohorts or, when feasible, pooling data with other TPP that are conducting similar analyses. More broadly, TPP can re-evaluate standards for what makes research evidence actionable. P-values from correlations or regression coefficients that minimize the likelihood of Type I errors (such as those less than 0.05) provide the strongest basis for evidence-based reform; however, to minimize the possibility of Type II errors and respond to findings that suggest a practically significant relationship, TPP can relax standards for designating research findings as actionable. While there must be continued scrutiny, to reduce the likelihood that TPP make programmatic changes that are not supported by evidence, this approach will also reduce the likelihood that TPP miss out on promising opportunities for reform.

Organizational Challenges: Even with the research capacity to leverage individual-level data, TPP cannot make evidence-based reforms without creating or supporting a "culture of evidence" amongst faculty, supervisors, and staff (Peck, Gallucci, Sloan, & Lippincott, 2009; Peck & McDonald, 2014). Essentially, TPP have to establish the collective values and institutional policies that recognize the importance of individual-level data (acquiring, analyzing, and using it for decision-making) and shift the conception of program reform away from disconnected changes made by single faculty members to coordinated and systemic efforts to improve recruitment/selection, curricular, and clinical practices in response to research evidence (Cochran-Smith & BCET, 2009; Peck, Gallucci, & Sloan, 2010; Peck & McDonald, 2014). To cultivate this culture of evidence, TPP faculty and staff must have an interest in program improvement—borne out of academic inquiry or concerns about what candidates take up from preparation experiences—and must view the shared data as valid and relevant to their practice (Peck & McDonald, 2014). Building a culture of evidence will impact the work of TPP faculty and staff and how faculty and staff view their work—as part of a larger, collective enterprise to improve the preparation of teacher candidates.

Perspectives from TPP

In the sections below, College of Education deans from three UNC system TPP share the research agenda they are pursuing with the individual-level data. While beneficial as stand-alone descriptions of evidence-based program reform, these perspectives can also serve a broader purpose as templates for other TPP considering programmatic changes.

North Carolina State University

STEM (science, technology, engineering, and mathematics) education is an area of emphasis at North Carolina State University (NCSU) and preparing STEM teachers with strong backgrounds in content and pedagogy is central to NCSU's mission to support a STEM teacher pipeline. Sustaining this pipeline is particularly important since outstanding STEM teachers are a key to preparing and motivating K-12 students to pursue post-secondary STEM opportunities (President's Council of Advisors on Science and Technology, 2010).

To strengthen STEM teacher education programs, NCSU is using the individual-level data to conduct drill-down studies examining the school placements of STEM graduates, the courses STEM graduates go on to teach, and the performance of elementary STEM graduates across STEM and non-STEM (English language arts) subjects. Specifically, NCSU is addressing the following sets of research questions:

(1) In comparison to state averages and non-STEM graduates, what are the characteristics of the schools in which NCSU's STEM graduates teach?

(2) What types of courses do NCSU's STEM graduates teach? Are they teaching advanced courses, such as calculus and physics, or regular courses, such as algebra and biology?

(3) Do NCSU's elementary education graduates have higher levels of content knowledge (as measured by licensure exams)? Are NCSU's elementary education graduates more effective mathematics and science teachers? Does NCSU's STEM-focus in elementary education compromise the performance of their graduates in English language arts?

The first question helps NCSU better align its coursework and student teaching placements to the types of students and schools STEM graduates encounter and allows NCSU to create a closer partnership with its most outstanding STEM graduates. The second question assists NCSU in aligning and setting content-area requirements and identifies which STEM graduates—those who were higher or lower performing as teacher candidates—teach a tested-grade/subject-area and have value-added data. Finally, NCSU recently created a STEM-focused elementary education program with high levels of STEM content-area requirements. The final set of questions allows NCSU to know whether this content focus produces graduates with higher levels of content knowledge, graduates who are more effective mathematics and science teachers, and graduates who are also effective in non-STEM subjects. Such data will help NCSU make informed decisions about the direction of its new elementary education program.

University of North Carolina Charlotte

UNC Charlotte (UNCC) is a large urban research institution with an explicit mission to prepare teachers for urban environments, as well as the surrounding rural and suburban school districts, with a focus on equity, excellence, and engagement with the community. The driving force behind UNCC's desire to use the individual-level data is the prior system studies illustrating that UNCC's elementary education graduates' value-added falls into the low or middle ranges of the value-added scores of all institutions in the UNC system.

To assess whether the college is fulfilling its mission and to understand why UNCC's elementary program graduates scored lower than expected, researchers at UNCC are employing the shared individual-level data to answer the following groups of research questions:

(1) Are UNCC's graduates more likely to teach in urban settings than graduates of system universities who do not have the same mission? How long do the teachers in urban settings stay in those settings? How effective are the teachers serving high-poverty populations? How effective are teachers who serve large populations of ELLs?

(2) Are UNCC's elementary program graduates' value-added scores predicted by entry characteristics (high school GPA, SAT scores, dispositions)?

(3) How do scores on mathematics and reading content licensure exams predict UNCC's elementary graduates' value-added scores? How do elementary graduates' course-taking patterns predict value-added scores? How do scores on key assignments during professional preparation predict elementary graduates' value-added scores?

The first group of questions assesses how well UNCC is addressing its stated mission of preparing professionals for challenging environments. Findings could have implications for reexamining the mission or program components to better meet the mission. The second set of questions assesses the relationship between candidate content knowledge or human capital and resulting student achievement, which could have implications for candidate recruitment and selection. Finally, the third set of questions assesses the relationship between and among the candidates' content knowledge, value-added scores, and program features. Answers to these questions will entail key implications for program faculty, as they grapple with how to better prepare elementary education candidates.

East Carolina University

In recent years East Carolina University (ECU), a pilot institution for edTPA in North Carolina and a recipient of a U.S. Department of Education Teacher Quality Partnership grant, has made significant efforts to implement evidence-based program reforms and evaluate the efficacy of these program revisions. To further this commitment to continuous improvement, ECU is using the shared individual-level data to pursue the following research questions:

(1) How much variance is in the value-added scores of ECU's graduates?

(2) What is the relationship between ECU graduates' value-added scores and the following: entry characteristics (e.g. GPA, test scores); academic major/concentration; the personnel training graduates (e.g. instructors, clinical teachers, university supervisors); the number, type, and length of graduates' clinical practice opportunities; the number, type, and scores of graduates' formative and summative program assessments; and the four-year GPA and Praxis II scores of graduates?

(3) What are the patterns in ECU graduates' attrition and changes in position (e.g. changing grades/subject-areas or schools/districts)?

The first question, assessing the variance in graduates' effectiveness, is a key consideration for ECU (and TPP generally) because if there is a large spread between the more and less effective program graduates, then ECU must address tough questions, such as: are program assessments valid and reliable, how consistently do faculty and staff monitor and grade candidate knowledge and skills, and how rigorous are the standards for demonstrating basic competency during student teaching? The second set of questions assesses the relationship between graduates' effectiveness and indicators of candidates' progress and performance. The goal of these analyses is to find patterns in the data that will inform program innovation. Finally, given that teacher mobility may adversely impact students, schools, and the teachers themselves—due to an inability to establish a collaborative and supportive group of practice, inconsistencies and gaps in induction/mentoring, and the lack of experience teaching a particular grade/subject-area—answers to the third question will provide insight into the types of support needed during teachers' induction period and identify gaps in ECU's preparation that may be contributing to graduates' early-career struggles.

Discussion

Recent efforts to hold TPP accountable for the performance of their graduates recognize the importance of teachers to student outcomes and the importance of teacher preparation to teacher performance (Bill and Melinda Gates Foundation, 2013; Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Goldhaber, Liddle, & Theobald, 2013; Nye, Konstantopoulos, & Hedges, 2004). If the goal of these accountability initiatives is to identify high and low-performing TPP, as measured by value-added scores, and put pressure on programs to reform, then the aggregate level data —e.g. the average value-added of a program's graduates— generated during these evaluations is sufficient for this purpose. However, if teacher educators want to develop a culture of evidence, create systems of continuous improvement, and adopt/adapt evidence-based program reforms, then TPP need individual-level data on the characteristics, work environments, and performance of their graduates. With such data TPP can assess whether graduates' preparation experiences are aligned with the types of schools and classrooms in which they work and how variation in graduates' programmatic components predicts teacher outcomes. Without such data, TPP may implement program revisions with no indication of whether or not those changes will improve the performance of program graduates.

Essentially, individual-level data sharing allows TPP to exercise greater agency in the program improvement process.

Despite its promise, individual-level data sharing cannot improve the quality of program practices and graduates if TPP do not possess additional resources to turn shared data into actionable evidence. Specifically, TPP need valid and reliable measures of teacher candidate progress and performance, robust data management systems, sufficient research expertise, and a faculty and staff committed to a coordinated and systemic view of TPP reform. For programs without such research capacity, TPP may need financial support from government agencies, teacher education groups, and/or philanthropic organizations. With such funds TPP could hire research coaches whose charge would include creating and improving TPP data systems; helping TPP specify a research agenda, analyze data, and interpret results; and most importantly, building the internal capacity of TPP to independently conduct analyses. Furthermore, TPP could use such funds to hold research conferences—where TPP came together to discuss their research activities and findings—and organize inter-institutional research collaborations. These types of structures would help TPP use the shared data to produce results and formulate/enact program reforms. Overall, individual-level data sharing represents a promising initiative to improve the quality of preparation practices and graduates; we call for the establishment of partnerships between TPP and researchers/state education agencies and the sharing of individual-level data with TPP.

References

Achenbach, T.M. (2005). Advancing assessment of children and adolescents: Commentary on evidence-based assessment of child and adolescent disorders. *Journal of Clinical Child and Adolescent Psychology, 34*(3), 541-547.

Bill and Melinda Gates Foundation. (2013). Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET project's three-year study. Available from: http://www.metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy, 1*(2), 176-216.

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.

Cochran-Smith, M. (2005). Teacher education and the outcomes trap. *Journal of Teacher Education, 56*(5), 411-417.

Cochran-Smith, M., & the Boston College Evidence Team. (2009). "Re-culturing" teacher education: Inquiry, evidence, and action. *Journal of Teacher Education, 60*(5), 458-468.

Council for the Accreditation of Educator Preparation. (2013). CAEP Accreditation Standards. Available from: http://caepnet.files.wordpress.com/2013/09/final_board_approved1.pdf

Danielson, C. (2013). *The framework for teaching evaluation instrument, 2013 edition.*

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.

Estabrooks, C. (2007). A program of research on knowledge translation. *Nursing Research, 56*(4), 4-6.

Fallon, D. (2006). Improving teacher education through a culture of evidence. Paper presented at the sixth annual meeting of the Teacher Education Accreditation Council, Washington, DC. Available from: http://www.teac.org/membership/meetings/Fallon%20remarks.pdf.

Feuer, M.J., Floden, R.E., Chudowsky, N., & Ahn, J. (2013). *Evauation of teacher preparation programs: Purposes, methods, and policy options.* Washington, DC: National Academy of Education.

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.

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.

Henry, G.T., Campbell, S.L., Thompson, C.L., Patriarca, L.A., Luterbach, K.J., Lys, D.B., & Covington, V. (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.

Henry, G.T., Patterson, K.M., Campbell, S.L., & Yi, P. (2013). UNC teacher quality research: 2013 teacher preparation program effectiveness report. Education Policy Inititative at Carolina. Available from: http://publicpolicy.unc.edu/files/2013/11/UNC_TQR_OverallProgramReport_Final.pdf

Henry, G.T., Purtell, K.M., Bastian, K.C., Fortner, C.K., Thompson, C.L., Campbell, S.L., & Patterson, K.M. (2014). The effects of teacher entry portals on student achievement. *Journal of Teacher Education, 65*(1), 7-23.

Hill, H.C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal, 48*(3), 794-831.

Kane, T., Rockoff, J., & Staiger, D. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review, 27*(6), 615-631.

Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2012). Teacher preparation programs and teacher quality: Are there real differences across programs? Available from: http://economics.missouri.edu/working-papers/2012/WP1204_koedel_et_al.pdf

National Council on Teacher Quality. (2014). 2014 teacher prep review: A review of the nation's teacher preparation programs. Available from: www.nctq.org/dmsView/Teacher_Prep_Review_2014_Report.

National Research Council. (2002). *Scientific research in education.* Washington, DC: National Academies Press.

Nye, B., Konstantopoulos, S., & Hedges, L. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237-257.

Peck, C.A., Gallucci, C., Sloan, T., & Lippincott, A. (2009). Organizational learning the program renewal in teacher education: A socio-cultural theory of learning, innovation, and change. *Educational Research Review, 4*(1), 16-25.

Peck, C.A., Gallucci, C., & Sloan, T. (2010). Negotiating implementation of high-stakes performance assessment policies in teacher education: From compliance to inquiry. *Journal of Teacher Education, 61*(5), 451-463.

Peck, C.A. & McDonald, M.A. (2014). What is a culture of evidence? How do you get one? And…should you want one? *Teachers College Record 116*, 1-27.

Peck, C.A., Singer-Gabella, M., Sloan, T., & Lin, S. (2014). Driving blind: Why we need standardized performance assessment in teacher education. *Journal of Curriculum and Instruction, 8*(1), 8-30.

President's Council of Advisors on Science and Technology (PCAST), *Prepare and Inspire: K-12 Education in Science, Technology, Engineering, and Math (STEM) for America's Future,* September 2010, http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-stemed-report.pdf

Rich, M. (2014, April 25). Obama administration plans new rules to grade teacher training programs. *The New York Times.* Retrieved from http://www.nytimes.com/2014/04/26/us/obama-administration-plan-seeks-to-rate-teacher-training-programs.html?_r=0.

Tennessee State Board of Education. (2012). 2012 report card on the effectiveness of teacher training programs. Available from: http://www.tn.gov/thec/Divisions/fttt/12report_card/PDF%202012%20Reports/2012%20Report%20Card%20on%20the%20Effectiveness%20of%20Teacher%20Training%20Programs.pdf

Tennessee State Board of Education. (2013). 2013 report card on the effectiveness of teacher training programs. Available from: http://www.tn.gov/thec/Divisions/fttt/13report_card/1_Report%20Card%20on%20the%20Effectiveness%20of%20Teacher%20Training%20Programs.pdf

Appendix

Appendix Table 1: Individual-Level Data Shared with the UNC System TPP

Category	Variable	Description
Employment	Lea_Sch	A variable indicating the school district and school an individual worked in as a teacher
	PP_Count	A variable indicating the number of pay periods an individual was paid as a teacher in a specific school during the fiscal year
	Min_PP; Max_PP	A pair of variables indicating the first and last pay period an individual was paid as a teacher in a specific school during the fiscal year
	FTE_PP_Sch; FTE_PP_All	A pair of variables indicating the number of full-time equivalency units an individual was paid as a teacher in a specific school and across all schools
Teacher Characteristics	Teaching_exp	A variable indicating a teacher's classroom teaching experience
	Graduate_deg	A variable indicating whether a teacher had earned a graduate degree at any point prior to the start of the academic year
	NBC	A variable indicating whether a teacher had earned NBC status at any point prior to the start of the academic year
	License_area; license_basis	Variables indicating the area(s) in which an individual holds a teaching license and the basis for each licensure area
	Test_code; test_score	Variables indicating the test type/code (e.g Praxis 0040) and the numerical score for a teacher
	Std_test_score	Variables indicating teachers' standardized test score values
	Tested	Variables indicating whether a teacher taught a tested-grade/subject during the academic year
Classroom Characteristics	Classes	A variable indicating the number of classes taught by a teacher in an academic year
	Grade	A variable indicating the average grade-level of students taught
	Num_students	A variable indicating the average class size for a teacher
	Class_ethnicity	A set of variables indicating the proportion of a teacher's students who are white, black, Hispanic, American Indian, or other
	Class_subject	A set of variables indicating the proportion of students a teacher taught in self-contained, math, reading, science, social studies, arts, vocational, PE, and all other classes
	Class_disabled; Class_advanced	A pair of variables indicating the proportion of a teacher's students who are disabled or academically advanced
	Class_FRL; Class_Reduced	A pair of variables indicating the proportion of a teacher's students who qualify for free or reduced-price lunches
	Class_islep; Class_waslep	A pair of variables indicting the proportion of a teacher's students who are currently receiving or previously received Limited English Proficient services
	Class_daysabs	A variable indicating the average number of days absent for a teacher's students
	Class_prior_ach (Standardized)	A set of variables indicating the average prior score for a teacher's students on available End-of-Grade and End-of-Course exams
	Class_prior_ach (Level)	A set of variables indicating the average prior achievement level (I, II, III, IV) for a teacher's students on available End-of-Grade and End-of-Course exams

Appendix Table 1: Individual-Level Data Shared with the UNC System TPP Cont.

Category	Variable	Description
School Characteristics	Urbanicity	A variable indicating a school's urbanicity status—city, suburb, town, or rural
	ADM	A variable indicating a school's average daily membership (school size)
	Pctfrpl	A variable indicating a school's percentage of students eligible for free or reduced price lunches
	Stsrate; Actper1k	A pair of variables indicating a school's short-term suspension rate and violent acts rate
	School_ethnicity	A set of variables indicting the proportion of a school's students who are white, black, Hispanic, American Indian, multiracial, or Asian
	Tot_PPX	A variable indicating the total per-pupil expenditures at the school
	Spending Categories	A set of variables indicating per-pupil expenditures directed to: regular, special and supplemental instruction; instructional support; professional development; student services; extracurricular activities; transportation; food service; school maintenance; and school leadership
	AYP Percentage	A variable indicating the percentage of AYP goals met at a school
	ABC_Status; ABC_Growth	A pair of variables indicating a school's North Carolina accountability and growth status
	Performance	A variable indicating a school's performance composite value—the percentage of End-of-Grade and/or End-of-Course exams passed
	Teacher Credentials	A set of variables indicating the proportion of a school's teachers who are fully licensed, novice (less than 3 years experience), NBC, or holding a graduate degree
	Prop_teachers	A variable indicating the average number of students, per teacher, at the school
	Stay_ratio	A variable indicating the proportion of a school's teachers that returned from the previous year
Teacher Outcomes	Will_return (Overall)	A variable indicating whether a teacher will return to North Carolina public schools in the following school year
	Will_return (School)	A variable indicating whether a teacher will return to the same North Carolina public school in the following school year
	ES_math_VA; ES_math_quintile	A pair of variables indicating a teacher's value-added in elementary grades mathematics and the quintile of the value-added estimate
	ES_read_VA; ES_read_quintile	A pair of variables indicating a teacher's value-added in elementary grades reading and the quintile of the value-added estimate
	ES_sci_VA; ES_sci_quintile	A pair of variables indicating a teacher's value-added in 5th grade science and the quintile of the value-added estimate
	MS_math_VA; MS_math_quintile	A pair of variables indicating a teacher's value-added in middle grades mathematics and the quintile of the value-added estimate

Appendix Table 1: Individual-Level Data Shared with the UNC System TPP Cont.

Category	Variable	Description
Teacher Outcomes	MS_read_VA; MS_read_quintile	A pair of variables indicating a teacher's value-added in middle grades reading and the quintile of the value-added estimate
	MS_sci_VA; MS_sci_quintile	A pair of variables indicating a teacher's value-added in 8th grade science and the quintile of the value-added estimate
	HS_math_VA; HS_math_quintile	A pair of variables indicating a teacher's value-added in high school mathematics courses (algebra 1, algebra 2, and geometry) and the quintile of the value-added estimate
	HS_eng_VA; HS_eng_quintile	A pair of variables indicating a teacher's value-added in high school English courses (English I and II) and the quintile of the value-added estimate
	HS_sci_VA; HS_sci_quintile	A pair of variables indicating a teacher's value-added in high school science courses (biology, chemistry, physical science, and physics) and the quintile of the value-added estimate
	HS_ss_VA; HS_ss_quintile	A pair of variables indicating a teacher's value-added in high school social studies courses (U.S. history and civics/economics) and the quintile of the value-added estimate