

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
© 2013 American Association of
Colleges for Teacher Education
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0022487113496431
jte.sagepub.com


Gary T. Henry¹, Shanyce L. Campbell², Charles L. Thompson²,
Linda A. Patriarca³, Kenneth J. Luterbach³, Diana B. Lys³, and
Vivian Martin Covington³

Abstract

Calls for evidence-based reform of teacher preparation programs (TPPs) suggest the question: Do the current indicators of progress and performance used by TPPs predict effectiveness of their graduates when they become teachers? In this study, the indicators of progress and performance used by one program are examined for their ability to predict value-added scores of program graduates. The study finds that rating instruments, including disposition surveys, clinical practice observation ratings, and portfolio assessments, each measure a single underlying dimension rather than the multiple constructs they were designed to measure. Neither these instruments nor teacher candidates' scores on standardized exams predict their later effectiveness in the classroom based on value-added models of student achievement. Candidates' grade point averages during their preparation program and number of math courses were positively associated with their students' math score gains. These findings suggest a need for better instruments to measure prospective teachers' progress toward proficiency.

Keywords

teacher research, school/teacher effectiveness, quantitative research, HLM (hierarchical linear modeling), education reform

A recent National Council for Accreditation of Teacher Education (NCATE; 2010) report made a bold pronouncement: “It is time to fundamentally redesign preparation programs to support the close coupling of practice, content, theory, and pedagogy” (p. iii). Yet, as another recent U.S. National Research Council (2010) study made clear, we do not have sufficient evidence about teacher preparation programs (TPPs) to know with any certainty what contributes to their effectiveness or “what makes clinical preparation effective” (NCATE, 2010, p. iv). As efforts to reform TPPs proceed, there is a clear need for evidence to inform the process. Over the longer term, the goal should be to institutionalize the use of evidence to guide continuous improvement of teacher preparation and thereby, of teaching and learning in our schools. In this article, we present the findings from a study designed to take some first steps toward that goal.

A logical first step in the pursuit of evidence-based improvement of TPPs is to examine the indicators of progress and performance routinely collected as teacher candidates' progress through a TPP to identify which of these indicators might account for variation in candidates' later effectiveness in the classroom as measured by value-added models (VAMs) of student achievement. Therefore, we assessed the predictive validity of a broad range of such

measures in one large TPP. Data on course taking and grades, professional behaviors and dispositions, performance assessments during student teaching, Praxis I exam scores, and comprehensive portfolios of the candidates' work at the end of their programs are routinely collected by this and many other TPPs and are examined in this study as potential predictors of effectiveness.

We define predictive validity as the extent to which the data gathered on teacher candidates in the course of their preparation are correlated with their effectiveness after they graduate and begin teaching. For the measure of effectiveness, we used multilevel VAMs that measure teachers' effectiveness, the changes in their students' test scores attributable to the teacher after adjusting for students' prior test scores, and a rich set of student, classroom, and school characteristics.

¹Vanderbilt University, Nashville, TN, USA

²University of North Carolina at Chapel Hill, USA

³East Carolina University, Greenville, USA

Corresponding Author:

Gary T. Henry, Department of Leadership, Policy and Organization,
Peabody College, Vanderbilt University, 106D Payne Hall, PMB #414,
230 Appleton Place, Nashville, TN 37203-5721, USA.
Email: gary.henry@vanderbilt.edu

While test scores are only one measure of teacher effectiveness, they are important indicators of how much students know and are able to do and are increasingly the focus of federal, state, and district policy interventions. We recognize that some indicators of teacher candidates' progress may relate to important aspects of teaching that do not affect student achievement scores directly and that other indicators may be required to establish a minimum level of quality rather than to predict their performance as teachers. For example, measures of professional conduct may be needed to ensure appropriate behaviors yet not predict student learning. However, it is reasonable to expect that some indicators of teacher candidates' progress collected by TPPs should be related to value-added measures of their later effectiveness as teachers. To the extent that current indicators of progress and performance do not predict later effectiveness, new and better indicators of candidates' strengths on entry and performance during the program will be needed to guide reform and continuous improvement of TPPs.

The overarching hypothesis that guided the development and implementation of this research project was straightforward: Teacher candidates who perform better on these indicators of progress during preparation should become more effective teachers when they enter the classroom. If so, the evidence from these assessments can be used to (a) provide feedback about the strengths and weaknesses of teacher candidates that relates directly to their ability to improve student achievement; (b) identify specific teacher candidates who need supplemental instruction, coaching, or mentoring; (c) redirect low performing teacher candidates into other fields; and (d) provide systematic information about knowledge, skills, and dispositions of effective teachers that are or are not being developed through the preparation program. If the assessments fail to predict effectiveness in the classroom, the assessments (a) may need to be redeveloped and tested for their ability to predict teacher effectiveness; (b) should be justified as measuring other aspects of effective teachers in ways that do not directly relate to improving the test scores of the students and should be subjected to similar validation efforts that rely on other measures of teacher effectiveness; or (c) should be replaced by alternative assessments with better predictive validity. Institutionalizing the use of genuinely predictive measures of teacher candidate progress from acceptance into candidacy through recommendation for licensure and certification should enable individual TPPs to improve continuously.

Study Background

Recently, states and school districts have enabled researchers to compile administrative longitudinal data sets that link individual teachers with their students, including information on the students' test scores and on teachers' education, experience, and certification status. In several states and districts, practicing classroom teachers have been linked to their

preparation programs, traditional university-based programs and alternative entry programs, such as Teach for America and Visiting International Faculty. Several states have reported the effectiveness of TPPs in terms of the average gain graduates added to their students' test scores (Gansle, Noell, & Burns, 2012; Henry et al., 2011). While providing information on the relative effectiveness of different TPPs represents a step forward, it does not provide information that TPP personnel can use to guide program improvement for their candidates. In other words, these analyses identify TPPs that need to be improved but do not provide specific information concerning what should be improved or how.

In addition to the studies estimating TPP effectiveness, studies estimating the effects of specific components of TPPs and individual teacher candidates' experiences have begun to emerge. In the first study of this type, Boyd and colleagues examined program requirements and survey information about the experiences of individual teachers during their preparation to determine whether there was a relationship between program requirements and subsequent classroom effectiveness as measured by VAMs (Boyd et al., 2006; Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009). The study sheds light on differences between programs on dimensions such as the extent and quality of candidates' experiences in schools or specific content coursework that are associated with varying levels of VAM effectiveness in the classroom (Boyd et al., 2009). Rather than a comparative study across programs, we develop a microlevel focus on the variation that occurs within a single program in the study presented in this article.

The purpose of this study is to begin to test an approach that could be used to provide an evidence base for continuous improvement of an individual TPP. As Diez (2010) has pointed out, a full assessment of the relationship between a TPP and later effectiveness in the classroom will require attention to (a) whether and to what extent teacher candidates actually learn what a TPP seeks to teach them, (b) the extent to which they subsequently put what they learn into practice in the classroom, and (c) the degree to which these classroom practices contribute to student learning. The present study takes a step toward such an assessment by examining the link between teacher candidates' performance in the TPP and their later effectiveness in promoting student learning, but does not address the extent to which teacher candidates subsequently apply in the classroom what they have learned in their TPP.

In one sense, this is a "proof of concept" study of one large TPP that produces effective elementary teachers. It capitalizes on a longitudinal database similar to ones that many states have recently developed to assess the effectiveness of individual teachers and TPPs (Henry, Kershaw, Smith, & Zulli, 2012). In earlier studies of the effectiveness of the state's TPPs, elementary teachers prepared by the large university-based TPP that is the focus of the present study were highly rated in terms of their effectiveness. For

example, in reading students taught by the program's graduates scored significantly higher on their end-of-grade (EOG) reading exams than students prepared by teachers from other programs, including other public universities in the state (Henry et al., 2011). These earlier studies identified the TPP as effective based on the effects of its graduates on student test scores, but could not provide information on what accounted for this effectiveness, such as whether routinely collected indicators of teacher candidates' progress predicted postgraduation effectiveness in the classroom.

To better understand the relationship between the indicators of teacher candidates' progress and the effectiveness of TPP graduates, program leaders and faculty collaborated with other researchers for this study. This large regional university was founded as a teacher training school and currently operates a large TPP accredited by the NCATE. The program was an excellent case for the intensive study for several reasons. First, the TPP was one of the largest producers of teachers in the state and region, graduating on average 753 teachers per year, 319 of whom were certified to teach elementary grade students between 2008-2009 and 2011-2012. Second, the TPP routinely collected the five types of progress indicators that were previously mentioned, all of which are aligned with the requirements of national accrediting and state agencies. Third, the program leaders and faculty made a commitment to hold itself accountable for the quality of the TPP and to demonstrate the program's effectiveness in impacting P-12 student learning through receipt of the American Association of State Colleges and Universities' Christa McAuliffe Excellence in Teacher Education Award. Finally, in a report on the effectiveness of TPPs in the state (Henry et al., 2011), the institution's TPP outperformed all other sources of teachers in elementary school reading and was rated highly in two consecutive analyses for the performance of its elementary teachers in raising their students' test scores.

In the following section, we provide a brief historical description of indicators of teacher candidates' progress and summarize the literature on indicators of teacher candidates' progress and performance that are currently used in TPPs. This section is followed by a discussion of the data and methods used in the study. We then describe our analysis and study results. We conclude with a discussion of the results and future directions.

A Brief History

Definitions and Indicators of Progress and Performance in Teacher Preparation

Preparing effective teachers has long been one of the goals for TPPs (Ogren, 2005), but definitions of effectiveness have varied substantially since the 19th century when many normal schools were established across the country. While some focus on learning and imparting content knowledge to

students has been a reasonably consistent part of the definition of what effective teachers should do, the definition has also included at various times instilling values, such as Protestant morality, belief in democratic tenants and capitalism, promoting social mobility and social justice, as well as individual self-esteem, and human emancipation. During the Sputnik era and again in the current era of standards-based education and accountability, there has been increased emphasis on fostering student achievement as a defining characteristic of an effective teacher.

In addition to content knowledge and general teaching or pedagogical skills, content-specific pedagogical knowledge (Shulman, 1986, 1987) and more recently the effective use of technology to teach specific content—technological pedagogical content knowledge (Mishra & Koehler, 2006)—have been incorporated into the skills that teacher candidates should develop to become effective teachers. In theory, the developing mastery of these skills among teacher candidates could be assessed during their time in a TPP and subsequently could be expected to correlate with teachers' effectiveness, including value-added measures of effectiveness in terms of raising their students' achievement test scores. In traditional TPPs, acquiring these skills is frequently tied to specific courses that teacher candidates are required to pass for certification. In addition to passing these courses, increasingly TPPs require candidates to prepare portfolios that use various types of evidence for a summative evaluation of the teacher candidates' readiness to teach, and to take standardized exams on content knowledge and pedagogy, such as Praxis exams.

Another type of assessment that TPPs utilize are ratings of professional dispositions, which are collections of behaviors and attitudes, including professionalism, relationships with others, and teaching qualities, have been seen as attributes of effective teachers that could promote learning, motivation, and the development of students (Flowers, 2006). Finally, clinical experiences, which offer teacher candidates opportunities to practice teaching, have been regarded as an important training method (McDiarmid & Clevenger-Bright, 2008) and evidence shows that the extent and quality of these experiences do correlate with value-added estimates of teachers' effectiveness (Boyd et al., 2009). At present, the NCATE and the Teacher Education Accreditation Council (TEAC) collectively accredit approximately 900 TPPs in the United States.¹ To gain NCATE or TEAC accreditation or reaccreditation, TPPs must provide various types of evidence that go beyond descriptive data to include information that demonstrates that teachers have attained the knowledge, skills, and dispositions needed to be effective teachers. More specifically, NCATE (2008) required that each TPP assess candidates' content knowledge, pedagogical content knowledge, and pedagogical skills as well as professional knowledge, skills, and dispositions "necessary to help all students learn" (p. 1). TEAC required trustworthy, reliable, and valid evidence indicating

that each TPP meets its quality standards, which include that candidates demonstrate (a) subject matter knowledge, (b) pedagogical knowledge, (c) caring and effective teaching skills, (d) ability to learn on their own, (e) understanding of multicultural perspectives, and (f) how to use technology.

As state accreditation is a prerequisite for national accreditation, it is important to consider state requirements to recognize all of the requirements on TPPs. In many states, to be eligible for an initial teaching license, teacher candidates are required to complete particular pedagogical and content knowledge courses; prepare a summative portfolio; demonstrate particular dispositions, behaviors, and skills; pass required Praxis exams; and receive satisfactory ratings on the assessments of their final clinical experiences. Dimensions on which the teacher candidates are typically assessed during their clinical experiences include lesson planning, content knowledge, management of instructional time, management of student behavior, instructional presentation, instructional monitoring and feedback, and dispositions, particularly attitudes toward students and colleagues. While the course requirements are generally specified, the methods of assessment for the other domains are often left to the individual TPP's discretion.

Prior Research on Indicators of Progress in a TPP

In this section, we briefly review the empirical literature on indicators of teacher candidates' progress with particular emphasis on the relationship of the indicators to effective teaching and student learning. We review the five categories mentioned above in turn: (a) evidence on coursework and grades, (b) professional behaviors and dispositions, (c) performance assessments during student teaching, (d) Praxis I exam scores, and (e) comprehensive portfolios of the candidates' work at the end of their programs.

Coursework and grades. One direct influence of TPPs occurs in the courses that programs offer and/or require. Wilson, Floden, and Ferrini-Mundy (2001) presented a thorough review of the literature on coursework and concluded that subject-specific coursework and pedagogy are important; however, more research is clearly needed. In their review, the only rigorous study that examined the impact of coursework on high school student achievement found diminishing returns to additional math courses on student performance (Monk, 1994). More specifically, teacher candidates who took more than five math courses had smaller gains on student achievement than those who took five or fewer courses. Monk (1994) also found that math pedagogy courses were associated with greater gains than math content courses. In a more recent study using a series of fixed effect models, Kukla-Acevedo (2009) examined the relationship between performance in a preparation program and later fifth-grade student learning and found that grade point average (GPA), math, and math education course hours were correlated with

fifth-grade students' math achievement. At the program level, Boyd et al. (2009) found that graduates of programs that required math and English language arts (ELA) courses positively influenced student achievement for 2nd-year teachers, but the results did not hold for 1st-year teachers. At the individual teacher level, the authors found that 2nd-year teachers who took more math content and pedagogy courses were more effective; however, a parallel finding did not hold for ELA. Examining actual coursework, Harris and Sass (2011) found little evidence that the amount of coursework taken in the subject that was taught, measured by credits, is associated with teacher effectiveness in math and reading.

Comprehensive portfolio assessments. Many institutions require teacher candidates to complete a portfolio as a condition of certification. The components of the assessment vary from institution to institution. Scholars conclude that, in a broad sense, portfolios allow teacher candidates to reflect on their teaching practices and demonstrate the skills and knowledge they have obtained while in the TPP (Delandshere & Arens, 2003; Wolf, 1991). While the use of portfolios is desirable for its ability to capture multiple items including authentic assessments, many question their predictive validity, suggesting that portfolio raters do not have the appropriate training and understanding of what teachers should know and do when rating teacher candidates (Delandshere & Petrosky, 1998). Recently, several organizations such as the Teacher Performance Assessment Consortium (TPAC) and the Interstate New Teacher Assessment and Support Consortium (INTASC) have begun to develop and pilot portfolio assessment tools that teacher candidates may be required to pass before receiving preliminary licensing. The predictive validity of the instruments has not yet been demonstrated.

Clinical experiences, student teaching, and placement. Educators and scholars alike agree that clinical experiences, the most common form of which is student teaching, are an essential component of TPPs (Boyd et al., 2009; Wilson et al., 2001). Many teacher candidates receive their clinical experiences during their final semester in a TPP. This experience has short- and long-term consequences on expectations and labor market decisions on entry into the teaching profession (Greenberg, Pomerance, & Walsh, 2011). In an examination of teacher preparation and student achievement, Boyd et al. (2009) found that among 1st-year teachers, congruence between student teaching placement and current job placement had a positive relationship with student performance in math. These results were consistent for 2nd-year teachers as well. Studying field placement and later retention, Ronfeldt (2012) found that teacher candidates who are placed in "easier-to-staff" schools for their clinical experiences are more likely to remain in that school and show greater student achievement gains than teacher candidates who are placed in "difficult-to-staff" schools.

Standardized assessments (Praxis/Scholastic Assessment Test [SAT]/American College Testing [ACT]). Prior to admission into a TPP, students are often screened using various standardized assessments including Praxis I Pre-Professional Skills Tests (PPSTs), SATs, and ACTs. The Praxis PPSTs are multiple-choice tests (except for the writing test, which includes an essay component) intended to measure basic reading, mathematics, and writing skills (Educational Testing Service [ETS], n.d.). The literature on standardized assessments of teacher candidates, specifically PPST, is divided into two lines of inquiry. First, several studies have examined the validity of PPSTs by estimating correlations with SAT, ACT, and undergraduate GPAs. The positive correlations between PPST, SAT, and ACT scores are well established in the literature (Mikitovics & Crehan, 2002; Pool, Dittrich, Longwell, Pool, & Hausfather, 2004; Wakefield, 2003). Pool and colleagues (2004) examined TPPs in three types of institutions—highly selective, selective, and regional state university—and found a strong positive relationship between SAT and PPST scores among the three institutions, 0.75, 0.71, and 0.70, respectively. Examining a single urban university, Mikitovics and Crehan (2002) found a strong relationship ($r = .65$) between ACT and PPST composite scores. In addition, several studies have examined the relationship between PPST scores and student teacher ratings and found no statistical relationship between the assessments, suggesting that PPST scores may not strongly predict whether a student is successful in a TPP (Dybdahl, Shaw, & Edwards, 1997; Hicken, 1992; Mikitovics & Crehan, 2002).

The second line of inquiry examines PPST as a gatekeeper among students of color and economically disadvantaged students. The research on racial, ethnic, and economic gaps in test performance has conclusively shown that with the exception of Asian students, students of color and economically disadvantaged students score worse than their White and economically advantaged counterparts (Lee, 2002). In a longitudinal study of Black and Hispanic teacher candidates' experiences with the PPST, Bennett, McWhorter, and Kuykendall (2006) found that race, ethnicity, economic status, and gender affect PPST passing rates and test taking persistence among students of color. In addition to Black and Hispanic students, Native American and economically disadvantaged students also experience negative consequences associated with the PPST as a TPP entry requirement (Wakefield, 2003; Watanabe, 2008).

Dispositions. Assessments of dispositions have become major components by which institutions measure the “readiness” of teacher candidates to enter the teaching profession. Institutions are evaluated on how well teacher candidates can “identify, define and operationalize” professional dispositions (NCATE, 2008, p. 90). However, because dispositions are based on internal characteristics that are difficult to define and assess, institutions and governing bodies have defined dispositions in various ways, which complicate this

measure for teacher candidates (Ginsberg & Whaley, 2003; Young & Wilkins, 2008). For example, INTASC describes “critical dispositions” as the “habits of professional action and moral commitments that underlie the performances [and] play a key role in how teachers do, in fact, act in practice” (Council of Chief State School Officers, 2011, p. 6). NCATE defines professional dispositions as “professional attitudes, values, and beliefs demonstrated through both verbal and non-verbal behaviors as educators interact with students, families, colleagues, and communities” (p. 89). NCATE explicitly indicates that institutions should evaluate dispositions based on observable behaviors with an emphasis on fairness and the belief that all students can learn.

The research on dispositions primarily addresses the reliability and validity of disposition instruments. In an examination of the Clinical Experience Rubric, an instrument developed to assess teacher candidates' dispositions, Flowers (2006) found three highly interrelated factors that teacher candidates were evaluated on—professionalism, teaching quality, and relationship with others. However, the factors were highly intercorrelated (between .82 and .89) and could be assessed with reasonable fit statistics as a one-factor construct. Another study of 32 field experience instruments found 13 disposition categories and 3 major dispositional themes—professional behaviors, professional way of thinking, and personality characteristics (Young & Wilkins, 2008). In our review of the literature, we did not find empirical research on the relationship between dispositional ratings of teacher candidates and the test scores of students they later taught.

As Diez (2010) has described in her careful explication of the chain of potential causation connecting teacher preparation with later effectiveness in the classroom, the chain is long, complex, nuanced, and rife with contingencies, and the study of it is replete with measurement and modeling challenges. In addition, as the foregoing literature review makes clear, efforts to date to trace these connections have addressed only segments of the chain, and have often done so by focusing on singular and relatively gross measures of the process, such as prospective teachers' course taking and GPAs.

Furthermore, student achievement is multiply influenced and heavily dependent on relatively stable factors in individual students' backgrounds. Some have even argued that the combination of the factors affecting teachers' ability to put what they learned during their preparation programs into practice (Diez, 2010) and the measurement and modeling challenges render efforts to assess the connection between teacher preparation and student achievement a futile enterprise. Yet in prior work, researchers have shown that with VAMs, it is possible to isolate the effects of teacher preparation routes and programs from the multiplicity of other variables affecting student learning (Boyd et al., 2009; Gansle et al., 2012; Henry et al., in press).

These studies demonstrate that it is possible to find differences in TPP effectiveness but as the framework provided by

Diez (2010) makes plain, they do not explain why these differences arise. A TPP's graduates' average VAM scores can tell us whether a program is performing well or underperforming, but they cannot tell us why. To improve TPPs, we need insights into the mechanisms leading to higher or lower performance. TPPs need measures of teacher candidates' progress toward effectiveness in the classroom. The approach that we take in this study is to examine existing measures of teacher candidate progress to see whether they are correlated with teachers' contribution to student test score gains.

We recognize that teacher preparation does not directly shape student achievement but through a long chain of indirect effects (Diez, 2010). But our strategy has been first to use value-added modeling to estimate the extent to which TPPs affect student achievement through their graduates and then to search for the predictors of the teachers' effectiveness. This approach has the triple advantage of (a) offering immediate evidence that teacher preparation does have substantively meaningful effects (an important finding in itself in the present environment of policy debate), (b) identifying more and less effective programs so that improvement efforts can be focused where they are needed most and guided by reference to more effective programs, and (c) providing a means to test whether teacher candidates' performance during their preparation process predict their effectiveness in the classroom. We do so by assessing the predictive validity of a series of measures and indicators of progress collected by a large TPP that has been found to prepare teachers who contribute to greater gains in their students' test scores than teachers prepared elsewhere.

Data and Method

Sample, Data, Measures, and Analysis

Study sample. For this study, we began by identifying all graduates of the large university TPP mentioned above who were in their first 5 years of teaching in the public schools of the state in which the university is located. The sample was further reduced to include only graduates who were teaching tested subjects (i.e., math and reading) at the elementary level. The majority of these teachers were licensed in elementary education (93%), but a small number were licensed in special education (6%), middle school science (0.40%), middle school social studies (0.40%), and high school English (0.40%). Of the 1,791 graduates with TPP data, 279 met the criteria for inclusion in this study.

Data. We assembled a unique data set that includes teachers' preservice measures of progress and performance collected by the TPP (i.e., the five types of TPP progress measures discussed above) and student, teacher, classroom, and school characteristics from a statewide administrative database for elementary school teachers in tested grades and subjects. Teacher data are linked to their students using official classroom rosters that document the teacher assigned to each

student by class period. The data for this analysis included TPP graduates between 2005-2006 and 2008-2009 who went on to teach in Grades 3 to 5. We estimate separate effects using student outcomes on statewide End of Grade (EOG) math and reading tests, with prior EOG test scores as controls for prior test performance in our VAMs.

The data contained some missing values across all measures for various reasons. For example, if a teacher candidate or rater does not complete all disposition forms, this will lead to missing data. If assessments were missed or skipped, missing data may result. In addition, any of the measures not entered into the data system result in missing data.

Missing data are a concern because missingness may threaten the validity of the findings due to inefficiency, loss of power, and parameter bias (Rose & Fraser, 2008). We first examine whether the missing values are missing completely at random (MCAR) by using Little's (1988) MCAR test. If the data are MCAR, complete case deletion is possible without producing biased estimates. However, the results suggest that the data are not MCAR; therefore, multiple imputation was used as a method to generate computed values for the missing values. In multiple imputation, missing values are replaced by values predicted by the relationships between the other variables in the data set and the target variable that is missing to which a small amount of random variation is added. The resulting imputed data set has no missing data and maintains the variances and covariances of the original variables. Consistent with prior literature, 10 data sets with imputed values were created, analyzed separately, and the 10 coefficients on each independent variable averaged to produce the final estimates (Rubin, 1987; Schafer, 1997; Schafer & Graham, 2002; for complete details on the imputation procedures used, see Rose & Fraser, 2008).

Measures. In this section, we describe the variables of interest, which include five indicators of teacher candidates' progress through the program, the outcome variables, and the covariates used in the analysis. The five indicators of progress are (a) data on course taking and grades, (b) ratings of professional behaviors and dispositions, (c) ratings of performance during student teaching, (d) Praxis I exam scores, and (e) ratings of comprehensive portfolios of the candidates' work at the end of their program.

Following Tabachnick and Fidell's (1989) recommendation, exploratory factor analysis (EFA) was used in the first stages of the analysis to determine whether items designed to measure three of the variables of interest—professional behaviors and dispositions, performance during student teaching, and comprehensive portfolios of the candidates' work—do indeed measure the intended constructs and whether each item made a reliable contribution to the measures. EFA is a statistical technique that uses the correlation among items on a survey instrument to determine the number of unique constructs that underlie the items and the extent to which each item is associated with each of the latent constructs. For this analysis, we used principal components

Table 1. Description of Coursework Variables.

Coursework variables	Math model	Reading model
Subject-specific content courses	Math content	Reading content
Subject-specific pedagogy courses	Math pedagogy	Reading pedagogy
Other content courses	Reading Humanities and fine arts Science Health and physical education Social studies	Math Humanities and fine arts Science Health and physical education Social studies
Other pedagogy courses	Reading pedagogy Science pedagogy English pedagogy Social studies pedagogy	Math pedagogy Science pedagogy English pedagogy Social studies pedagogy
General pedagogy courses	Instructional technology Learning and development	Instructional technology Learning and development
Professional studies core courses	Technology Education Introduction to diversity Foundations of American Education	Technology Education Introduction to diversity Foundations of American Education

analysis with orthogonal rotation. Conducting the EFA for the three types of assessments at each administration point established the basis for examining whether the attributes measured relate to student test performance.

Coursework and grades. Data on coursework include course counts within subject areas. The course counts were aggregated into seven categories of courses separately for the reading and math models. In the reading model, the number of courses taken in reading content, reading pedagogy, nonreading content, nonreading pedagogy, general pedagogy, professional studies, and additional courses is included. The math model includes the number of courses taken in math content, math pedagogy, nonmath content, nonmath pedagogy, general pedagogy, professional studies, and additional courses. As shown in Table 1, nonreading content includes health and physical education, humanities and fine arts, and science, math, and social studies. Nonreading pedagogy includes math pedagogy, science pedagogy, English pedagogy, and social studies pedagogy. Similar groupings of courses are used to create the nonmath content and nonmath pedagogy categories. General pedagogy courses include instructional technology and learning and development courses. Professional studies core courses include technology in education, introduction to diversity, and foundations of American education. Analyses were also conducted without the aggregation; however, there was no significant change in the results.

Total GPA is composed of the candidate's GPA over the set of courses taken while in the TPP, generally the final 2 years of the candidate's undergraduate studies. GPA was calculated by the division of total points acquired by total course hours earned. Courses taken before candidate is accepted into upper division are not included in the calculation of the GPA.

Professional behaviors and dispositions. At three distinct points, the TPP examined in this study uses a survey to assess each teacher candidate's dispositions (e.g., attitudes toward teaching, students, and families; sensitivity; responsibility; initiative; respect; and willingness to share). At the first point of data collection, candidates complete the survey as a self-assessment. This occurs early in the candidate's preparation. The second point occurs during candidates' student teaching. The clinical or supervising teacher uses the disposition survey to assess the teacher candidate. The final point occurs near the end of candidates' student teaching, when the university supervisor completes the final evaluation of candidates. The disposition survey uses a 4-point rating scale of "always/almost always," "usually," "sometimes," or "rarely/never" to rate candidates on 20 items that have been conceptually grouped into three constructs: (a) professional demeanor, (b) professional commitment, and (c) professional interactions. Based on an EFA (described below), this variable was measured by an index of the survey responses from the final assessment conducted by the university supervisor.

Performance assessment during student teaching. The TPP has a two-semester student teaching internship requirement where candidates complete the traditional "student teaching" during the second semester. The assessment of student teaching used in this study occurs in the final semesters of the TPP, during which teacher candidates are assigned to public schools for 15 weeks. Teacher candidates are required to teach full-time for 3 consecutive weeks within the 15-week period. At four points during this 15-week student teaching experience, all teacher candidates are evaluated using the same instrument, which is labeled a progress report. Teacher candidates were rated as "above satisfactory," "satisfactory," "needs improvement," or "unsatisfactory" on four constructs: planning, knowledge of subject, professional attitudes and

Table 2. Covariates Used in the Analysis.

Student covariates	Classroom and teacher covariates	School covariates
Prior test scores (math and reading)	Class size	School size (ADM)
Classmates' prior test scores	Heterogeneity of prior achievement within the class	Violent acts per 1,000 students
Days absent	Teacher out-of-field status	Suspension rates
Structural mobility	Single-year indicators for teacher experience	Total per-pupil expenditures
Within-year mobility		Average district teacher supplement
Race/ethnicity		Racial/ethnic composition
Gender		Concentration of poverty
Poverty status		
Gifted		
Disability		
Currently limited English proficient		
Was limited English proficient		
Overage for grade		
Underage for grade		

Note. ADM = Average Daily Membership.

relationships, and instruction. These constructs each contain four separate items: (a) management of instructional time, (b) management of student behavior, (c) presentation, and (d) instructional feedback and monitoring. Based on the EFA (described below), this variable was measured by an index of the ratings from the second evaluation period.

Praxis I exam scores. Praxis I in mathematics, reading, and writing scores were included in the analyses. The state where the TPP is located requires that candidates for admission into a TPP achieve a combined score of 522 or a passing score of 173 in the mathematics, 176 in the reading, and 173 in the writing components (ETS, n.d.). However, some students may be exempt from taking one or more of the tests because they have achieved specific scores on either the SAT or ACT. Consequently, not every teacher candidate will have a complete set of Praxis I scores, and due to these exemptions, some teacher candidates did not have any Praxis I scores. Prior to imputation, approximately 46% of the teacher candidates were missing one or more Praxis I scores in reading, writing, or mathematics.

Comprehensive portfolios. By the end of candidates' senior year, a portfolio that consists of four parts—candidate work sample of instructional practice, classroom management, impact on student learning, and technology skills—is required. The portfolio is evaluated by the university supervisor and receives either an “above expectations,” “meets expectations,” or “below expectations” rating on three constructs that have between 3 and 7 underlying items. Based on the EFA, this variable was measured by an index of 13 items (2 items were excluded from the latent measure, formative assessment, and reflection items from the candidate work sample ratings, due to low loadings that are discussed later).

Outcome measures. Two outcome measures are used to examine teachers' impact on student learning, math, and

reading achievement test scores. In the state where the TPP is located, public school students in Grades 3 to 5 are required to take EOG exams in math and reading each year. These exams are criterion-referenced multiple-choice tests that assess student knowledge and skills in accordance with the state's standards. The math and reading test scores are standardized within year, grade, and subject. Standardized scores have a mean of zero and a standard deviation of one.

Covariates. Because student test performance can be influenced by student, teacher/classroom, and school variables, we included these covariates in our analyses to isolate teacher effects from the effects of these other variables. Table 2 contains a complete list of these covariates. Five teacher-level covariates were included in the analyses. SAT/ACT scores and high school rank, expressed as a percentile, were included as measures of selection and academic ability of the TPP candidates. Several candidates in the sample were participants in the TPP partnership with local community colleges to prepare cohorts of teachers in an off-campus location. This covariate was included in the analyses because it represents a distinct group of teachers whose experience differed from that of traditional TPP candidates. Consistent with prior research on teacher candidate placement, two postbaccalaureate measures were created (Boyd et al., 2009; Cook & van Cleaf, 2000; Hebert & Worthy, 2001; Ronfeldt, 2012). The first is a dichotomous variable indicating whether student teaching occurred in the school where they currently teach. The second is a dichotomous variable indicating whether student teaching occurred in the same grade in which the teacher currently teaches.

Analytic methods. The purpose of the analysis of student test scores was to examine the extent to which ratings on the five types of progress and performance indicators were correlated with student test score gains for the students taught

by graduates of the TPP being studied. As the TPP graduates were not randomly assigned to schools, classes, or the students they teach, we implemented a hierarchical linear model (HLM) with students nested within classrooms nested within schools and rich covariates at each level designed to adjust for variables that may influence the test scores of the students taught by the TPP graduates and that correlate with the measures of teacher candidate progress or performance. This analysis uses a value-added modeling approach to estimate teachers' contributions to student test score gains. The model takes into account students' prior math and reading test scores, as well as student, classroom, teacher, and school characteristics in attempts to isolate the portion of student test gains attributable to the teacher. Equation 1 illustrates the estimation of the teacher effectiveness in reduced form:

$$Y_{icst} = \beta_0 + \beta_1 Y_{icst-1} + \delta_1 \mathbf{X}_{ics} + \theta_1 \mathbf{T}_{cs} + \varphi_1 \mathbf{S}_s + u_{00i} + r_{0cs} + e_{ics}, \quad (1)$$

where Y_{icst} is the standardized math and reading test score at time t for student i in classroom c in school s ; Y_{icst-1} is prior math or reading test score for student i in classroom c in school s ; \mathbf{X}_{ics} is a vector of individual characteristics for student i in classroom c in school s ; \mathbf{T}_{cs} is a vector of teacher and classroom characteristics in classroom c in school s ; \mathbf{S}_s is a vector of school characteristics in school s ; u_{00i} represents the error between schools; r_{0cs} represents the error between classrooms within schools for classroom c in school s ; and e_{ics} represents the error specific to student i in classroom c in school s .

Findings

In this section, we begin by providing descriptive information on our sample of teachers from the TPP being studied; then we provide the results of the EFA and the predictive validity analysis.

Descriptive Findings

The majority of the teachers in the study were in their 1st year of teaching. As is true of teachers across the state, the majority of the teachers were White females. The means and standard deviations for the variables of greatest interest in the study are presented in Table 3. Eighteen percent of the teachers taught in the same school where they conducted their student teaching and 23% taught the same grade level in which they taught during student teaching. The teachers in the sample taught in schools with varying racial and ethnic student populations; however, on average, teachers taught in schools with high concentrations of either Black or White student populations. On average, teachers taught in schools where 62% of the student population was eligible for the free or reduced lunch program. As students, the teacher took more reading and reading pedagogy courses than math and math pedagogy courses.

Table 3. Descriptive Statistics of Program Graduates Employed in the State's Public Elementary Schools.

Variables	M	SD
Subject-specific content courses (reading)	3.87	1.78
Subject-specific pedagogy courses (reading)	3.67	2.18
Other content courses (reading)	15.42	6.35
Other pedagogy courses (reading)	5.73	1.63
Subject-specific content courses (math)	1.93	1.19
Subject-specific pedagogy courses (math)	2.00	0.92
Other content courses (math)	14.20	3.26
Other pedagogy courses (math)	17.35	6.76
General pedagogy courses	11.64	2.61
Professional studies core courses	3.84	0.73
Additional courses	5.26	3.15
Total GPA	3.44	0.36
SAT/ACT	996.09	112.43
High school rank	62.19	29.91
Praxis I—reading	177.61	5.49
Praxis I—writing	174.79	4.26
Praxis I—math	178.90	6.39
Portfolio ratings	31.39	4.91
Disposition ratings index	77.22	5.16
Student teaching ratings index	24.22	3.01
Community college partnership	0.09	—
Same schools as student teaching experience	0.18	—
Same grade as student teaching experience	0.23	—

Note. GPA = grade point average; SAT = Scholastic Assessment Test; ACT = American College Testing.

EFA Findings

The EFA findings indicate that items on each of the three instruments used to measure teacher candidates' performance and progress measure a single underlying construct rather than the three or four constructs the instruments were intended to measure (see Tables 4-6). In other words, the responses on the items are so highly correlated that they are effectively measuring only one underlying construct and can be combined into a single factor or latent measure. For the two instruments that were administered more than once—the disposition survey and the student teacher rating—a separate factor analysis was conducted on each of the three administrations. In both cases, each administration produced a single underlying factor. For the analysis presented below, the construct from a single administration was chosen based on the amount of variance explained by the individual factor and the amount of missing data. Analysis of alternative administrations of the instruments produced very similar results.

Table 4. Portfolio Ratings Factor Loadings.

Variables	Loadings
Candidates work sample of instructional practices	
Unit plan	0.818
Lesson plans	0.810
Related student work	0.767
Formative assessment	0.060
Reflection	0.525
Classroom management	
Unit plan	0.713
Daily protocols/routines	0.859
Levels of consequences	0.883
Intervention strategies	0.843
Preventive management/motivation tech	0.838
Implementation plan	0.821
Reflection	0.785
Impact on student learning	
Assessment design	0.709
Data analysis	0.799
Reflection	0.770

Predictive Validity Findings

As previously noted, separate analyses for math and reading were conducted for predictive validity analysis (see Table 7). We examined the bivariate correlations between the variables of interest before conducting the multivariate analysis. Correlations were generally low but in two cases were higher than the absolute value of .40. The correlation between reading content courses and professional course was .41. The correlation between total GPA and nonmath content courses was $-.41$. The variance inflation factors did not indicate significant issues with multicollinearity. The analysis of the null model showed that for math, 21% of the variance was between schools, 6% was between classrooms within schools, and 71% was between students within classrooms. For reading, the figures were similar: 17% of the variance was between schools, 5% was between classrooms within schools, and 74% was between students within classrooms.

Overall, few of the indicators of progress and performance for the teacher candidates were correlated with the increases in their students' test scores after they became classroom teachers. One exception was that a one-course increase in math content courses taken as a teacher candidate is associated with an increase of about 3% of a standard deviation in students' math performance. In contrast, a one-course increase in math pedagogy courses taken as a teacher candidate is associated with a reduction of about 5% of a standard deviation in students' math achievement. In the reading model, taking more nonreading content pedagogy courses (e.g., math pedagogy, science pedagogy, and social studies pedagogy courses) is associated with 2% of a standard deviation reduction in reading achievement. This

finding seems to be attributable to the significantly negative effect of more math pedagogy courses on TPP graduates' effectiveness. On average, one additional professional studies core course was associated with an increase in reading test scores of about 11% of a standard deviation. Total GPA in upper division courses has a positive relationship with students' math achievement, but the coefficient for total GPA on reading achievement is not statistically significant. Teacher candidates' Praxis I scores, portfolio ratings, student teacher performance ratings, or professional disposition ratings were not associated with value-added measures of their students' learning in either math or reading.

Surprisingly, the teacher candidates' SAT scores were not related to the value-added measures of their students' learning after controlling for the other variables in the model. Moreover, controlling for the other covariates, the teacher candidates' high school rank was negatively and weakly correlated with value-added measures of their students' learning. It appears that the total GPA variable was a better predictor of math value-added scores than any of the other measures that appeared to be related to the teacher candidates ability measures, including SAT, high school rank, and Praxis I math scores.

A handful of other covariates produced noteworthy findings. Teaching in the same grade where the teacher candidate had student-taught significantly increased reading achievement by 4% of a standard deviation. The effect was larger, though not statistically significant at the $p < .05$ level, for math scores. In math and reading, teaching in the same school where a candidate did her student teaching had no statistical association with value-added to student achievement.

Discussion and Conclusion

In summary, the predictive validity of the five indicators of progress or performance used by the TPP examined in this study was low. Disposition surveys, student teaching ratings, and the summative portfolio assessment did not measure the multiple, underlying constructs they were intended to measure. Instead they provided what could be considered a global rating of the teacher candidates. Furthermore, none of these instruments produced measures that predicted the candidates' later effectiveness as teachers in the state's public schools as measured by value-added to their students' performance on math or reading achievement tests. Teacher candidates' grades in their final 2 years of coursework were positively associated with value-added scores of students' math achievement but not with reading achievement. The number of math courses taken by teacher candidates did predict their students' value-added test score gains, but the association with the number of math pedagogy courses taken was negative and significant. For reading, the number of professional studies core courses was positively related to teachers' value-added and the number of other content courses was

Table 5. Disposition Ratings Factor Loadings.

	Loadings
Professional demeanor	
The candidate demonstrates adherence to standards of professional ethics	0.978
The candidate wears professional attire for teachers when in a school setting	0.960
The candidate demonstrates reliability by performing assigned tasks or duties on time without prompting	0.966
The candidate displays a positive attitude toward teaching and interactions with students and families	0.984
The candidate responds to frustration and stress with poise and seeks positive outlets for emotions	0.962
The candidate demonstrates resourcefulness, initiative, and independence	0.954
The candidate accepts responsibility for successes and mistakes and seeks solutions to problems	0.971
Professional commitment	
The candidate establishes an environment of respect for diversity in professional relationships and through culturally responsive teaching	0.981
The candidate respects students as valued individuals by focusing professional decision making on student needs rather than personal preference	0.981
The candidate thinks critically; perceiving multiple sides of an issue or problem to develop creative solutions and make appropriate decisions	0.970
The candidate solicits suggestions and feedback and seeks opportunities for professional growth	0.966
The candidate responds positively to constructive criticism and suggestions by integrating professional feedback into practice	0.979
The candidate examines critically his or her perspective, experiences, and effectiveness and reflects on ways to improve student performance	0.978
The candidate persists in seeking new and more effective teaching strategies to help all children achieve success	0.967
Professional interactions	
The candidate listens and responds thoughtfully to the ideas and perspectives of others	0.970
The candidate demonstrates diplomacy, tact, and sensitivity toward the feelings and opinions of others	0.979
The candidate demonstrates awareness of impact of own words/actions on students by monitoring and adjusting personal behavior accordingly	0.980
The candidate shares ideas and materials willingly and contributes actively to positive group functioning	0.980
The candidate articulates ideas clearly and comfortably in conversation, discussion, or presentation, demonstrating conventions of standard spoken English and awareness of audience	0.974
The candidate articulates ideas clearly and coherently in writing demonstrating conventions of standard written English and awareness of audience	0.968

Note. One factor was retained for all rating periods.

Table 6. Student Teaching Performance Ratings Factor Loadings.

	Loadings
Planning	0.799
Knowledge of subject area(s)	0.835
Instruction	
Management of instructional time	0.846
Management of student behavior	0.821
Presentation	0.848
Instructional monitoring and feedback	0.791
Professional attitudes and relationships	0.769

Note. One factor was retained for all rating periods.

negatively related to teachers' value-added. Otherwise, course taking had no statistically significant effect on later effectiveness in the classroom.

Overall, these findings suggest an urgent need to identify or develop valid and reliable assessment instruments to measure the performance and progress of teacher candidates.

Valid and reliable instruments would allow TPP administrators to (a) provide feedback to teacher candidates about their strengths and weaknesses that directly relate to their ability as teachers to improve student achievement; (b) identify specific teacher candidates who need supplemental instruction, coaching, or mentoring; (c) redirect low performing teacher candidates into other fields; and (d) track the development of teacher candidates' knowledge, skills, and dispositions as they move through their preparation programs—in effect showing the value-added by the programs to the effectiveness of their graduates in raising student achievement. Without valid and reliable instruments, it will remain extremely difficult to improve TPPs in ways that increase student learning. In addition to providing better instruments, it is important that university supervisors and clinical teachers have the knowledge and skills to reliably assess teacher candidates using these tools. This may require substantial training and periodic retraining on the use of the instruments.

The puzzling coursework results highlight the need for additional research to make definitive conclusions regarding

Table 7. Average Effects of Preservice Measures of Progress and Performance on Program Graduates Effectiveness in Elementary Schools.

Variables	Math model	Reading model
	Unstandardized coefficient (SE)	Unstandardized coefficient (SE)
Subject-specific content courses	0.033** (0.009)	-0.004 (0.006)
Subject-specific pedagogy courses	-0.053** (-0.013)	-0.006 (0.005)
Other content courses	-0.001 (0.002)	-0.001** (0.002)
Other pedagogy courses	-0.002 (0.006)	-0.023 (0.009)
General pedagogy courses	0.001 (0.007)	-0.005 (0.005)
Professional studies core courses	0.037 (0.032)	0.112** (0.023)
Additional courses	-0.001 (0.004)	-0.002 (0.003)
Total GPA	0.126** (0.041)	0.008 (0.032)
SAT/ACT	0.000 (0.000)	0.000 (0.000)
High school rank	-0.001** (0.001)	-0.001** (0.001)
Praxis I—reading	-0.002 (0.002)	-0.001 (0.002)
Praxis I—writing	-0.002 (0.003)	-0.001 (0.002)
Praxis I—math	0.001 (0.002)	0.001 (0.002)
Portfolio ratings	0.001 (0.003)	0.003 (0.002)
Disposition rating index	0.002 (0.003)	0.001 (0.002)
Student teaching rating index	0.003 (0.004)	-0.002 (0.004)
Community college partnership	0.001 (0.078)	-0.015 (0.062)
Same schools as student teaching experience	-0.016 (0.035)	-0.009 (0.028)
Same grade as student teaching experience	0.050* (0.026)	0.041** (0.021)

Note. GPA = grade point average; SAT = Scholastic Assessment Test; ACT = American College Testing.

*Significant at the .10 level. **Significant at the .05 level.

the impact of course taking on student learning. Previous work has shown that beyond a certain point, additional courses taken by teacher candidates are sometimes associated with actual reductions in subsequent student achievement, but it remains unclear which courses offer diminishing returns and how they may do so. As there is a relatively fixed total on the number of courses teacher candidates must take to graduate, taking one course may reduce or eliminate the opportunities to take another. Is it the course taken that somehow impairs effectiveness, or the lack of a particular course that was forgone? Might it be that teacher candidates who are prone to take more mathematics pedagogy courses, for example, are less able to explain mathematical content to their students than the teachers who take an additional mathematics course? Additional research on these issues is essential for unpacking the current findings.

Another issue is the discrepancy between what is being taught and what is taken up by the teacher candidates. The curricula for TPPs may cover skills needed for the teacher candidates to improve student achievement, but the candidates may be unable to put those skills into practice after they graduate and reach the classrooms as teachers. It seems reasonable to attempt to assess uptake of the skills that are learning objectives for TPP coursework as a means of measuring the quality of the instruction in the TPP including the clinical experiences, as recommended by Diez (2010). As the U.S. National Research Council (2010) reported, there is limited

information on the qualifications of these individuals who often serve as clinical faculty. Perhaps, instead of qualifications of clinical faculty, which are measures of inputs, measures of effectiveness in terms of their teaching using value-added measures or their ability to increase teacher candidates' skills measured over the course of their clinical experiences should be used to assess the quality of these faculty.

A more valid and reliable assessment system for teacher candidates could also be used to provide information on the quality of instruction and supervision in TPP programs. The edTPA (formerly the Teacher Performance Assessment), developed by the Stanford Center for Assessment, Learning, and Equity (SCALE) and the American Association of Colleges for Teacher Education (AACTE) and field tested in some 25 states, shows promise as a summative measure, but as of spring 2013, the predictive validity of the instrument has not yet been fully established.² Other assessments of mastery of course learning objectives and ability to apply the skills in practice will be needed to fill out the assessment system.

The findings suggest that placement in the same grade where student teaching was performed has positive impacts on student learning. The placement of teachers on the job is typically beyond the control of TPPs. However, opportunities for classroom involvement that might include more experiences in different grades or summer orientation sessions that

provide grade-specific information for planning could be a possible avenue to prepare teachers better for their specific assignments. Some new teacher support programs do provide “backward planning” support for beginning teachers in which they begin to break the annual student learning goals for their grade and subject into shorter term objectives and plan for lessons to facilitate the students achieving the goals. A tighter relationship between the preparation program and the program graduate as they transition into the classroom could help to make novice teachers more effective and provide useful feedback for the improvement of the TPP.

An implication of Diez’ analysis, one that we readily concede, is that relatively gross variables such as the number of courses taken in a subject or overall GPA may not measure what is really important about teacher preparation. More nuanced factors such as interactions between syllabus-specific aspects of courses, pedagogical approach and faculty members’ teaching skill, and features of the receiving schools and classrooms may be at play. We began with the “low hanging fruit” of existing administrative data, not because we were convinced that the variables whose effects we estimated were necessarily the correct ones, but because these are typical of the measures of teacher candidates that are readily available in most TPPs and they are the only types of data that are routinely collected in most TPP concerning teacher candidate performance. Finer grained, more nuanced measures will be time-consuming and expensive to collect. More measurement-intensive studies should be undertaken, especially those that include measures of constructs that are likely to influence teachers effectiveness in the classroom, but they are likely to take a decade or more to produce results to develop and test teacher candidate assessment instruments that are reliable and usable for program improvement purposes.

In this era of accountability, many TPPs are beginning to systemize the data collection process and use the data in an effort to build internal capacity to assess the quality of the program. Currently, TPPs have used assessment tools and classroom teaching strategies with *hopes* that teacher candidates would leave the program prepared for the teaching profession. If TPPs begin to monitor the effectiveness of their programs using student achievement data and other student outcome measures from the classrooms in which their graduates teach *and* develop or identify and implement reliable and valid instruments to assess the progress and performance of their teacher candidates, the veil can be lifted and programs will have *actual* evidence of areas that are successful and those where improvement is necessary. Until then, all they can do is to hope that program reforms improve the effectiveness of their graduates. This study represents a step in that process by showing that the instruments used to assess the progress and performance of teacher candidates in one large TPP did not provide reliable information on the constructs they were designed to measure and lacked predictive validity.

A clear limitation is that the study examined the instruments and indicators from only one TPP and only at the

elementary level. It does establish that predictive validity studies can be done for individual programs and with the widespread availability of teacher value-added scores, the study can be replicated elsewhere. While the predictive validity of the measures of teacher candidates’ progress and performance can be assessed by individual programs, developing instruments and testing them for validity may better be undertaken by consortia of programs or state or federal organizations, thus reducing the costs of development and testing significantly. The edTPA, mentioned above, is an example of this type of effort. In addition to joint development and testing of such instruments, collaboration in collecting and analyzing comparable data across programs would provide larger samples and greater variation in the data—advantages for program and improvement purposes.

Another limitation is that the study included student mathematics and reading test scores as its only outcome measures. Student learning and teachers’ ability to facilitate and enhance their students’ learning is an important outcome and VAMs based on these scores measure teacher effectiveness on this one outcome. However, teachers must do much more than raise test scores to be truly effective. They must establish warm, productive relationships with their students and colleagues. They must provide students with opportunities for active engagement with the content and provide leadership within the classroom, school, and community. Of course, these outcomes and skills are not as readily measured as student test scores and may be difficult to attribute to specific teachers rather than the faculty and school leaders as a whole. The considerable challenge of measurement development should not constrain efforts to develop a broader range of measures of teacher effectiveness and to use them in studies of the effectiveness of TPPs.

In sum, the present study points to work that is needed if TPPs are to make evidence-based improvements in their programs. Developing instruments with face validity is not sufficient to guide the needed improvements. Instruments that yield measures with high levels of predictive validity are required if progress is to be made in preparing more effective teachers.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The University of North Carolina General Administration Teacher Quality Research project partially supported the research that is reported in this article.

Notes

1. In October 2010, the governing bodies of National Council for Accreditation of Teacher Education (NCATE) and Teacher

Education Accreditation Council (TEAC) voted to create the Council for the Accreditation of Educator Preparation (CAEP) to consolidate the accreditation work of the two agencies; but because the CAEP is not expected to complete the consolidation and/or expansion of their standards until December 2012, in this article we focus on the two accrediting bodies.

2. A report addressing the issues of reliability and validity is projected for completion during the summer or fall of this year (retrieved from <http://edtpa.aacte.org/about-edtpa>, May 2013).

References

- Bennett, C. I., McWhorter, L. M., & Kuykendall, J. A. (2006). Will I ever teach? Latino and African American students' perspectives on PRAXIS I. *American Educational Research Journal, 43*(3), 531-575.
- Boyd, D. J., Grossman, P., Lankford, H., Loeb, S., Michelli, N. M., & Wyckoff, J. (2006). Complex by design: Investigating pathways into teaching in New York City schools. *Journal of Teacher Education, 57*(2), 155-166.
- Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis, 31*(4), 416-440.
- Cook, D. W., & van Cleef, D. W. (2000). Multicultural perceptions of 1st-year elementary teachers' urban, suburban, and rural student teaching placements. *Urban Education, 35*, 165-174.
- Council of Chief State School Officers. (2011, April). *Interstate Teacher Assessment and Support Consortium (InTASC) model core teaching standards: A resource for state dialogue*. Washington, DC: Author.
- Delandshere, G., & Arens, S. A. (2003). Examining the quality of the evidence in preservice teacher portfolios. *Journal of Teacher Education, 54*, 57-73.
- Delandshere, G., & Petrosky, A. R. (1998). Assessment of complex performances: Limitations of key measurement assumptions. *Educational Researcher, 27*(2), 14-24.
- Diez, M. E. (2010). It is complicated: Unpacking the flow of teacher education's impact on student learning. *Journal of Teacher Education, 61*(5), 441-450.
- Dybdahl, C. S., Shaw, D. G., & Edwards, D. (1997). Teacher testing: Reason or rhetoric. *Journal of Research and Development in Education, 30*(4), 248-254.
- Educational Testing Service. (n.d.). Retrieved from <http://www.ets.org/praxis/nc/requirements/>
- Flowers, C. (2006). Confirmatory factor analysis of scores on the clinical experience rubric: A measure of dispositions of preservice teachers. *Educational and Psychological Measurement, 66*(3), 478-488.
- 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.
- Ginsberg, R., & Whaley, D. (2003). Admission and retention policies in teacher preparation programs: Legal and practical issues. *The Teacher Educator, 38*(3), 169-189.
- Greenberg, J., Pomerance, L., & Walsh, K. (2011). *Student teaching in the United States*. Washington, DC: National Council on Teacher Quality.
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics, 95*(7-8), 798-812.
- Hebert, E., & Worthy, T. (2001). Does the first year of teaching have to be a bad one? A case study of success. *Teaching and Teacher Education, 17*(8), 897-911.
- Henry, G. T., Bastian, K. C., Fortner, C., Kevin, K., David, C., Purtell, K. M., & Zulli, R. A. (in press). Teacher preparation policies and their effects on student achievement. *Education Finance and Policy*.
- Henry, G. T., Kershaw, D. C., Smith, A. A., & Zulli, R. A. (2012). Incorporating teacher effectiveness into teacher preparation program evaluation. *Journal of Teacher Education, 63*(5), 335-355.
- Henry, G. T., Thompson, C. L., Bastian, K. C., Fortner, C. K., Kershaw, D. C., Marcus, J. V., & Zulli, R. A. (2011). *UNC teacher preparation program effectiveness report*. Chapel Hill, NC: Carolina Institute for Public Policy.
- Hicken, S. (1992). The Pre-Professional Skills Test: How it affects teacher aspirants and predicts performance in teacher preparation programs. *The Urban Review, 24*(4), 253-261.
- Kukla-Acevedo, S. (2009). Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement. *Economics of Education Review, 28*(1), 49-57.
- Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher, 31*(1), 3-12.
- Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198-1202.
- McDiarmid, G. W., & Clevenger-Bright, M. (2008). Rethinking teacher capacity. In M. Cochran-Smith, S. Feiman-Nemser, D. J. McIntyre, & K. E. Demers (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (3rd ed., pp. 134-156). New York, NY: Routledge.
- Mikitovics, A., & Crehan, K. D. (2002). Pre-Professional Skills Test scores as college of education admission criteria. *Journal of Educational Research, 95*(4), 215-223.
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017-1054.
- Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*(2), 125-145.
- National Council for Accreditation of Teacher Education. (2008). *Professional standards for the accreditation of teacher preparation institutions*. Washington, DC: Author.
- National Council for Accreditation of Teacher Education. (2010). *Transforming teacher education through clinical practice: A national strategy to prepare effective teachers*. Washington, DC: Author.
- Ogren, C. A. (2005). *The American state normal school: An instrument of great good*. New York, NY: Palgrave Macmillan.
- Pool, J., Dittrich, C., Longwell, E., Pool, K., & Hausfather, S. (2004). An analysis of SAT and PRAXIS I performance of teacher education candidates at three different types of institutions. *Action in Teacher Education, 26*(2), 60-68.
- Ronfeldt, M. (2012). Where should teachers learn to teach? Effects of field placement school characteristics on teacher retention and effectiveness. *Educational Evaluation and Policy Analysis, 34*(1), 3-26.
- Rose, R. A., & Fraser, M. W. (2008). A simplified framework for using multiple imputation in social work research. *Social Work Research, 32*(3), 171-178.
- Rubin, D. B. (1987). Comment. *Journal of the American Statistical Association, 82*(398), 543-546.

- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London, England: Chapman & Hall.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147-177.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.
- Shulman, L. S. (1987). Assessment for teaching. An initiative for the profession. *Phi Delta Kappan*, 69(1), 38-44.
- Tabachnick, B., & Fidell, L. (1989). *Using multivariate statistics*. Northridge, CA: Harper Collins.
- U.S. National Research Council. (2010). *Preparing teachers: Building evidence for sound policy*. Washington, DC: National Academies Press.
- Wakefield, D. (2003). Screening teacher candidates: Problems with high-stakes testing. *Educational Forum*, 67(4), 380-388.
- Watanabe, S. (2008). "Because we do not know their way": Standardizing practices and peoples through habitus, the NCLB "highly-qualified" mandate, and PRAXIS I examinations. *Journal of American Indian Education*, 47(1), 118-135.
- Wilson, S., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations* (Working paper). Seattle, WA: University of Washington, Center for the Study of Teaching and Policy. Retrieved from <http://depts.washington.edu/ctpmail/PDFs/TeacherPrep-WFFM-02-2001.pdf>
- Wolf, K. (1991). The schoolteacher's portfolio: Issues in design, implementation, and evaluation. *Phi Delta Kappan*, 73(2), 129-136.
- Young, A., & Wilkins, E. (2008). Dispositional statements on student teacher evaluation instruments: Commonalities across institutions. *The Teacher Educator*, 43(3), 198-215.
- Leadership, Policy and Organization, Peabody College, Vanderbilt University. His research interests include teacher quality, educational reform and evaluation.
- Shanyce L. Campbell** is a graduate research fellow at the Education Policy Initiative at Carolina and doctoral student in the Department of Public Policy at the University of North Carolina at Chapel Hill. Her research interests include teacher quality, teacher distribution, educational inequality, and program evaluation.
- Charles L. Thompson** is a director of Teacher Quality Research in the Carolina Institute for Public Policy at the University of North Carolina at Chapel Hill.
- Linda A. Patriarca** is the dean of the College of Education and professor of special education at East Carolina University. Her research interests include evidence-based practices, collaboration, and organizational change.
- Kenneth J. Luterbach** is an assistant professor in the Department of Mathematics, Science, and Instructional Technology Education, College of Education, East Carolina University. His interests include research and development of instruction, particularly instruction and computer applications for learning and performance, teacher education, and systemic school reform.
- Diana B. Lys** is the director of Assessment and Accreditation in the College of Education at East Carolina University. Her research interests include assessment of student learning outcomes, innovations in teacher preparation, middle-level education, and collaborative partnerships in education.
- Vivian Martin Covington** is the executive director of the Office of Teacher Education in the College of Education at East Carolina University. Her research interests include teacher recruitment and retention; partnership among schools, community colleges, and the university; and assessment of teacher education programs.

Author Biographies

Gary T. Henry holds the Patricia and Rodes Hart Chair and serves as professor of public policy and education in the Department of