



## Evaluating Prospective Teachers: Testing the Predictive Validity of the edTPA.

Dan Goldhaber  
American Institutes of Research  
Center for Education Data & Research  
University of Washington Bothell

James Cowan  
American Institutes of Research

Roddy Theobald  
American Institutes of Research

## Contents

---

<b>Acknowledgements .....</b>	<b>iii</b>
<b>Abstract.....</b>	<b>iv</b>
<b>I. Background: The Teacher Education Accountability Movement .....</b>	<b>1</b>
<b>II. Assessment of Prospective Teachers and the Role of the edTPA .....</b>	<b>4</b>
<b>III. Data and Analytic Approach .....</b>	<b>8</b>
<b>IV. Results .....</b>	<b>15</b>
<b>V. Policy Implications.....</b>	<b>24</b>
<b>VI. Conclusion .....</b>	<b>25</b>
<b>References.....</b>	<b>29</b>
<b>Tables .....</b>	<b>33</b>
<b>Figures.....</b>	<b>38</b>

## Acknowledgements

---

This work is supported by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER) (IES grant #R305C120008), the Bill and Melinda Gates Foundation (grant #OPP1128040), and an anonymous foundation.

We wish to thank Trevor Gratz for outstanding research assistance and Joe Koski, John Krieg, Gerhard Ottehenning, Ray Pecheone, Amy Vaughn, Jennifer Wallace, and Andrea Whitaker for helpful comments.

CEDR working papers have not undergone final formal review and should be cited as working papers. They are intended to encourage discussion and suggestions for revision before final publication. The views expressed in this paper do not necessarily reflect those of American Institutes for Research or the University of Washington. Any and all remaining errors are our own.

Suggested citation:

Goldhaber, D., Cowan, J., & Theobald, R. (2016). Evaluating Prospective Teachers: Testing the Predictive Validity of the edTPA. CEDR Working Paper 2016-2. University of Washington, Seattle, WA.

*© 2016 by Dan Goldhaber, James Cowan, Roddy Theobald. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission, provided that full credit, including © notice, is given to the source.*

You can access other CEDR publications at  
<http://www.CEDR.us/publications.html>

## **Evaluating Prospective Teachers: Testing the Predictive Validity of the edTPA**

Dan Goldhaber, James Cowan, Roddy Theobald

CEDR Working Paper No. 2016-2

May 2016

### **Abstract**

Given the rapid policy diffusion of the edTPA, a performance-based, subject-specific assessment of teacher candidates, it is surprising that there is currently no existing large-scale research linking it to outcomes for inservice teachers and their students. We use longitudinal data from Washington State that include information on teacher candidates' scores on the edTPA to provide estimates of the extent to which edTPA performance is predictive of the likelihood of employment in the teacher workforce and value-added measures of teacher effectiveness. While edTPA scores are highly predictive of employment in the state's public teaching workforce, evidence on the relationship between edTPA scores and teaching effectiveness is more mixed. Specifically, continuous edTPA scores are a significant predictor of student mathematics achievement in some specifications, but when we consider that the edTPA is a binary screen of teaching effectiveness (i.e., pass/fail), we find that passing the edTPA is significantly predictive of teacher effectiveness in reading but not in mathematics.

## I. Background: The Teacher Education Accountability Movement

It is fair to say that teacher education programs are facing significant scrutiny over the inservice performance of their graduates.<sup>1</sup> About 75% of the roughly 100,000 novice teachers who enter the public school workforce each year are trained in a traditional college or university setting, and there is significant policy concern that the preparation that prospective teachers receive is not adequate to ensure they are ready to teach on their first day in a classroom.<sup>2</sup> Former Education Secretary Arne Duncan, for instance, stated that: “By almost any standard, many if not most of the nation’s 1,450 schools, colleges and departments of education are doing a mediocre job of preparing teachers for the realities of the 21st century classroom” (U.S. Department of Education, 2009).<sup>3</sup>

Given this environment, it is not surprising that there are a number of new initiatives designed to hold teacher education programs (TEPs) more accountable, either through direct measures of the training they provide teacher candidates or based on output measures, such as the value added of candidates who enter the teaching workforce.<sup>4</sup> One of the ways that TEPs and states have responded to this increased accountability pressure is by adopting the edTPA, a performance-based, subject-specific assessment that is administered to teacher candidates during their student teaching assignment. There

---

<sup>1</sup> The focus on teacher education is well-founded given empirical evidence about the large impact of teacher quality on student outcomes; see, for instance, Aaronson et al. (2007), Chetty et al. (2014b), Hanushek and Rivkin (2010), and Rivkin et al. (2005).

<sup>2</sup> For more on teacher preparation and the evaluation of teacher preparation institutions, and the increasing reliance on alternative routes, see, for instance, Crowe (2011) and National Research Council (2010).

<sup>3</sup> More recently a collaboration between the National Council on Teacher Quality and *U.S. News and World Report* published ratings of the nation’s teacher training programs. This review, while suggesting a few bright spots was largely a scathing indictment of the training most teachers receive, concluding that teacher training programs “have become an industry of mediocrity, churning out first-year teachers with classroom management skills and content knowledge inadequate to thrive in classrooms” (Greenberg et al., 2013).

<sup>4</sup> See, for instance: recent efforts at market-based accountability through publicized ratings of TEPs (Greenberg et al., 2013), the U.S. Department of Education’s proposed state and institutional reporting requirements in the extension of the Higher Education Act (Sawchuk, 2014), and the standards adopted by the major national accrediting body, the Council for the Accreditation of Educator Preparation (CAEP, 2013). For more background on research tying student achievement to the institution from which teachers receive their teaching credential, see Goldhaber (2014).

has been remarkably rapid policy diffusion of this assessment from its initial field testing in 2012 to full implementation (Gottlieb et al., 2016): The edTPA is now used by over 600 TEPs in 40 states, and passing the edTPA is a requirement for licensure in seven states.<sup>5</sup> Yet despite the rapid adoption of this assessment, critics of the edTPA (e.g., Greenblatt & O'Hara, 2015) point out that there is currently no existing large-scale research linking edTPA scores to outcomes for inservice teachers and their students.<sup>6</sup>

There are several theories of action for how the edTPA might improve the quality of the teacher workforce (e.g., Hill et al., 2011). First, the edTPA can be used as a high-stakes screen that prevents low-performing teacher candidates from receiving teaching credentials; this is how the edTPA is currently used in states in which the assessment is a determining factor in whether teacher candidates are eligible to participate in the labor market.<sup>7</sup> This use of the edTPA requires predictive validity around the cut point adopted for labor market participation.

The edTPA might also improve the quality of the teaching workforce by affecting candidate teaching practices. Indeed, the edTPA is described by its developers as an “educative assessment” that “supports candidate learning and preparation program renewal” (edTPA, 2015). This could occur at the individual teacher candidate level if, for instance, participation in the edTPA directly influences the teaching practices of teacher candidates. Alternatively, this could occur at the TEP level if, for instance, participation in the influences the training provided by TEPs. Finally, the edTPA might be used for hiring purposes; for instance, school systems might be more likely to hire teacher applicants with higher edTPA scores. Each of these potential mechanisms for workforce improvement requires that the edTPA provides a signal of quality teaching; i.e., that there is predictive validity away from the cut point such

---

<sup>5</sup> See <http://edtpa.aacte.org>

<sup>6</sup> The edTPA was initially developed by researchers at Stanford University’s Center for Assessment, Learning, and Equity (SCALE), who have a stated commitment to “predictive validity studies that follow candidates into employment if the state database enables linking teachers to classrooms” (SCALE, 2015).

<sup>7</sup> For a full summary of edTPA participation across the country, see edTPA (2015), p. 13.

that differences in edTPA performance (at the candidate or institution level) might be indicative of teacher quality.

In this paper we use longitudinal data from Washington State that includes information on teacher candidates' scores on the edTPA to provide estimates of the extent to which edTPA scores are predictive of the likelihood of entry into the teacher workforce and value-added measures of teacher effectiveness (i.e., predictive validity). Specifically, we test different theories of action for how the edTPA might improve the quality of the teacher workforce by considering the predictive validity of the edTPA as both a screen and a signal of future teacher effectiveness.

Despite the fact that the edTPA was not consequential for some of the teacher candidates in our sample, we find that edTPA scores—both in terms of passing status and continuous scores—are highly predictive of the probability that a teacher candidate is employed the following year in the state's public teaching workforce. Evidence on the connection between performance and value-added measures of teacher effectiveness is more mixed. Continuous edTPA scores provide a signal of future teaching effectiveness in mathematics in some specifications, but are not statistically significant in reading. When we consider the edTPA as a binary screen of teaching effectiveness (i.e., pass/fail), we find that passing the edTPA is significantly predictive of teacher effectiveness in reading but not in mathematics.

The rest of the paper proceeds as follows: In Section II, we provide additional information regarding teacher licensure and the edTPA in particular. We describe our data and analytic approach in Section III, present our findings in Section IV, outline some extensions in Section V, and offer concluding remarks in Section VI.

## II. Assessment of Prospective Teachers and the Role of the edTPA

There are various ways that teacher candidates are typically assessed and judged to be eligible—that is, licensed—to teach in public schools.<sup>8</sup> Licensure in many states requires that prospective teachers graduate from an approved TEP and complete some preservice student teaching, although the last decade has also seen an increased reliance on teachers entering the profession through state-approved alternative routes.<sup>9</sup> Forty-nine of 50 states also require potential teachers to pass licensure tests that cover basic skills, content knowledge, and/or professional knowledge.<sup>10</sup>

The edTPA, by design, is quite different from traditional question-and-answer licensure tests: It is a portfolio-based, subject-specific assessment akin to the National Board for Professional Teacher Standards (NBPTS) assessment of inservice teachers.<sup>11</sup> The edTPA was initially developed by researchers at Stanford University’s Center for Assessment, Learning, and Equity (SCALE) and has been further developed and distributed through a partnership between SCALE, the American Association of Colleges for Teacher Education (AACTE), and Evaluation Systems (a member-organization of the Pearson Education group). The edTPA was initially introduced in two large-scale field tests in 2011–12 and 2012–13, and was “operationally launched” in 2013–14 (Pecheone et al., 2013). The edTPA relies on the scoring of teacher candidates who are videotaped while teaching three to five lessons from an instructional unit to one class of students, along with assessments of teacher lesson plans, student work samples and evidence of student learning, and reflective commentaries by the candidate. Candidates

---

<sup>8</sup> Note that the terms licensure and certification are often used interchangeably in the teacher labor market.

<sup>9</sup> For more background on the teacher licensure process and alternative routes into teaching, see Goldhaber (2011).

<sup>10</sup> A number of studies explore the relationship between teacher performance on licensure tests and student test achievement. Many (e.g., Clotfelter et al., 2007, 2010; Goldhaber, 2007; Goldhaber et al., 2016a; Hendricks, 2015) but not all (Buddin and Zamarro, 2010) find a significant positive relationship between *some* tests and test achievement. Goldhaber (2007), for example, finds that students in North Carolina assigned to a teacher who passed the state’s licensure test perform .02-.06 standard deviations higher on end-of-year tests, all else equal, than students assigned to a teacher who failed the test. While licensure tests are typically only used as a pass/fail signal, prior research indicates that these tests also provide predictive value about future teacher effectiveness for candidates whose scores are far from the employment eligibility cut point.

<sup>11</sup> For empirical evidence on the relationship between NBPTS assessments and student achievement, see Goldhaber and Anthony (2007) and Cowan and Goldhaber (2015).

pay a \$300 fee to take the edTPA and often take several months to prepare their portfolios for submission (e.g., Jette, 2014).

The edTPA is a subject-specific assessment with different versions aligned with 27 different teaching fields (e.g., “Early Childhood”, “Secondary Mathematics”, etc.). Each of these versions of the edTPA contains 15 different rubrics, each of which is scored on a 1–5 scale; the rubrics have equal weight so the range of possible summative scores (for tests with no incomplete rubric scores) is 15 to 75.<sup>12</sup> The 15 rubrics that are used to calculate a candidate’s summative score in Washington State are grouped into three areas: Planning (e.g., “Planning for Subject-Specific Understandings”); Instruction (e.g., “Engaging Students in Learning”); and Assessment (e.g., “Analysis of Student Learning”).<sup>13</sup> Teacher candidates in Washington State are also scored on three additional Student Voice rubrics (e.g., “Eliciting Student Understanding of Learning Targets”), which are designed to incorporate student-produced material into a teacher’s evaluation. For reasons discussed in the next section, these rubric scores are not currently used in computing a candidate’s summative score.<sup>14</sup>

Proponents of the edTPA argue that the assessment is an authentic measurement tool that can be used to predict teacher candidates’ success in the classroom (e.g., Darling-Hammond et al., 2009; edTPA, 2015; Hill et al., 2011). While the edTPA is designed to assess individual teacher candidates, it is also thought to inform improvements in TEPs. Some states are, in fact, using the average edTPA performance of teacher candidates at an institution as a measure of institutional quality and/or in the accreditation process. In addition, the use of the edTPA is heavily promoted by AACTE, which touts the

---

<sup>12</sup> Candidates may receive an incomplete score on any of the 15 rubrics for having technical issues with the upload, uploading an incomplete file, having an edited video, or uploading material that is not related to the handbook. If a candidate received only one incomplete score, it counts as a zero in the calculation of the final summative score; but the summative score is incomplete if the candidate receives an incomplete on two or more rubrics.

<sup>13</sup> We performed a principal component analysis on the 15 rubric scores and found that the rubric scores load onto three factors that align closely with these areas.

<sup>14</sup> The national edTPA handbook for elementary education also includes three additional Mathematics Assessment rubrics (e.g., “Analyzing Whole Class Misunderstandings”) that have not been adopted in Washington State.

assessment as a means of improving “...the information base guiding the improvement of teacher preparation programs [and] strengthen[ing] the information base for accreditation and evaluation of program effectiveness.”<sup>15</sup>

Claims about the predictive validity of the edTPA are primarily based on small-scale pilot studies of the edTPA’s precursor, the Performance Assessment for California Teachers (PACT).<sup>16</sup> Specifically, Newton (2010) finds positive correlations between PACT scores and future value-added for a group of 14 teacher candidates, while Darling-Hammond et al. (2013) use a sample of 52 mathematics teachers and 53 reading teachers and find that a one-standard deviation increase in PACT scores is associated with a .03 standard deviation increase in student achievement in either subject.<sup>17</sup> Beyond the fact that these estimates are based on small sample sizes, however, there are several substantive differences between the edTPA and PACT in terms of scoring, implementation, and standards alignment.<sup>18</sup>

As described in the next section, the administrative data we utilize for our research allows us to leverage a larger sample size of teachers (over 200 in both mathematics and reading) than the PACT studies cited above. Each of these teachers took the edTPA after its full national implementation in the 2013–14 school year. It is important to note, however, that the edTPA did not become consequential in

---

<sup>15</sup> <http://edtpa.aacte.org/about-edtpa#Goals-1>. The edTPA is also promoted as an alternative to value-added based assessments of TEPs; AACTE has taken the position that value-added measures should not be used to assess TEPs (<https://aacte.org/news-room/press-releases-statements/486-aacte-on-behalf-of-800-teacher-prep-programs-submits-comments-on-proposed-federal-regulations>).

<sup>16</sup> The 2014 edTPA administrative report states that “Preliminary data from studies by Benner and Wishart (2015) has revealed that edTPA scores predict candidates’ ratings of teacher effectiveness, as measured by a composite score that combines students’ performance data and classroom observations” (edTPA, 2015). However, these data have never been published, and follow-up documentation from the authors suggests that these relationships are more mixed than this quote suggests (personal communication, May 2016).

<sup>17</sup> Darling-Hammond et al. (2013) report nearly identical point estimates as those reported in this paper but with substantially more precision using a considerably smaller sample than is available in this paper. We attempted to replicate their findings using differing assumptions regarding the appropriate level of clustering and could only estimate coefficients with similar levels of precision in models that assume independent errors across students in the same classroom. We attempted to compare modeling choices directly, but in discussions with the authors, we were unable to do so as they no longer have their data files (personal communication, February 2016).

<sup>18</sup> See <http://www.ctc.ca.gov/commission/agendas/2012-09/2012-09-2F.pdf>

Washington State until January 2014<sup>19</sup>, so candidates who failed the test in fall 2013 (as well as candidates who failed after January 2014 but subsequently re-took and passed the test) provide an opportunity to observe candidates who failed the test but still entered the public teaching workforce.

While this study is one of the first to provide evidence on the validity of the edTPA as a measure of classroom performance, it is important to distinguish the validity of the edTPA as an assessment of teaching practice from its efficacy as a teacher licensing tool. In particular, while validity is a significant prerequisite for using the edTPA to support effective licensure policy, extrapolating from these results to the effects of particular policies requires imposing additional assumptions beyond those that we test here.

In particular, four features of common licensure policies limit such additional conclusions. First, licensure policies may change the population of potential teachers if candidates view the test as costly. There is some evidence from changes to state licensing provisions that licensure tests discourage some candidates with high academic achievement or outside wage offers from pursuing teaching as a profession, although evidence on overall effects on student achievement is mixed (Angrist & Guryan, 2008; Larsen, 2015; Wiswall, 2007). Second, policies typically allow candidates to attempt the assessment multiple times. In the second half of the 2013–14 school year (when the edTPA was consequential), 4% of test takers failed the edTPA the first time they took it, but about half of these candidates eventually passed the test. Third, the matching of teacher candidates to teaching positions may provide additional screening beyond what is required by law. For example, it is not clear that the small number of teachers in our sample who never pass the edTPA would obtain employment even in the absence of testing requirements. Finally, licensure systems like the edTPA might have system-wide effects on teacher quality. If participation in the edTPA raises overall performance, the signaling effects

---

<sup>19</sup> See <http://assessment.pesb.wa.gov/faq/edtpa-policies>

we estimate here may understate the overall effects of implementing testing requirements. The policy effects of national implementation of the edTPA, and similar authentic licensure assessments, therefore remains an important area for future research.

### III. Data and Analytic Approach

#### 1. Data

Our research uses administrative data on teacher candidates provided by Washington State's Professional Educator Standards Board (PESB), as well as data on Washington State students, teachers, and schools maintained by the Office of the Superintendent of Public Instruction (OSPI). The PESB data includes scores on each individual edTPA rubric (as well as the final summative score) for *all* teacher candidates who took the edTPA in Washington State, not just those who ultimately are employed in the teacher workforce. As described in the previous section, the 15 rubrics used to compute the summative score can be combined into three subscores: Planning (rubrics 1–5), Instruction (rubrics 6–10), and Assessment (rubrics 11–15).<sup>20</sup> While the three Student Voice rubrics can be combined into a separate subscore, Washington State does not count these rubrics toward a candidate's consequential score. Consequently, we do not consider them in our primary analysis.<sup>21</sup>

Washington State participated in the edTPA field test in the 2012–13 school year (see Pecheone et al., 2013), and the PESB data include teacher candidate scores from this pilot year and two subsequent school years (2013–14 and 2014–15) after the full national implementation of edTPA. Because there were substantive changes to the assessment between the pilot year and full implementation (edTPA, 2015), and because inservice data are not yet available for teacher candidates

---

<sup>20</sup> The correlations between the three subscores range from 0.598 to 0.661.

<sup>21</sup> That said, when we consider scores on the Student Voice rubrics as a predictor of teaching effectiveness in mathematics and reading, the coefficients are positive but not statistically significant. Results are available from the authors upon request.

who took the edTPA in 2014–15, our primary results focus on the 2,362 teacher candidates from Washington State TEPs who took the edTPA in the 2013–14 school year. In most cases, we consider edTPA scores from each candidate’s first test administration, although in cases where a candidate received an incomplete score and subsequently resubmitted his or her materials within a month, we disregard the initial incomplete score and consider a candidate’s subsequent submission.<sup>22</sup>

We link these edTPA scores to data from OSPI that include test scores on other licensure tests that teacher candidates must also pass in order to be eligible to teach, such as the Washington Educator Skills Test-Basic (WEST-B), an assessment of basic skills in reading, writing, and mathematics that has been a requirement for admission into Washington State TEPs since 2002.<sup>23</sup> Among teacher candidates in the edTPA sample, 60.29% entered the state’s public teaching workforce in the 2014–15 school year (defined as being employed in a certificated teaching position), and for these 1,424 teacher candidates, the OSPI data also include information about their school assignments, race, gender, and ethnicity.

For the subset of 277 teacher candidates who enter the workforce and teach mathematics or reading in Grades 4–8 (i.e., grades and subjects in which both current and prior test scores are available, or the value-added sample), we can investigate the relationship between edTPA performance and student achievement. Specifically, we observe annual student test scores in mathematics and reading in Grades 3–8 (also provided by OSPI) on the state’s Measures of Student Progress (MSP) examination in 2012–13 and 2013–14 and Smarter Balanced Assessment (SBA) in the 2014–15 school year.<sup>24</sup> We

---

<sup>22</sup> As discussed in the previous section, if a candidate received only one incomplete score, that score counts as a zero in the calculation of the final summative score; but the summative score is incomplete if the candidate receives an incomplete on two or more rubrics. We therefore drop incomplete scores in cases where the candidate resubmits materials within a month of the score reporting date. We experimented with models that consider all incomplete scores as failures and found similar results.

<sup>23</sup> Some alternative licensing exams may be submitted instead of taking the WEST-B. Thus, not all prospective teachers take the WEST-B (RCW 28A.410.220 & WAC 181-01-002).

<sup>24</sup> About one third of Washington State schools participated in the state’s Smarter Balanced Assessment pilot in the 2013–14 school year, so test scores are not available in 2013–14 for students in these schools. We discuss our approach to these missing data in the analytic approach section.

standardize these scores within grade and year and connect them to additional student demographic information (gender, race/ethnicity, special education status, free/reduced-priced lunch eligibility, and English learner status) and, through a unique link in the state's Comprehensive Education Data and Research System (CEDARS) data system, to data on the student's teachers in mathematics and reading (described above).<sup>25</sup>

**Table 1** summarizes data for prospective teachers who took the edTPA assessment in 2013–14 for all candidates (columns 1–6) and for candidates who appear in the teaching workforce in 2014-15 (columns 7–12). Within each set of columns, we present summary statistics for all individuals within the group (columns 1 and 7) and by quintile of performance on the edTPA (columns 2–6 and 8–12).<sup>26</sup> In column 1, we see that the overall first-time pass rate on the test, 93.9%, was quite high because Washington State had set a low cut score of 35, but this passing rate would have been only 86.5% had the state used its future cut score of 40.

The summary statistics for teacher candidates by quintile of performance on the edTPA (columns 2-6) make it clear that there is a correlation between edTPA performance and the WEST-B basic skills licensure tests that are required for entry into Washington State's TEPs.<sup>27</sup> It is also immediately clear that teachers who perform better on the edTPA are more likely to be employed in Washington State's public schools in the subsequent year: Only 50.8% of first-quintile (lowest-quintile) teachers are observed teaching versus 64.6% of fifth-quintile (top quintile) teachers.

---

<sup>25</sup> CEDARS data includes fields designed to link students to their individual teachers, based on reported schedules. However, limitations of reporting standards and practices across the state may result in ambiguities or inaccuracies around these links. We limit the student sample to students who received instruction from a single teacher in that subject and year.

<sup>26</sup> Note that the quintiles in this table are based on edTPA scores across multiple test types; but all models include fixed effects for test type (so candidates are compared only with other candidates who took the same test type).

<sup>27</sup> The correlations between continuous edTPA scores and the three WEST-B subtests are moderate ( $r = 0.20$  in mathematics and reading,  $r = 0.25$  in writing).

Turning to the summary statistics for candidates who appear in the teaching workforce, we observe fairly large differences among these individuals in terms of edTPA performance by teacher race. Specifically, Hispanic teachers are about twice as likely to score in the lowest quintile of the edTPA as in any other quintile.<sup>28</sup> However, since we do not observe race for teacher candidates who do not enter the workforce, we cannot make broader comparisons of performance across all test takers. Nonetheless, this raises the possibility that there may be tradeoffs between the edTPA requirement and the diversity of credentialed teachers in the state.

## 2. Analytic Approach

To investigate the relationship between edTPA scores and the probability of workforce entry, we first define  $p_{jkt}$  as the probability that teacher candidate  $j$  who took edTPA test type  $k$  in 2013–14 appears as a Washington State public school teacher in the 2014–15 school year and estimate a simple logit model:

$$\log\left(\frac{p_{jk}}{1-p_{jk}}\right) = \alpha_0 + \alpha_1 TPA_{jk} + \alpha_k + \varepsilon_{jk} \quad (1)$$

In the base specification of the model in equation 1,  $TPA_{jk}$  is a binary variable indicating whether teacher candidate  $j$  passed the edTPA on the first test sitting. Given that all specifications include fixed effects for test type  $k$ , all coefficients can be interpreted as relative to other teacher candidates who took the same test type.<sup>29</sup> Although the coefficient of interest  $\alpha_1$  is on the log odds scale, we present all estimates as average marginal effects. We also estimate three other specifications of the model in equation 1 in which: (1)  $TPA_{jk}$  is an indicator for whether candidate  $j$  *would have* passed the edTPA at the state’s future (and higher) cut score; (2)  $TPA_{jk}$  is a continuous variable indicating the edTPA score of

---

<sup>28</sup> This is consistent with research showing that performance on licensure tests varies across teacher candidate subgroups (Goldhaber and Hansen, 2010).

<sup>29</sup> As discussed in section II, there are 27 different versions of the edTPA, so this ensures that candidates are only compared to other candidates who completed the same test type.

candidate  $j$  (standardized relative to all test takers); and (3)  $TPA_{jk}$  is a vector of scores for candidate  $j$  across the three subscores on the test (each standardized relative to all test takers).

To investigate the predictive validity of the edTPA in terms of predicting the achievement of a teacher candidate's future students, we estimate value-added models (VAMs) intended to separate the impact of teacher characteristics (such as edTPA scores) from other variables that influence student test performance (see Koedel et al. [2015] for review). Specifically, we estimate variants of the following VAM:

$$Y_{ijgst} = \beta_0 + \beta_1 Y_{i,t-1} + \beta_2 X_{it} + \beta_3 C_{ist} + \beta_4 Z_{jt} + \beta_5 TPA_{jk} + \beta_g + \beta_k + \varepsilon_{ijkgst} \quad (2)$$

In equation 2,  $Y_{ijgst}$  is the SBA score of student  $i$  in grade  $g$ , subject  $s$ , and year  $t$  (the 2014-15 school year for all students), while in the classroom of teacher  $j$  who took edTPA test type  $k$ .  $Y_{i,t-1}$  is a vector of student  $i$ 's prior year test scores in mathematics and reading. The student test scores in both  $Y_{ijgst}$  and  $Y_{i,t-1}$  are standardized by test, grade, and year across all test takers. Therefore, the units of the coefficients on the right side of equation 2 are standard deviations of student performance (relative to other scores on the same test in the same grade and year).  $X_{it}$  is a vector of student covariates for student  $i$ , in year  $t$ , which includes indicators for race/ethnicity, gender, free or reduced-priced lunch eligibility, gifted/highly capable, limited English proficiency (LEP), special education, and learning disabled.  $C_{ist}$  is a vector of aggregated student characteristics in the student's classroom, while  $Z_{jt}$  an indicator for whether or not a teacher possesses an advanced degree in year  $t$ .<sup>30</sup> All specifications include fixed effects for grade  $g$  and test type  $k$ , so all results can be interpreted as relative to other students in the same grade whose teachers took the same edTPA test type.

---

<sup>30</sup> Note that we do not need to control for teaching experience because every teacher in the VAM sample is a first-year teacher.

The different specifications of the model in equation 2 correspond to the different theories of action discussed in the introduction. When we investigate the edTPA as a screening mechanism intended to prevent low-performing teachers from entering the workforce,  $TPA_{jk}$  is an indicator for whether candidate  $j$  passed the edTPA on the first test administration (or, in a related specification, *would have* passed the edTPA at the state’s future cut score). When we investigate the signal value of edTPA scores (i.e., the extent to which a candidate’s score could be used as a proxy for future teaching effectiveness),  $TPA_{jk}$  is the standardized edTPA score of candidate  $j$  (or, in a separate specification, a vector of standardized scores for candidate  $j$  across the test’s three subscores).

We estimate specifications with only test type fixed effects (so teachers are compared to other teachers who took the same test type), with test type and TEP fixed effects (so teachers are compared to other teachers who took the same test type and graduated from the same TEP), and with test type and school district fixed effects (so teachers are compared to other teachers who took the same test type and are teaching in the same school district).<sup>31</sup> We estimate equation 2 by ordinary least squares (OLS) and cluster standard errors at the teacher level to account for correlation between the errors of students taught by the same teacher.

One challenge in estimating all of these specifications is that approximately one-third of students in Grades 4–8 have missing prior-year test scores because their school participated in Washington State’s Smarter Balanced Assessment pilot in the 2013–14 school year (and the state did not collect their scores). We therefore estimate three types of models: (1) a listwise deletion model that drops all students with missing prior-year test scores (possible in Grades 4–8); (2) an imputation model that uses twice-lagged test scores to impute lagged test scores for students with missing test scores (possible in Grades 5–8); and (3) a stacked model that considers any student with either once-lagged

---

<sup>31</sup> We also experiment with school fixed effects models, but a relatively small number of teachers in the VAM sample teach in the same school as compared with other teachers who took the edTPA.

scores, twice-lagged scores, or both and uses missing-value dummies to account for missing data (possible in Grades 4–8). We present primary results from the stacked models because they are based on the largest sample sizes, but estimates from the other models show that the results are not sensitive to these sample considerations.<sup>32</sup>

The broader VAM literature (e.g., Chetty et al., 2014; Kane et al., 2013) suggests that the VAMs described above account for the potential non-random sorting of students to teachers in the sample. A second concern, however, is the potential for sample selection bias. As is the case with other licensure tests (see Goldhaber et al., 2016a), sample selection is a concern if teacher characteristics not captured by the edTPA are relevant for hiring decisions and contribute to teacher effectiveness. The literature on teacher hiring suggests that this is likely to be the case. For example, administrative and survey evidence suggests that references, interviews, and personality traits are important predictors of employment outcomes, and that several of these measures are related to student achievement (Goldhaber et al., 2014a; Harris and Sass, 2014; Jacob et al., 2016; Rockoff et al., 2011). Consequently, teachers who perform poorly in the domains measured by the edTPA but who appear in our sample are likely hired because they possessed some compensating skill or skills that make them more effective teachers. In other words, the candidates we observe with low scores are probably disproportionately high-performing teachers.

We explore this issue empirically in Section IV below, but we argue that two factors are likely to limit the selection bias in our application. First, we examine the edTPA at a time when it was not fully binding in Washington State. Given the lower cut score and the ability of failing teacher candidates to retake the assessment, the selection probabilities between initial passing candidates and initial failing candidates are not as substantial as they would be if the testing requirement was fully binding.

---

<sup>32</sup> Results are available from the authors upon request.

Second, while non-tested teacher skills appear related both to hiring decisions and to teacher effectiveness, this relationship is not particularly strong. For example, analyses of the kinds of subjective data available to hiring authorities suggest that, when combined with observable and objective measures of teacher skill, these measures explain only 10% to 20% of the variation in teacher effectiveness (Goldhaber et al., 2014a; Jacob et al., 2016; Rockoff et al., 2011). Results from Jacob et al. (2016) suggest a similar relationship to the probability that a candidate for a position is hired.

## IV. Results

In this section, we describe our primary research findings on the extent to which edTPA scores predict: The likelihood of being in the Washington State public teacher workforce (**Table 2**); teacher effectiveness in reading (**Table 3**); and teacher effectiveness in mathematics (**Table 4**). Effectiveness results are also illustrated in scatterplots relating candidate edTPA scores to teacher value added in reading (**Figure 1**) and mathematics (**Figure 2**). Before discussing our primary findings, however, a few peripheral findings are worth brief mention.<sup>33</sup> In terms of predicting employment in the Washington State teacher labor market, we find both that individual TEPs are associated with different probabilities of employment and that candidates who took the edTPA in a STEM area are more likely to be employed than are candidates who took an elementary edTPA assessment. Both findings echo earlier results from Goldhaber et al. (2014a).

When estimating student achievement models, we find that underrepresented minority students (black and Hispanic), participants in the free and reduced-price lunch program, and students with reported learning disabilities score lower than their reference groups, all else equal. The magnitudes of these findings are quite similar to what has previously been found in Washington State

---

<sup>33</sup> The coefficients we discuss are not reported in the tables but are available from the authors upon request.

(e.g., Goldhaber et al., 2013) and other states (e.g., Rivkin et al., 2005). Similar to the employment models, TEPs explain a significant portion of student achievement gains in both mathematics and reading. This finding is similar to evidence from Washington State and other states in terms of the variation in teacher effectiveness that can be attributed to TEPs (Boyd et al., 2009; Goldhaber et al., 2013; Mihaly et al., 2013).

## 1. edTPA as Predictor of Workforce Entry

Table 2 reports several specifications of models predicting the likelihood of being employed in the Washington State public school teacher labor market the year after a candidate takes the edTPA assessment (see equation 1 above).<sup>34</sup> All coefficients are reported as average marginal effects; so the estimate in column 1, for example, means that teacher candidates who passed the edTPA at the Washington State cut score are 15.2 percentage points more likely to enter the public teaching workforce than are teacher candidates who failed the edTPA at the Washington State cut score, all else equal (i.e., compared with other candidates who took the same test type). The estimated marginal effect is somewhat smaller when candidates are compared with other candidates from the same TEP (column 2), and when we consider candidates who would have passed the test at the future Washington State cut score (columns 3 and 4). These relationships are not surprising given that passing the edTPA is a licensure requirement for some candidates in our sample. Not surprisingly, these relationships are even stronger when we restrict the sample only to teacher candidates who took the edTPA after it became consequential.<sup>35</sup>

Columns 5–8 consider continuous measures of edTPA performance as predictors of workforce entry. These continuous scores are standardized across all test takers, so the average marginal effect in

---

<sup>34</sup> In Tables 2-4, the “Teachers” row reports the number of teachers who identify the model presented in each column.

<sup>35</sup> Results available from authors upon request.

column 5 means that a one standard deviation increase in a candidate's edTPA score is associated with a 5.9 percentage point increase in the probability that an average teacher candidate is employed in the teacher workforce the following year. Columns 7 and 8 report specifications in which the three subscores of the edTPA are separately included in the model and show that the positive relationship between the total score and the likelihood of being in the labor market is driven largely by the assessment and instruction subscores. When we consider quintiles of edTPA scores, we find that scoring in the top quintile of the edTPA is associated with a 14 percentage point increase in the probability that a candidate will be employed in the following year, as compared with a candidate who scored in the bottom quintile.

To help visualize the relationship between edTPA scores and the probability of teaching employment, **Figure 1** plots the observed probability of employment associated with each edTPA score, along with a polynomial best-fit line.<sup>36</sup> Two patterns are worth noting. First, the relationship between edTPA scores and probability of employment is relatively steep and linear in the lower range of edTPA scores—with no discontinuity at the current passing score of 35—suggesting that, at least at the lower end of the distribution, continuous edTPA scores reflect some candidate trait or traits that are predictive of employment. Second, the relationship is much weaker in the upper range of the distribution of edTPA scores, which means that the probabilities of employment are similar for candidates within the range of relatively high edTPA scores.

Although the results in Table 2 and Figure 1 demonstrate a strong relationship between edTPA scores and the probability that a teacher candidate is employed in Washington State's K–12 public teaching workforce, it is not possible to disentangle preferences of teacher candidates and employers in interpreting these findings. As noted above, some districts may use edTPA to help them decide among

---

<sup>36</sup> The best fit line is estimated from a logit at the teacher candidate level, with the order of polynomial chosen to minimize the AIC of the regression.

teacher applicants. On the teacher candidate side, moreover, these findings may reflect the fact that more dedicated teacher candidates perform better on the assessment and are also more likely to enter the profession.

## 2. edTPA as a Screening Mechanism

Columns 1–6 of Tables 3 and 4 summarize the relationship between passing the edTPA (either at the current Washington State cut score of 35 or future cut score of 40) and teacher effectiveness in reading or mathematics, respectively. We estimate these screening models using data from the classrooms of teachers employed in the year following their edTPA administration. Given that many teacher candidates do not find teaching positions and that only a minority of teachers work in tested grades and subjects, this is a necessarily small subset of the total number of teacher candidates sitting for the edTPA. We may therefore worry that such selection biases our results. The concern is that teachers who perform poorly on the edTPA but still obtain teaching positions likely have other skills that are valued in the workplace but are not observed in our data, suggesting that the coefficients reflecting the relationship between edTPA performance and teacher effectiveness are biased downward; i.e., a lower bound on the true relationship. As discussed in the previous section, there are good reasons to believe that sample selection bias is a minimal concern, but this motivates the bounding exercise described in subsection 4D. As we describe in that section, while our findings are imprecise, the qualitative results are largely robust to sample selection issues.

The models in Tables 3 and 4 correspond to equation 2, and include lagged test scores and other student background controls (the specific independent variables used in each model specification are reported in notes below the table), but they exclude other teacher candidate variables as we are focused only in assessing the pass/fail screening value of the edTPA assessment.<sup>37</sup> However, the

---

<sup>37</sup> The argument for not including other teacher variables is that states do not utilize any other information when making an up or down decision on teacher candidates; thus no other variables should be included in the models that

coefficients in Tables 3 and 4 change very little when the models include additional teacher controls (such as WEST-B scores). We also note that results are very consistent between the primary specifications reported in Tables 3 and 4 and the more conservative specifications that either only use students with non-missing prior year test scores or non-missing twice-lagged test scores.<sup>38</sup>

Column 1 of Table 3 demonstrates that teacher candidates who pass the edTPA at the Washington State cut score are more effective in reading instruction, all else equal, than teacher candidates who fail the edTPA on their first test administration. Specifically, students assigned to a teacher who passed the edTPA score 0.252 standard deviations higher, all else equal, than students who failed the edTPA. This relationship is large and statistically significant in all specifications—i.e., comparing candidates to other candidates from the same TEP (column 2) or who teach in the same school district (column 3)—and are more modest but still statistically significant when we consider whether candidates would have passed the test at the future Washington State cut score. We interpret these results as suggesting that the edTPA has strong predictive validity in reading as a screen at these cut points. Our point estimates for the edTPA screening effect in mathematics in columns 1–6 of Table 4, on the other hand, are smaller and generally statistically insignificant. Although positive in all specifications, the screening coefficient in mathematics is statistically significant in only one specification (column 5).

The differences between the screening coefficients in reading and the corresponding coefficients in math are statistically significant, and these differences are reflected in Figures 2 and 3, which plot estimated teacher value added and edTPA test scores for all teachers in our sample. The lines plotted in these figures show local linear estimates of the relationship between teacher value added and

---

identify the value of the edTPA pass/fail signal. The effect of these teacher characteristics on student achievement is captured by the licensure test variables (as well as the other variables included in the model) through their partial correlations.

<sup>38</sup> These results are available from the authors upon request.

edTPA test scores.<sup>39</sup> While these figures do *not* control for candidate test type (and thus candidates are being compared to all other candidates regardless of test type), they illustrate that candidates who fail the edTPA at the current Washington State cutoff (35) and future Washington State cutoff (40) tend to be considerably less effective in reading (Figure 2), but less so in mathematics (Figure 3). The predicted effectiveness in reading increases sharply before the cut points, but predicted effectiveness in mathematics changes relatively little in this same range. As demonstrated by the scatter plot, we observe a smaller number of teachers with failing scores in the reading sample than in the mathematics sample and these teachers are more likely to have low value added.

### 3. The Signal Value of edTPA Performance

The value of the edTPA as a signal of teacher quality is an important policy issue. Recall that the edTPA is described as an “educative assessment,” and this is much more plausible if there is predictive validity to the assessment away from the cut point (suggesting that changes in performance by candidates or institutions are indeed predictive of teacher effectiveness). Additionally, whether inservice teachers with higher edTPA scores are more effective is an important policy question given that school systems may wish to consider an applicant’s edTPA scores in making hiring decisions.

Columns 7–12 of Tables 3 and 4 report the estimated relationships between continuous measures of candidate edTPA performance and student achievement in reading and mathematics, respectively. Columns 7–9 of Table 3 illustrate that we find little evidence that edTPA scores throughout the distribution are predictive of teacher effectiveness in reading. Specifically, the coefficient in column 7 means that a one standard deviation increase in a candidate’s edTPA score is correlated with a 0.02 standard deviation increase in student performance in the candidate’s classroom in his or her first year

---

<sup>39</sup> We estimate teacher value added using the same specification as equation 2, but omitting the edTPA scores and teacher controls. We then estimate local linear regressions of estimated teacher value added on edTPA scores using the `np` package in R (Hayfield & Racine, 2008).

teaching, but this relationship is not statistically significant. The weak relationship between continuous edTPA scores and teacher effectiveness in reading is reflected in Figure 2, as there is little increase in predicted teacher effectiveness within the range of passing scores (i.e., above a 40). We note, however, that this relationship is positive and statistically significant when we focus solely on candidates who took the edTPA after it became consequential in January 2014.<sup>40</sup>

On the other hand, columns 7–12 of Table 4 provide some evidence that edTPA scores provide a signal of future teacher effectiveness in mathematics.<sup>41</sup> Specifically, when candidates are compared across TEPs and districts (column 7), a one standard deviation increase in a candidate’s edTPA score is correlated with a 0.03 standard deviation increase in student performance in the candidate’s classroom in his or her first year teaching, and this relationship is marginally statistically significant. This is reflected in the generally positive slope of the local-linear fit line in Figure 3.

The relationship between edTPA scores and mathematics teaching effectiveness is stronger when candidates are compared to other candidates from the same TEP (column 8), but weaker when candidates are compared to other candidates who are teaching in the same school district (column 9). As discussed in Goldhaber et al. (2013), it is possible that the district fixed effects in the model in column 9 capture district-level effects that are attributed to teachers in the estimates reported in columns 7 and 8; but it is also possible that these effects remove average differences in teacher quality among different school districts that should be attributed to teachers. Given that we cannot distinguish between these possibilities, we simply conclude that the predictive validity of the edTPA as a signal of future teaching effectiveness in mathematics is stronger when comparisons are made across districts rather than within districts.

---

<sup>40</sup> Results are available from the authors upon request.

<sup>41</sup> Note, however, that the differences between the signal coefficients in math and the corresponding coefficients in reading are not statistically significant.

Finally, columns 9–12 of Table 4 consider the three edTPA subscores as joint predictors of teacher effectiveness in mathematics, and suggest that candidate performance on the Planning rubrics are driving the relationships in columns 7–9. This is an interesting finding, as the Planning subscores were less predictive of the probability of employment than were the other two subscores (see Table 2).

#### 4. Sample Selection and Bounding the Estimates

As we note above, biases due to sample selection could occur if: 1) the teaching skills measured by the edTPA are valued in the labor market; and 2) teaching skills that are not measured by the edTPA are important for hiring decisions and student test scores. These conditions would hold if principals or other district officials balance teaching skills reflected in the edTPA with other signals of teacher quality when making hiring decisions.<sup>42</sup> If this were the case, teacher candidates who perform poorly on the edTPA but are nonetheless hired likely possess teaching skills that are uncorrelated with the teacher credentials included in the effectiveness models described above, so hired teachers with failing scores on the edTPA would constitute an especially qualified sample of the failing teachers. Hence, their classroom effectiveness would not be representative of the full set of teachers failing the edTPA.

In order to obtain an estimate of the potential magnitude of sample selection bias in these estimates, we conduct a bounding exercise in the spirit of Lee (2008). The basic idea behind the bounding exercise is that, in a worst-case scenario for our analysis, the teachers hired because they demonstrate greater performance on the edTPA than the failing sample are the lowest performing teachers we observe among passing teachers. If this were the case, we can obtain an upper bound on the difference in mean performance between those who passed and those who failed the edTPA by trimming the bottom of the value added distribution for teachers who passed the test so that the selection probability matches the failing sample. Similarly, if selection operates in the other direction,

---

<sup>42</sup> Hiring officials might, for instance, derive important information about teacher applicants from interviews or references.

we could generate a lower bound by trimming the top of the passing group's value-added scores. In our application, this involves estimating value-added scores for each teacher using a model similar to equation 2, omitting the edTPA score and teacher covariates, and then trimming the top (bottom) 20% of the passing sample in mathematics and 50% of the passing sample in reading. While sample selection may plausibly generate substantial biases, the main qualitative results hold. Our results suggest the point estimates for the screening effect lie between 0.05 and 0.40 for reading and between -0.09 and 0.09 in mathematics.

This exercise produces bounds on the difference in value added under the assumption that unobserved skills are independent of performance on the test. For example, this assumption permits mean differences in unobserved skills across performance categories but imposes a homoscedasticity requirement. There is some reason to doubt that this holds exactly. For instance, the local linear regressions in Figures 2 and 3 provide some evidence of a nonlinear relationship between assessment scores and teacher value added in the tails of the edTPA distribution. In the absence of this assumption, the bounding exercise has a more limited interpretation. In particular, the bounds apply only to the difference in effectiveness among the subset of teachers who would have obtained a tested position under either outcome under the licensing restrictions in place at the time of the test. Given the relatively low threshold set by Washington State during the period we consider, only 2% of the testing sample never received a passing score. Hence, in practice, this procedure should closely approximate bounds for the screening effect among teachers who would have obtained teaching positions in the absence of the testing requirement.

The caveat that this considers only eligibility based on initial scores is, however, crucial. Our estimates do not provide evidence on the overall effects of licensure testing requirements on the teacher workforce. Instead, they provide a measure of the extent to which performance on the edTPA is

a useful measure of teacher performance. As we discuss above, two features of the testing regime inhibit this comparison. First, the test permits retakes so that about half of teachers failing their first assessment become eligible upon later examination.<sup>43</sup> Second, teachers' effort on the test is likely endogenous to the testing conditions. Changes to the licensing rules, such as limiting the number of retakes, would therefore change the underlying relationship between teacher effectiveness and performance on the test. Despite the fact that licensure tests appear correlated with productivity, the direct evidence on their efficacy as a workforce improvement tool is more mixed (Angrist and Guryan, 2008; Larsen, 2015; Wiswall, 2007).

## V. Policy Implications

In this study, we find that teachers failing the edTPA under the future Washington State passing threshold have lower value added in reading than teachers who passed the test at this cut score. We find no statistically significant difference between those who pass and those who fail in mathematics, although changes in the assessment score are predictive of teacher performance. These results generally hold when a licensure test of candidates' basic skills is included in the model, which suggests that portfolio-based assessments such as the edTPA contain information about teaching practice that is not captured by these basic-skills tests. Although our point estimates are imprecisely estimated due to the small samples employed in this study, the magnitudes of the signal estimates are roughly similar to those observed in studies of other licensure tests (Clotfelter et al., 2007; Darling-Hammond et al., 2013; Goldhaber, 2007; Goldhaber et al., 2016a,b; Hendricks, 2015).

In order to put the results in perspective, we estimate the probability that a teacher candidate failing the edTPA is a low-performing teacher (defined as being in the bottom 20% of value added) or a

---

<sup>43</sup> In fact, only 2% teacher(s) in our sample never pass the edTPA.

high-performing teacher (defined as being in the top 20% of value added).<sup>44</sup> The results of this test are in Table 5. If the passing the edTPA provided no predictive power for value added, we would expect 20% of teachers who fail the test to be in each of these categories. Not surprisingly given the null screening results in mathematics, we find that 19% of mathematics teachers who fail the edTPA are in the low-performing category. On the other hand, we find that 46% of reading teachers who fail the edTPA are in the low-performing category, far higher than the 20% we would expect by chance alone. That said, if the edTPA really were used as a one-time, high-stakes test for employment eligibility, screening these candidates who would become ineffective teachers comes at the cost of screening out some candidates who would become effective teachers. Specifically, 8% of reading teachers and 14% of math teachers who fail the edTPA are in the high-performing category (top 20% of value added); neither of these proportions is statistically different than the 20% we would expect by chance.

We can also use these results to do a back-of-the-envelope calculation of the monetary costs associated with screening ineffective teachers from the labor market. Specifically, these results suggest that the edTPA identifies one bottom quintile reading teacher for every 17 assessed candidates, while it identifies one bottom quintile mathematics teacher for every 39 candidates. Given the monetary cost to candidates of taking the edTPA, this suggests a cost of \$5,100 to identify an ineffective reading teacher and \$11,700 to identify an ineffective mathematics teacher.

## **VI. Conclusion**

Given that this is the first predictive validity study of the edTPA, and given the nuanced findings we describe above, we are hesitant to draw broad conclusions about the extent to which edTPA implementation will improve the quality of the teacher workforce. Instead, we relate our findings back

---

<sup>44</sup> We obtain similar results if we instead estimate these conditional probabilities using the simulation method suggested by Jacob and Lefgren (2008).

to the different theories of action for how the edTPA *might* improve teacher workforce quality, but we stress that even these conclusions come with important caveats that policymakers and practitioners should weigh as they interpret these results.

The first theory of action is that the edTPA can be used as a screen to prevent ineffective teacher candidates from entering the workforce. The screening results in reading—demonstrating predictive validity around the current and future Washington State cut points used for licensing decisions—generally suggest that this theory of action is promising in terms of improving overall workforce quality in reading. But as we discuss in the previous section, this screening comes at a cost, as candidates who fail the edTPA but become high-performing teachers will also be screened out of the workforce.

The screening results are far less promising for mathematics, perhaps because the edTPA is, in part, focusing on candidates' writing capacities, which are more likely to be related to a teacher's ability to teach reading than mathematics.<sup>45</sup> It is also important to recognize that the screening theory of action is predicated on teacher candidates failing the assessment. It is unclear that this screening theory of action can actually work in a setting in which candidates are able to take the test multiple times in order to pass, as the ability of the assessment to predict teacher effectiveness is likely to be low for candidates with multiple retakes (Cowan & Goldhaber, 2015).

The second theory of action is that the edTPA could improve the quality of *all* teaching candidates through the experience of the assessment or programmatic changes that are related to information TEPs receive about teacher candidate performance. This is much more likely if the edTPA scores can serve as a signal of quality teaching beyond just at the cut point required to participate in the

---

<sup>45</sup> For example, edTPA scores are more highly correlated with WEST-B writing scores ( $r=0.25$ ) than WEST-B reading or mathematics scores ( $r=0.20$ ).

labor market. In this case, it is the modest but statistically significant results in mathematics that suggest promise for this theory of action and the weaker results in reading that suggest caution.

We believe there are a number of potential next steps that are not possible to pursue with the data used in this study but that would be valuable to policymakers. One is to investigate the degree to which the different rubric scores within the edTPA might be reweighted (or modified) in order to increase the relationship between summative edTPA scores and student achievement or teacher value added. The samples in Washington State are currently insufficient for optimal weighting exercises (e.g., Goldhaber et al., 2014a), but such exercises are possible with additional years of data and/or data from other states. A second next step might be to assess how edTPA scores are related to other, broader measures of teacher performance, such as observational ratings. This is not currently possible using Washington State's administrative data; but it may be possible elsewhere. Policymakers may also be interested in whether edTPA requirements have a differential impact on minority teacher candidates, as has been found with other credential tests (e.g., Goldhaber and Hansen, 2010). We could not test this directly given that we only observe the race of test-takers who enter the teaching workforce, but the large disparities in edTPA performance by teacher race suggest that the edTPA could have implications for the diversity of the teacher workforce. Finally, given concerns about the fairness of teacher observations across classroom contexts (Steinberg and Garrett, 2016) and recent calls to place more student teachers in disadvantaged schools (Krieg et al., 2016), policymakers would benefit from evidence about whether edTPA scores vary substantially across teacher candidates in different kinds of student teaching positions.

A final caveat to these conclusions—and an essential issue for policymakers to weigh in interpreting these results—is whether the results we reference above justify the investments that candidates, states, and TEPs have made in the edTPA. While the monetary costs associated with the

edTPA are easily quantifiable (e.g., \$300 per teacher candidate), there are also less easily quantifiable time-commitment costs for both candidates and programs. We know very little regarding whether these costs might affect the pool of people who seek to become teachers. We therefore view the interpretation of these results as very much in the eye of the beholder, and we hope this early analysis spurs an evidence-based discussion about the potential promise and drawbacks of edTPA implementation.

## References

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago Public High Schools. *Journal of Labor Economics*, 25(1), 95–135.
- Angrist, J. D., & Guryan, J. (2008). Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review*, 27(5), 483–503.
- 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.
- Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, 66(2), 103–115.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593–2632.
- Chetty, R., Friedman, J. & Rockoff, J. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcome in adulthood. *American Economic Review*, 104, 2633–2679.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6), 673–682.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources*, 45(3), 655–681.
- Council for the Accreditation of Educator Preparation (2013). *CAEP Accreditation Standards*. Retrieved from <http://caepnet.org/standards/introduction>
- Cowan, J. & Goldhaber, D. (2016). National Board Certification and teacher effectiveness: Evidence from Washington State. *Journal of Research on Educational Effectiveness*. Advance online publication.
- Crowe, E. (2011). *Getting better at teacher preparation and state accountability: Strategies, innovations, and challenges under the federal Race to the Top Program*. Center for American Progress.
- Darling-Hammond, L. (2009). Teaching and the change wars: The professionalism hypothesis. In A. Hargreaves & M. Fullan (Eds.), *Change Wars* (pp. 45–68), Bloomington, IN: Solution Tree.
- Darling-Hammond, L., Newton, S. P., & Chung Wei, R. (2013). Developing and assessing beginning teacher effectiveness: The potential of performance assessments. *Educational Assessment, Evaluation and Accountability*, 25(3), 179–204.
- edTPA (2015). *Educative assessment & meaningful support: 2014 edTPA administrative report*. September 2015.

- Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources*, 42(4), 765–794.
- Goldhaber, D. (2011). Licensure: Exploring the value of this gateway to the teacher workforce. *Handbook of the Economics of Education*, 3, 315–339.
- Goldhaber, D. (2014). *What do value-added measures of teacher preparation programs tell us?* Carnegie Knowledge Brief 12.
- Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics*, 89(1), 134–150.
- Goldhaber, D., Gratz, T., & Theobald, R. (2016a). *Assessing the use of licensure tests as an indicator of teachers' science and mathematics content knowledge for teaching*. AERA 2016 Conference Paper.
- Goldhaber, D., Grout, C., & Huntington-Klein, N. (2014a). *Screen twice, cut once: Assessing the predictive validity of teacher selection tools* (CEDR Working Paper No. 2014-9). Seattle, WA: Center for Education Data and Research.
- Goldhaber, D., & Hansen, M. (2010). Race, gender, and teacher testing: How informative a tool is teacher licensure testing? *American Educational Research Journal*, 47(1), 218–251.
- Goldhaber, D., Krieg, J., & Theobald, R. (2014b). Knocking on the door to the teaching profession? Modeling the entry of prospective teachers into the workforce. *Economics of Education Review*, 42, 106–124.
- Goldhaber, D., Krieg, J., & Theobald, R. (2016b). Does the match matter? Exploring whether student teaching experiences affect teacher effectiveness and career paths. CALDER Working Paper 149.
- Goldhaber, D., Liddle, S., and Theobald, R. (2013). The gateway to the profession: Evaluating teacher preparation programs based on student achievement. *Economics of Education Review*, 34, 29–44.
- Gottlieb, J., Hutt, E., & Cohen, J. (2016). *Diffusion in a vacuum: The case of edTPA*. AEFPP 2016 Conference Paper.
- Greenberg, J., McKee, A., & Walsh, K. (2013). *Teacher prep review: A review of the nation's teacher preparation programs*. National Council on Teacher Quality.
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *The American Economic Review*, 100(2), 267–271.
- Harris, D. N., & Sass, T. R. (2014). Skills, productivity and the evaluation of teacher performance. *Economics of Education Review*, 40, 183–204.

- Hayfield, T., & Racine, J. S. (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5).
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153–161.
- Hendricks, M. D. (2015). *Public schools are hemorrhaging talented teachers. Can higher salaries function as a tourniquet?* AAFP 2015 Conference Paper.
- Hill, D., Hansen, D., & Stumbo, C. (2011). *Policy considerations for states participating in the Teacher Performance Assessment Consortium (TPAC)*. Council of Chief State School Officers, April 2011.
- Jacob, B., Rockoff, J. E., Taylor, E. S., Lindy, B., & Rosen, R. (2016). *Teacher applicant hiring and teacher performance: Evidence from DC Public Schools* (No. 22054). Cambridge, MA: National Bureau of Economic Research.
- Jette, A. (2014, July 19). 10 tips for edTPA success. *Education Week Teacher*. Retrieved from [http://www.edweek.org/tm/articles/2014/07/29/ctq\\_jette\\_edtpa.html](http://www.edweek.org/tm/articles/2014/07/29/ctq_jette_edtpa.html)
- Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, 47, 180-195.
- Krieg, J., Theobald, R., and Goldhaber, D. (2016). A foot in the door: Exploring the role of student teaching assignments in teachers' initial job placements. *Educational Evaluation and Policy Analysis* 38, 364-388.
- Larsen, B. (2015). *Occupational licensing and quality: Distributional and heterogeneous effects in the teaching profession*. Retrieved from [http://web.stanford.edu/~bjlarsen/Larsen%20\(2015\)%20Occupational%20licensing%20and%20quality.pdf](http://web.stanford.edu/~bjlarsen/Larsen%20(2015)%20Occupational%20licensing%20and%20quality.pdf)
- Mihaly, K., McCaffrey, D., Sass, T. R., & Lockwood, J. R. (2013). Where you come from or where you go? Distinguishing between school quality and the effectiveness of teacher preparation program graduates. *Education Finance and Policy*, 8(4), 459–493.
- National Research Council. (2010). *Preparing teachers: Building evidence for sound policy*. Committee on the Study of Teacher Preparation Programs in the United States. Washington, DC: National Academies Press.
- Newton, S. P. (2010). *Preservice performance assessment and teacher early career effectiveness: Preliminary findings on the Performance Assessment for California Teachers*. Stanford, CA: Stanford Center for Assessment, Learning, and Equity.
- Pecheone, R., Shear, B., Whittaker, A., & Darling-Hammond, L. (2013, November). *2013 edTPA field test: Summary report*. Stanford Center for Assessment, Learning, and Equity.
- Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005) Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, 6(1), 43–74.

Sawchuk, S. (2014, December). U.S. rules aim to heighten tracking of ed. schools' performance. *Education Week*.

Steinberg, M. & Garrett, R. (2016). Classroom composition and measured teacher performance: What do teacher observation scores really measure? *Educational Evaluation and Policy Analysis* 38, 293-317.

U.S. Department of Education. (2009). *U.S. Secretary of Education Arne Duncan says colleges of education must improve for reforms to succeed*. <http://www.ed.gov/news/press-releases/us-secretary-education-arne-duncan-says-colleges-education-must-improve-reforms-succeed>

Wiswall, M. (2007). *Licensing and occupational sorting in the market for teachers*. Retrieved from [http://www.econ.nyu.edu/user/wiswall/research/wiswall\\_teacher\\_licensing.pdf](http://www.econ.nyu.edu/user/wiswall/research/wiswall_teacher_licensing.pdf)

## Tables

**Table 1. Summary Statistics**

	Teacher Candidate Sample						Teacher Sample					
	1	2	3	4	5	6	7	8	9	10	11	12
	All	Q1	Q2	Q3	Q4	Q5	All	Q1	Q2	Q3	Q4	Q5
Total Score	46.339 (6.960)	37.272 (4.535)	44.100 (0.804)	47.045 (0.813)	50.398 (1.133)	56.204 (3.069)	46.960 (6.726)	37.537 (4.398)	44.122 (0.797)	47.035 (0.814)	50.368 (1.124)	56.240 (3.003)
Planning Subscore	15.924 (2.644)	12.962 (2.216)	15.218 (1.234)	16.138 (1.256)	17.284 (1.410)	19.102 (1.624)	16.102 (2.584)	13.088 (2.182)	15.135 (1.297)	16.095 (1.304)	17.243 (1.458)	19.113 (1.628)
Instruction Subscore	15.460 (2.416)	12.869 (1.771)	14.783 (1.203)	15.515 (1.121)	16.473 (1.389)	18.623 (1.678)	15.661 (2.334)	12.974 (1.733)	14.857 (1.133)	15.574 (1.132)	16.465 (1.395)	18.579 (1.642)
Assessment Subscore	14.922 (2.948)	11.395 (2.317)	14.055 (1.293)	15.361 (1.286)	16.605 (1.291)	18.473 (1.638)	15.165 (2.896)	11.432 (2.317)	14.094 (1.317)	15.327 (1.350)	16.627 (1.287)	18.542 (1.603)
% Passing Current WA Cut Score (35)	0.939	0.743	1.000	1.000	1.000	1.000	0.951	0.759	1.000	1.000	1.000	1.000
% Passing Future WA Cut Score (40)	0.865	0.436	1.000	1.000	1.000	1.000	0.889	0.456	1.000	1.000	1.000	1.000
WEST-B Reading	271.016 (16.139)	266.396 (17.522)	271.200 (16.709)	271.656 (15.307)	272.459 (14.961)	274.659 (14.221)	271.482 (15.997)	265.681 (17.806)	271.269 (17.118)	271.901 (14.711)	273.364 (14.623)	275.667 (13.276)
WEST-B Writing	264.340 (18.049)	257.273 (19.124)	264.049 (17.832)	265.408 (17.259)	267.332 (15.523)	269.685 (17.224)	264.910 (17.550)	256.710 (18.740)	264.166 (17.113)	266.214 (16.468)	267.905 (15.455)	270.266 (16.578)
WEST-B Math	279.548 (17.649)	274.924 (20.092)	280.000 (16.231)	280.233 (16.797)	281.775 (15.768)	282.064 (17.452)	280.093 (17.288)	274.978 (18.889)	280.261 (16.128)	280.838 (15.973)	282.431 (15.946)	282.357 (18.206)
Female							0.766	0.759	0.718	0.789	0.775	0.794
White							0.801	0.777	0.777	0.824	0.815	0.815
Asian							0.037	0.023	0.047	0.025	0.046	0.044
Black							0.015	0.007	0.019	0.007	0.023	0.017
American Indian							0.002	0.000	0.003	0.004	0.000	0.003
Pacific Islander							0.005	0.007	0.003	0.007	0.003	0.003
Hispanic							0.066	0.121	0.063	0.070	0.040	0.037
Multiracial							0.074	0.065	0.088	0.063	0.073	0.081
Entering Workforce	0.602	0.508	0.599	0.615	0.674	0.646	1.000	1.000	1.000	1.000	1.000	1.000
Read VAM sample	0.088	0.058	0.086	0.121	0.095	0.091	0.139	0.107	0.135	0.190	0.136	0.132
Math VAM sample	0.087	0.072	0.084	0.110	0.095	0.077	0.137	0.134	0.132	0.173	0.136	0.111
Teacher Candidates	2376	569	501	447	432	427	1508	307	319	284	302	296

\*NOTE: Standard deviations of continuous variables in parentheses. 413 teacher candidates and 185 teachers are missing WEST-B scores, with the distribution of missing scores relatively uniform across quintiles.

**Table 2. Models Predicting Public Teaching Employment**

	edTPA as a Screen				edTPA as a Signal			
	1	2	3	4	5	6	7	8
Passing in WA (35)	0.152*** (0.042)	0.112** (0.042)						
Future Passing Score (40)			0.137*** (0.030)	0.112*** (0.030)				
Total Score (Standardized)					0.059*** (0.011)	0.045*** (0.011)		
Assessment Subscore (Standardized)							0.034* (0.015)	0.030* (0.014)
Planning Subscore (Standardized)							-0.003 (0.014)	-0.004 (0.014)
Instruction Subscore (Standardized)							0.035* (0.014)	0.023 (0.013)
TEP effects		X		X		X		X
Teachers	2384	2384	2384	2384	2384	2384	2384	2384

NOTE: +p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001. All models controls for teacher degree level and test type effects. Average marginal effects calculated from logit model in equation 1.

**Table 3.** Value-Added Results in Reading (Stacked Model)

	edTPA as a Screen						edTPA as a Signal					
	1	2	3	4	5	6	7	8	9	10	11	12
Passing in WA (35)	0.252*** (0.073)	0.191* (0.080)	0.247*** (0.065)									
Future Passing Score (40)				0.203*** (0.058)	0.149** (0.054)	0.169** (0.058)						
Total Score (Std.)							0.022 (0.017)	0.003 (0.018)	0.006 (0.018)			
Assessment Subscore (Std.)										0.031+ (0.017)	0.017 (0.016)	0.050* (0.020)
Planning Subscore (Std.)										0.020 (0.014)	0.018 (0.014)	-0.007 (0.014)
Instruction Subscore (Std.)										-0.025+ (0.014)	-0.028+ (0.016)	-0.031+ (0.019)
TEP effects		X			X			X			X	
District effects			X			X			X			X
Teachers	208	206	175	208	206	175	208	206	175	208	206	175

NOTE: +p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001. All models control for student prior performance (either two years of lagged scores or just lagged or twice lagged score with a missing value dummy for the other) and demographics, classroom-level student demographics, teacher degree level, and grade and test type effects. All standard errors are clustered at the teacher level.

**Table 4.** Value-Added Results in Mathematics (Stacked Model)

	edTPA as a Screen						edTPA as a Signal					
	1	2	3	4	5	6	7	8	9	10	11	12
Passing in WA (35)	0.038 (0.071)	0.061 (0.068)	0.063 (0.058)									
Future Passing Score (40)				0.053 (0.045)	0.085* (0.043)	0.038 (0.037)						
Total Score (Std.)							0.029+ (0.015)	0.036* (0.016)	0.016 (0.014)			
Assessment Subscore (Std.)										-0.004 (0.026)	0.003 (0.026)	0.016 (0.019)
Planning Subscore (Std.)										0.060* (0.025)	0.071** (0.025)	0.002 (0.021)
Instruction Subscore (Std.)										-0.028 (0.021)	-0.041+ (0.021)	-0.002 (0.022)
TEP effects		X			X			X			X	
District effects			X			X			X			X
Teachers	204	203	177	204	203	177	204	203	177	204	203	177

NOTE: +p<.10; \*p<.05; \*\*p<.01; \*\*\*p<.001. All models control for student prior performance (either two years of lagged scores or just lagged or twice lagged score with a missing value dummy for the other) and demographics, classroom-level student demographics, teacher degree level, and grade and test type effects. All standard errors are clustered at the teacher level.

**Table 5.** Conditional Probabilities of Teacher Effectiveness Given edTPA Performance

	<b>Stacked Math Sample</b>		<b>Stacked Reading Sample</b>	
	Fail	Pass	Fail	Pass
Bottom Quintile	0.190 (0.088)	0.202 (0.029)	0.462** (0.110)	0.185 (0.028)
Top Quintile	0.143 (0.087)	0.202 (0.030)	0.077 (0.111)	0.205 (0.029)

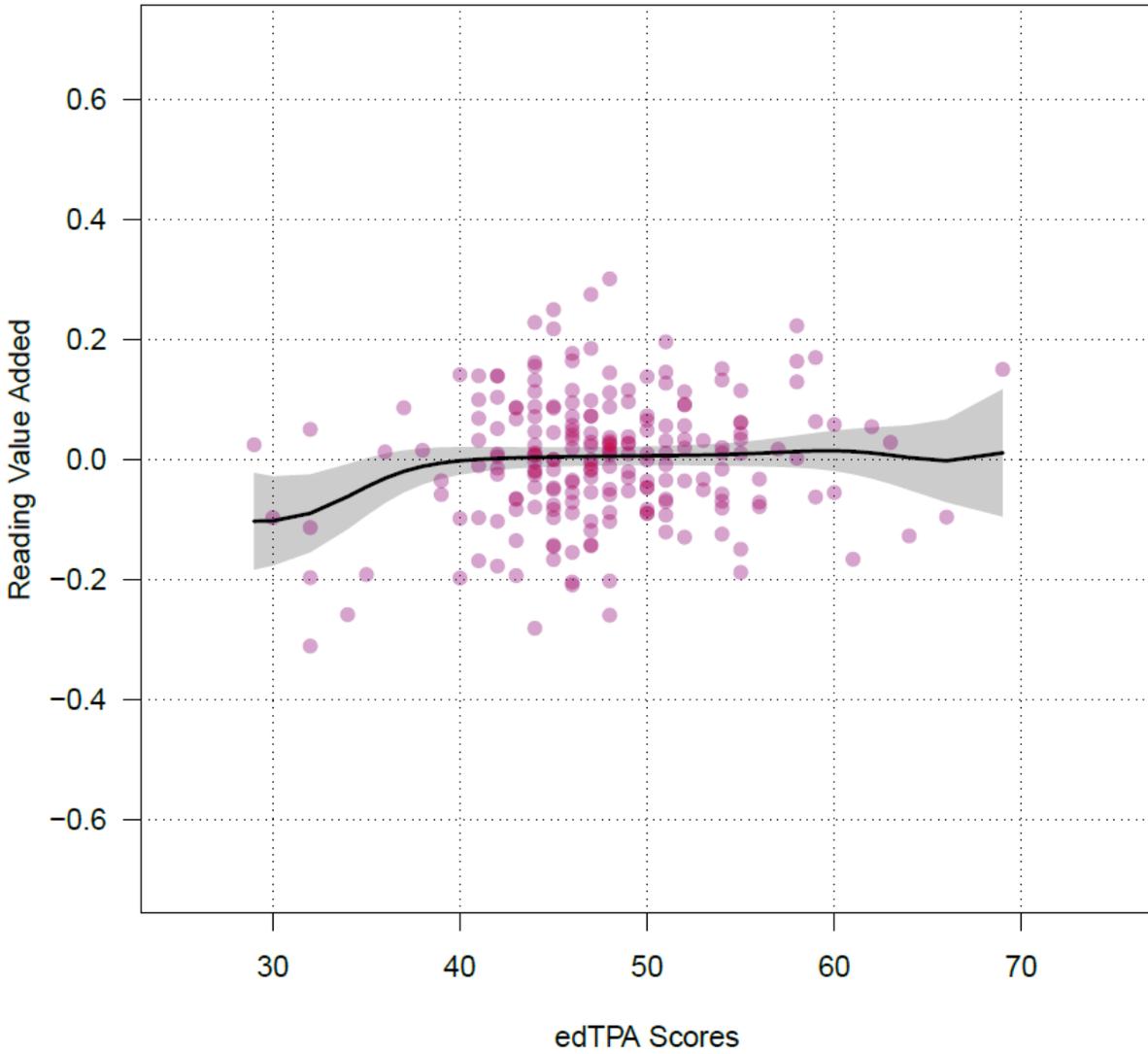
NOTE: + $p < .10$ ; \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ . Each cell gives the probability that a teacher with the indicated performance on the edTPA falls into each quintile of the value-added distribution. Standard errors in parentheses. The test of significance is against the null hypothesis that the proportion is 0.2.

# Figures

**Figure 1.** Relationship Between edTPA Scores and Probability of Public Teaching Employment



**Figure 2.** Relationship Between edTPA Scores and Reading Value Added



**Figure 3.** Relationship Between edTPA Scores and Mathematics Value Added

