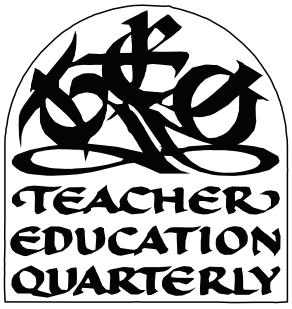

Does Standardized Writing Predict Teacher Performance?

No Relationship Found Between Standardized Assessments for State Certification

Kirsten Koetje

Abstract

The researcher analyzed data from two online cohorts of preservice teachers at a university in Washington State. Teacher education policies vary from state to state, with varying requirements for basic skills and standardized performance tests for teacher licensure. At the time of this study, Washington State required minimum scores for both basic writing skills and teacher performance to earn state residency certification (licensure). The researcher conducted a correlational analysis to determine if standardized writing scores upon admission to the program predicted performance on the national Educative Teacher Performance Assessment (edTPA). Contrary to the researcher's hypothesis, academic writing ability did not have a predictive positive relationship with edTPA performance, $r = -.004$, $p = .98$.

Introduction

Many scholars claim that of the variables over which school systems have some control, teacher quality and personal factors contribute the most significantly

Kirsten Koetje is an assistant professor in the School of Education at Seattle Pacific University. Email address:koetjek@spu.edu

to student outcomes (Hattie, 2012). Educational researchers and policy makers have long agreed that quality teaching does indeed make a difference on student outcomes as more data emerge on the subject (Aaronson et al., 2007; Angrist & Guryan, 2008; Clotfelter et al., 2010; Goldhaber et al., 2017). Although research in the last half century has substantiated many out-of-school influences, a growing release of data has attempted to correlate teacher attributes with student outcomes (Raudenbush, 2004) and tease out those things over which education policy makers may have some control. Educator preparation programs (EPPs) have faced fierce criticism throughout the decades, from political entities to teachers themselves (Cochran-Smith et al., 2018; Darling-Hammond, 2006a, 2006b; Parkes & Powell, 2015). This has created an era of increased teacher accountability and standardized testing (Cochran-Smith et al., 2018).

Educational observers often critique EPPs for having theoretical learning in coursework that seems to diverge from the experiences of the classroom. As one example, Riley (2020) explained that a group of five EPPs wanted to be strategic in implementing reading science in their elementary educator programs. These EPPs reported that although a majority of candidates (60%–80%) had opportunities to learn about reading science in coursework, a majority of those teachers (50%–60%) received no practical training on actually employing the reading science principles in the classroom (Riley, 2020). Though knowledge is necessary for teaching, it is not sufficient. Reflection in teacher education remains a cornerstone practice for growth (Darling-Hammond, 2006a, 2006b; Nagro et al., 2016; Schieble et al., 2015). Self-reflective strategies promote effective educator actions and positive classroom behaviors. Although reflection remains foundational in teacher preparation for the purpose of improving teacher performance, reflection alone is not enough (Brownell et al., 2019; Mena-Marcos et al., 2013). Because P–12 student outcomes rely in part on quality teacher performance, educational leaders and teacher trainers have sought to measure this concept of "effective performance" in standardized ways.

Teacher Testing and Performance

One of the data analysis techniques that has grown with access to available data, computing technology, and the focus on teacher quality is value added measures (VAM). VAM considers student achievement outcomes, usually on standardized test scores, which control for student characteristics like socioeconomic status (SES), race, and prior achievement, along with school context factors, such as overall school SES and achievement rates. VAM models attempt to measure and compare achievement outcomes while factoring in variables that are outside of school personnel's control. Merely comparing test means or percentage of students meeting proficiency from one school to the next has serious limitations. For instance, how can a teacher who inherits several students who are reading below grade level be assumed to get to the same finish line as a teacher who inherits several students

who are already performing above grade level? VAM models use equations that factor in a multitude of student and/or school variables and estimate a hypothetical trajectory of that student's achievement into the future (Raudenbush, 2004). Any deviation from that projected trajectory can be seen as a "value added measure" or an intervention that made a difference in that child's trajectory—for good or ill.

Proponents of teacher licensure tests claim that such testing helps to professionalize teaching and align it more with expectations of lawyers and doctors (Goldhaber et al., 2017). Teacher licensure policy also produces modest wage increases in the profession, around 3%–5% (Angrist & Guryan, 2008). Some evidence has also suggested that the requirements of basic skills scores are moving the needle on the overall average qualifications of teacher recruitment in Washington State as measured by the rising trend of Washington Educator Skills Test–Basic (WEST-B) scores, higher average SAT scores, and teacher candidates coming from more selective undergraduate institutions (Goldhaber et al., 2017). There have been documented modest positive correlations with teacher test scores and future student outcomes (Goldhaber et al., 2017). For instance, Goldhaber et al. used various VAM models for new teachers in Washington State and were able to factor in P–12 students' prior test scores. Similar studies were conducted in Chicago public schools (Aaronson et al., 2007) and in North Carolina (Clotfelter et al., 2010) on their end-of-course required tests, which reported positive correlations between teacher test scores and student achievement. Aaronson et al. (2007) reported that a 1 standard deviation increase of teacher quality as measured by licensure scores can result in a 22% increase in a student's math learning for the year. This effect was even stronger with lower to mid-performing math students and for African Americans. These studies documented low to moderate positive correlations among teachers with higher licensure scores and the achievement of their future students, while controlling for a variety of variables.

There is much more to being an effective teacher than simply "knowing" a discipline as demonstrated on standardized tests. A teacher candidate must go from being aware and reflecting on classroom practice to improving instruction through deliberate action (Mena-Marcos et al., 2013; Nagro et al., 2016; Schieble et al., 2015). Many states have mandated that preservice teachers demonstrate this reflective skill along with their actual teaching aptitude in a performance portfolio. The Educative Teacher Performance Assessment, better known as the edTPA, emerged as the national standardized teacher performance measure in recent years. The edTPA involves self-reflective practice in authentic situations by mandating classroom artifacts, such as lesson plans, video clips, and P–12 student work samples.

The edTPA is a standardized national performance assessment for those seeking teacher certification. Some states require it for certification or licensure, whereas others use it as a formative tool. The assessment purports to evaluate authentic evidence of teacher skills, rather than relying solely on content knowledge tests (Greenblatt & O'Hara, 2015). The edTPA involves the teacher candidate creating a portfolio of artifacts from student teaching internship and commentary revolv-

ing around three classroom tasks: *planning* (Task 1), *instruction* (Task 2), and *assessment* (Task 3). The artifacts that candidates submit to the external scoring agency include three to five consecutive lesson plans, lesson materials, video clips, assessments, scoring criteria or rubrics, and student work samples with teacher feedback. These artifacts come from the candidate's own student teaching internship classroom and thus provide evidence of the preservice teacher enacting their pedagogical knowledge. In addition to these artifacts, candidates provide 40–60 pages of explanation and analytical commentary on their instructional implementation. Santagata and Sandholtz (2019) pointed out that performance tests like the edTPA or the Performance Assessment of California Teachers (PACT) require preservice teachers to demonstrate their own teaching, rather than simply analyzing someone else's teaching or proving content knowledge on an exam. In other words, the edTPA requires teachers to enact their pedagogical learning and content knowledge in actual P–12 classrooms, rather than in hypothetical or simulated situations.

The content of the edTPA was developed by the Stanford Center for Assessment, Learning, and Equity. A broad network of professional educators, including more than 1,000 educators and more than 450 institutions of higher education, informed its development (American Association of Colleges of Teacher Education [AACTE], 2017). All three edTPA tasks include written commentary by the candidates, asking them to justify their original decisions and to reflect on the actual implementation of the planned lessons (Goldhaber et al., 2017; Parkes & Powell, 2015; Pecheone et al., 2016). With all the written commentary, explanation, and justification, the edTPA proves to be both an applied teaching task and a substantial writing task (Greenblatt & O'Hara, 2015; Santagata & Sandholtz, 2019). Some researchers have contended that although teacher performance is the goal, the reading, writing, and technical savvy required for uploading the final portfolio may conflate final scores alleging to measure teaching competence (Greenblatt & O'Hara, 2015; Kim, 2019; Santagata & Sandholtz, 2019). In sum, they question the edTPA's validity for measuring teaching competence.

Some evidence suggests that teachers passing performance assessments adds modest predictive value to future students' math and reading scores (Goldhaber et al., 2017; Newton, 2010). For instance, students assigned to a teacher who passed the edTPA scored .252 standard deviations higher in reading than their counterparts assigned to a teacher who failed the edTPA, when controlling for other factors (Goldhaber et al., 2017). Newton (2010) conducted a VAM study of PACT scores to see if preservice teacher performance could be linked to future student achievement on standardized tests. He controlled for student variables like SES and prior achievement to see if teachers who performed well on the PACT had a more positive effect on student learning trajectories than their lower performing teacher counterparts. Newton found that the main predictors of future student achievement, though modest, came from preservice candidates scoring high in the PACT areas of (a) assessment and (b) descriptions of student language development (part of

academic language rubrics in the edTPA). VAM studies like Newton's and Goldhaber et al.'s (2017) attempt to quantify this vague concept of "teacher quality."

Proponents of the edTPA claim that it provides an authentic and predictive task of teaching practice that requires candidates to reflect on student learning (Darling-Hammond, 2006a; Goldhaber et al., 2017; Parkes & Powell, 2015). However, many question the validity and authenticity of a task that is highly edited and curated for the purpose of passing a performance test. For instance, because teacher candidates may choose to edit out instructional missteps, choose only their best lessons, select favorite subject matter, and even choose the students to highlight in video submissions and work samples, critics challenge the claim of "authenticity" (Greenblatt & O'Hara, 2015). Promoters of edTPA policy advocate its real classroom context in P–12 classrooms as demonstrative of necessary teaching skills, rather than relying solely on pen-and-paper tests or contrived preparation program case studies or role-plays (AACTE, 2017; Darling-Hammond, 2006a; Newton, 2010; Pecheone et al., 2016).

edTPA and Academic Writing

So common is the notion among EPPs that candidates' success on the edTPA depends on strong academic writing skills that Whittaker et al. (2018) highlighted this claim in their report responding to various edTPA critiques. They countered that scorers are trained to be aware of this potential bias regarding writing quality. As of the date of their rebuttal report in January 2018, no empirical evidence had been published to support the critique that strong writing predicted strong edTPA performance (Whittaker et al., 2018). They cited an unpublished study that showed no correlation between standardized writing scores for program entrance (e.g., WEST-B Writing in Washington State) and edTPA performance.

In reviewing the literature for this article, I did not find published empirical data on the presumed positive association between strong writing and edTPA performance. This common conception of strong academic writing influencing edTPA scores continues to propagate without empirical support. This article aims to provide evidence on this assertion. Some of the literature has suggested that students with stronger academic backgrounds will have an advantage on the edTPA (Greenblatt & O'Hara, 2015; Santagata & Sandholtz, 2019). This study attempted to focus specifically on writing skill, separate from grade point average (GPA), by using WEST-B Writing scores as a standardized measure. At the time of this study, passage of the WEST-B or an approved alternate was an admission requirement for entrance into an EPP in Washington State.

Method

I conducted a correlational analysis using 2 years of cohort data, from 2017–2018 and 2018–2019, from the postbaccalaureate online teacher certification program at a liberal arts university in Washington State. During these years, Washington

law required passage of the edTPA to earn residency teacher certification. In 2021, amid the COVID-19 global pandemic, the Washington legislature permanently eliminated the edTPA as a state certification requirement.

Research Question and Hypothesis

Owing to the lack of empirical evidence on the presumed association of strong writing and teacher performance, I established the following research question:

> RQ. What is the relationship between teacher candidates' writing abilities as measured by WEST-B Writing scores and teacher candidates' performance on the edTPA?
>
> H_0. There is no relationship between standardized writing scores and performance on the edTPA.

I predicted a positive association between the predictor variable of standardized writing score as measured by WEST-B Writing and edTPA performance due to the large writing demand of the portfolio assessment. Santagata and Sandholtz (2019) referenced a modest correlation between GPA and edTPA scores. However, WEST-B Writing scores offer a more targeted measure specific to the skill of writing. The edTPA involves a significant amount of writing, and hence there is a common conception among EPP stakeholders that strong writing skills influence edTPA scores. Although the writing bias is a prevalent notion in the EPP community, I did not find any empirical evidence to support this assumption. On the contrary, unpublished evidence points to no relationship between writing skill and teacher performance (cited in Whittaker et al., 2018). Because I wanted to examine the relationship between writing ability and teacher performance assessment in a measured way, I selected WEST-B Writing scores rather than GPA.

Participants and Variables

Participant data came from graduate teacher education preservice students who completed online programs between 2017 and 2019. The university collects all the data for program purposes. For the purposes of this study, these data are considered archival. A total of 49 students had edTPA scores from these cohorts. The students represented various subject area endorsements, as shown in Table 1. Of the 49 candidates, 34 had WEST-B Writing scores due to the alternatives that were used to meet the basic skills program entrance requirements. Students may meet EPP entrance requirements with sufficient writing scores on the SAT or ACT or on various other state-approved writing exams in lieu of WEST-B Writing scores. Ten endorsement areas were represented in this data set of edTPA scores. In Washington State, special education teacher candidates must also have a general education endorsement, such as elementary or math. University program policy instructs special education interns pursuing dual endorsements to complete the edTPA in special education.

To earn certification in Washington, interns needed to pass the edTPA in one of their endorsement areas. Table 1 shows the endorsement for which the candidate completed the edTPA.

The online students took coursework from the same pool of online instructors and were assigned to a common set of online supervisors, thereby minimizing programmatic variance. However, teacher internship placements varied substantially, from urban to rural, public to private, and honors courses to self-contained special education classrooms. Moreover, each candidate had a different mentor teacher during internship. Thus internship setting and assigned mentor are both potentially confounding variables not included in this study, which limits the findings. Other researchers have suggested that classroom setting and school SES impact edTPA scores (Greenblatt & O'Hara, 2015). Although the school sites differed, the durations and expectations of the internship remained similar. Teacher candidates in these programs began a yearlong internship in late August and completed a full year's residency within the same classroom(s) through June.

The university assigns field supervisors to candidates for the duration of the year. These supervisors observe 8–10 lessons throughout the year and provide additional coaching. Once a school identifies an appropriate classroom mentor, the university confirms that the mentor selection meets statutory requirements, such as state certification and a minimum of 3 years of experience. The field supervisors, with mentor teacher and candidate input, assess the teacher candidate's performance and growth throughout the year. Most of the interns in the online program live geographically far from campus, though within state boundaries. Some supervisors live geographically closer to online teacher candidates. Some supervisors also have a stronger preference for live visits and a higher willingness and availability to travel. Thus online candidates could conduct all their classroom observations via instructional recordings and teleconference debriefs or do a blend of some

Table 1

Intern edTPA Endorsements

Endorsement	*Number of. interns*
Elementary Literacy	6
Elementary Mathematics	4
K–12 Performing Arts	1
Secondary English Language Arts	4
Secondary Social Studies	1
Secondary Math	3
Secondary Science	2
Special Education	19
Visual Arts	4
World Language	5
Total	49

traditional live visits with some instructional video recordings. The blend of live and recorded observations varied between each candidate–supervisor arrangement to reach the expectation of 8–10.

WEST-B Writing

During the 2 years of data collection, candidates had to show both basic skills proficiency and subject matter proficiency by passing certain tests to earn Washington State residency teacher certification. The basic skills tests include a battery of three tests assessing math, writing, and reading known as the WEST-B. Candidates could waive any of the three subcategories with sufficient SAT or ACT scores or state-approved equivalents from other states. Washington, as of 2019, no longer has a "cut score" for these basic tests but still requires candidates to take them. During these 2 years of data collection, the passing score for WEST-B Writing was 240. Of the 49 identified candidates from the two cohorts, 34 had WEST-B scores. Other candidates had used prior college entrance exam scores or out-of-state teacher entrance exam scores as substitutions for the WEST-B Writing requirement. This study did not attempt to determine an equivalent scale for each of the possible alternatives for meeting the state's writing requirement for entering an EPP, nor did it attempt to find *z*-scores for each test. I did not have access to all the alternative scores. This study limited the correlation to WEST-B Writing scores. If a student took the WEST-B multiple times, I input the first attempt in the analysis in an attempt to standardize the scores as much as possible for all candidates (i.e., using all first attempts).

edTPA Scores

In Washington State at the time of this study, teacher candidates had to pass the edTPA with a total score of 40 to earn certification (34 for languages). For most subject areas, there are 5 rubrics for each of 3 tasks—Task 1, planning; Task 2, instruction; and Task 3, assessment—for a total of 15 rubrics. Each rubric has a possible score of 5, making 75 a perfect score. World languages (WL) and classical languages only have 13 total rubrics and thus a lower possible score (65) and a lower passing score (34). Because the entirety of the language disciplines focus on language itself, they do not include the specific rubrics that call out teaching components of academic language. With these passing scores, candidates needed to earn an average of 2.67 on each rubric, though scorers only give whole scores, to earn certification. As with the WEST-B Writing scores, if a student took the edTPA multiple times, the first attempt with complete data was used as the data point. For instance, if a student received an error code for any of the rubrics, that edTPA score was not included in the data set.

Pearson, an external organization, scores the edTPA portfolio. Approximately 10% of portfolios are double-scored, and Pearson reports multiple reliability coefficients

that indicate a high level of internal scoring consistency, such as Cronbach's alpha of .91 (Pecheone et al., 2016). The total score on 15 rubrics determined passing or failing in Washington State. However, because those in world and classical languages have two fewer rubrics, the data analysis required an imputation technique to estimate those nonexistent values. I chose individual candidate rubric mean substitution as the preferred method. The average rubric score for each WL student's portfolio was added twice again to their total score to make it equivalent to a 15-rubric, 75-point possible score. Omitting the WL students (n = 5) from the data set was not a desirable choice because I wanted to include them as fully participating students in the online cohorts. The WL edTPA intentionally leaves out the two rubrics on academic language because the entire portfolio involves teaching communicative language skills; the interns endorsing in WL are not a random subsample. Thus I chose mean substitution for these nonexistent data using the individual students' own mean rubric scores, not group means. For example, if a WL participant earned a 39 over 13 rubrics, that averaged to a 3 for each of the 13 rubrics for that particular candidate. Dummy scores of 3 were placed into those two missing rubrics for academic language and added to the total score (39 + 6 = 45). Mean substitution as a method for estimating absent values brought the total possible maximum score to 75 over 15 rubrics. Each of the five WL students had their own personalized mean substitution added twice to their scores to retain what variability existed among students. Although the edTPA is reported in whole numbers, the researcher rounded to the nearest tenth for this imputation technique (e.g., a WL student's total score of 42 divided by 13 rubrics would result in a 3.2 mean being added for the two missing rubrics). Personalizing the mean substitution as a form of prior (current) knowledge about the students' performance and rounding to the nearest tenth were both done to preserve some of the variability within rubric scores (Tabachnick & Fidell, 2014). Mean substitution is considered a conservative estimation technique for missing values because it reduces the possible variability (Tabachnick & Fidell, 2014).

Results

I collected the relevant data on the 49 teacher candidates from two online cohorts, 2017–2018 and 2018–2019. Descriptive statistics for the preliminary data analysis are shown in Table 2. Before conducting the correlational analysis, I calculated descriptive statistics to analyze the appropriateness of the data. There were no major concerns with the data set.

Table 2
Descriptive Statistics of WEST-B and edTPA

	n	*Range*	*Min.*	*Max.*	*Mean*	*SD*	*Skew*	*Kurt*
WEST-B	*34*	*68*	*229*	*297*	*268.88*	*17.44*	*−0.34*	*.78*
edTPA	*49*	*24*	*32*	*56*	*45.99*	*5.76*	*−0.51*	*−.14*

The research question focused on the relationship between teacher candidates' writing abilities as measured by WEST-B Writing scores and teacher candidates' performance on the edTPA. There was no statistically significant relationship between those two variables as the correlation was near zero. I failed to reject the null hypothesis. In this data set, academic writing ability, as measured by the WEST-B, had no relationship with edTPA performance, $r = -.004$, $p = .98$.

Students who met the basic writing entrance requirement another way, such as with strong high school SAT or ACT scores, were not included in the correlational analysis. To determine if there was a difference in edTPA performance for students who used the WEST-B Writing for EPP entrance and those who did not, I performed an independent-samples *t*-test. The results of the Levene's test, $F(1, 47) = .008$, $p = .931$, indicate that the two groups had equivalent variance. On average, students who used an alternative to the WEST-B Writing requirement ($n = 15$, $M = 48$, $SE = 1.39$) did perform better on the edTPA compared to those who used the WEST-B ($n = 34$, $M = 45$, $SE = 0.99$) by an average of 3 points. However, this difference of 3 points, 95% CI [−6.44, 0.62], was not significant, $t(47) = -1.66$, $p = .10$. It represented a medium-sized effect ($d = 0.54$).

Discussion

Although there is a common conception of strong academic writers being favored in the edTPA, it is currently unsupported with empirical data. It seems an intuitive relationship because the edTPA requires much writing and commentary, but I did not find empirical support for this idea either in the literature review or in the correlational analysis. A few studies referred to this relationship (Greenblatt & O'Hara, 2015; Kim, 2019; Santagata & Sandholtz, 2019), but none actually provided empirical evidence. One report identified this particular claim as a critique of the edTPA (Whittaker et al., 2018), but the authors refuted this argument by pointing out that no empirical evidence has been published that supports this idea. Whittaker and colleagues referred to an unpublished study that showed similar correlational results as the present study, meaning there was no predictive value or relationship between standardized writing scores and edTPA performance. In this study, the correlation was essentially zero, indicating a lack of relationship between the two. The edTPA developers train scorers regarding this potential writing bias and inform scorers to be aware of this potential partiality. Scorers are instructed to mark portfolios according to the rubric language and not take off points for grammatical errors or weak writing. The assessment also permits bullet point explanations in the commentary sections (Whittaker et al., 2018). This study provides data that challenges the current conception held in many EPPs that the edTPA is biased toward strong writers. Perhaps academic writing skills are not the conflating variable that EPPs imagine.

While the correlational analysis did not provide evidence of a positive relationship between strong academic writing and edTPA performance, the *t*-test of the

two groups who did and did not use the WEST-B Writing as an entry requirement indicated some difference. The medium positive effect size in favor of those who had used alternative writing proficiency scores to enter into the EPP (d = 0.54) raises the possibility of something going on. Although the difference between the two groups was not significant, the effect size is substantial enough to cause curiosity. It was a smaller sample size and could be vulnerable to a Type II error (Field, 2013). It is possible that the population of students who took the WEST-B Writing differed academically from those who used alternative scores (primarily SATs or ACTs) to waive the writing requirement. For instance, students who attended 4-year universities as freshmen typically had SAT and ACT scores, whereas many community college transfer students did not. Those who used their SAT or ACT scores to waive the writing requirement were those who had sufficiently high scores. Santagata and Sandholtz (2019) did cite undergraduate GPA as having a small, positive correlation with edTPA performance. It is possible that the group who did not have to take the WEST-B Writing was somehow academically stronger as a group compared to those who took the WEST-B, thereby removing many strong writers from the analysis and limiting the correlational finding. If the populations did differ, it may be that a group of students who had stronger academic writing backgrounds were not included in the sample because they were essentially "excused" from the WEST-B Writing requirement by meeting it through alternative means.

However, of the 49 students, 34 students were accounted for in the WEST-B Writing variable. Scores ranged from 229 to 297 on the first attempt, with 300 being the maximum score possible and 240 considered passing. This range encompasses a decent amount of variability in academic writing ability among the included group as measured by this standardized test. The data captured from these cohorts for a correlation analysis did not provide evidence of academic writing as a predictor of edTPA performance. This contradicts a very common conception held within EPPs but does not contradict other empirical findings from the literature review (Greenblatt & O'Hara, 2015; Kim, 2019; Santagata & Sandholtz, 2019). The literature review produced a noticeable lack of empirical support for this common notion of writing bias in the edTPA. The lack of support for academic writing predicting edTPA performance may provide some confidence that the edTPA does not conflate writing ability with emerging teaching competencies. Rather, this lack of relationship adds to the validity of the assessment as a measure of teacher capabilities as defined by the existing edTPA rubrics.

To conclude, teacher quality remains a complex concept that cannot be easily measured. It seems common sense that a floor should be established in terms of who should be able to teach our nation's children—not just anyone should be granted access. However, are licensure tests a reasonable method for ensuring this quality? The VAM studies referenced in this article from Washington, North Carolina, and Chicago public schools do lend some support to a positive correlation between teacher test scores and future student achievement, particularly for

lower performing students. Nevertheless, licensure scores are a noisy predictor, as Angrist and Guryan (2008) argued, and can be a barrier for some to entering the profession. Educational variables tend to be noisy, as there are so many other contributing factors. Broader representation of Black teachers in the workforce is another positive contributor to student success, particularly for African American boys (Rosen, 2018). Licensure testing is clearly not the sole answer to the question of teacher quality and subsequent student outcomes. The edTPA attempts to measure teacher quality, either for summative or formative purposes. Though it comes with the very high demand of time, money, and stress for candidates to complete, it may be offering some insight into candidates' pedagogical enactment and be less linked to academic writing as many have assumed it to be.

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. https://doi.org/10.1086/508733

American Association of Colleges of Teacher Education. (2017). *edTPA state policy FAQs.* http://edtpa.aacte.org/state-policy

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. https://doi.org/10.1016/j.econedurev.2007.03.002

Brownell, M. T., Benedict, A. E., Leko, M. M., Peyton, D., Pua, D., Richards-Tutor, C., & Maheady, L. J. (2019). A continuum of pedagogies for preparing teachers to use high-leverage practices. *Remedial and Special Education*, *40*(6), 338–355. https://doi.org/10.1177/0741932518824990

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school. *Journal of Human Resources*, *45*(3), 655–681. https://doi.org/10.1353/jhr.2010.0023

Cochran-Smith, M., Carney, M. C., Keefe, E. S., Burton, S., Wen-Chia, C., Fernandez, M. B., Miller, A. F., Sanchez, J. G., & Baker, M. (2018). *Reclaiming accountability in teacher education*. Teacher's College, Columbia University.

Darling-Hammond, L. (2006a). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education*, *57*(2), 120–138. https://doi.org/10.1177/0022487105283796

Darling-Hammond, L. (2006b). *Powerful teacher education: Lessons from exemplary programs.* Jossey-Bass.

Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock 'n' roll* (4th ed.). Sage.

Goldhaber, D., Cowan, J., & Theobald, R. (2017). Evaluating prospective teachers: Testing the predictive validity of the edTPA. *Journal of Teacher Education*, *68*(4), 377–393. https://doi.org/10.1177/0022487117702582

Greenblatt, D., & O'Hara, K. (2015). Buyer beware: Lessons learned from edTPA implementation in New York State. *Teacher Education Quarterly*, *42*(2), 57–67.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge. https://doi.org/10.4324/9780203181522

Kim, N. (2019). Challenges of teaching and preparing edTPA. *Art Education*, *72*(1), 21–27. https://doi.org/10.1080/00043125.2019.1534439

Mena-Marcos, J., García-Rodríguez, M., & Tillema, H. (2013). Student teacher reflective writing: What does it reveal? *European Journal of Teacher Education*, *36*(2), 147–163. https://doi.org/10.1080/02619768.2012.713933

Nagro, S. A., deBettencourt, L. U., Rosenberg, M. S., Carran, D. T., & Weiss, M. P. (2016). The effects of guided video analysis on teacher candidates' reflective ability and instructional skills. *Teacher Education and Special Education*, *40*(1), 7–25. https://doi.org/10.1177/0888406416680469

Newton, S. (2010). *Preservice performance assessment and teacher early career effectiveness: Preliminary findings on the performance assessment for California teachers*. Stanford Center for Assessment, Learning, and Equity.

Parkes, K. A., & Powell, S. R. (2015). Is the edTPA the right choice for evaluating teacher readiness? *Arts Education Policy Review*, *116*(2), 103–113. https://doi.org/10.1080/10632913.2014.944964

Pecheone, L., Whittaker, A., & Klesch, H. (2016). *Educative assessment and meaningful support 2015: edTPA administrative report.* Stanford Center for Assessment, Learning, and Equity. https://secure.aacte.org/apps/rl/res_get.php?fid=3013&ref=rl

Raudenbush, S. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, *29*(1), 121–129. https://doi.org/10.3102/10769986029001121

Riley, B. (2020). Drawing on reading science without starting a war. *Educational Leadership*, *77*(5), 16–22.

Rosen, J. (2018, November 12). *Black students who have one Black teacher are more likely to go to college.* https://hub.jhu.edu/2018/11/12/black-students-black-teachers-college-gap/

Santagata, R., & Sandholtz, J. H. (2019). Preservice teachers' mathematics teaching competence: Comparing performance on two measures. *Journal of Teacher Education*, *70*(5), 472–484. https://doi.org/10.1177/0022487117753575

Schieble, M., Vetter, A., & Meacham, M. (2015). A discourse analytic approach to video analysis of teaching: Aligning desired identities with practice. *Journal of Teacher Education*, *66*(3), 245–260. https://doi.org/10.1177/0022487115573264

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Pearson.

Whittaker, A., Pecheone, R., & Stansbury, K. (2018). Fulfilling our educative mission: A response to edTPA critique. *Education Policy Analysis Archives*, *26*(30). https://doi.org/10.14507/epaa.26.3720