

JAEPR  
Copyright © 2024  
ISSN: 1930-9325

# JAEPR

VOLUME 9, 2024  
ISSN:1930-9325

**Journal of Applied Educational and Policy Research**

**EDITOR  
RICHARD G. LAMBERT**

**MANAGING EDITORS  
QIAO LIU  
KRISTIN VILLANUEVA**

 **THE CENTER FOR EDUCATIONAL  
MEASUREMENT AND EVALUATION**

## RESEARCH ARTICLE

---

### Using recent North Carolina Teacher Licensure Exams and Teacher Effectiveness results to Understand the Association Between edTPA Scores and First-Year Teacher Effectiveness?

Theodore Kaniuka<sup>1</sup>, Ashely Johnson<sup>1</sup>, Bradley Mills<sup>1</sup>, Kevin Bastain<sup>\*</sup>, Emily Haire<sup>1</sup>

<sup>1</sup>*Fayetteville State University*

<sup>\*</sup>*The University of North Carolina at Chapel Hill*

Evaluating pre-service teachers is a common and integral part of every Teacher Preparation Program. Recently, some states have implemented evaluation and performance criteria to assess these programs' effectiveness on criteria such as number of candidates passing pre-service licensure exams, and in-service performance. North Carolina has committed to using the Educative Teacher Performance Assessment (edTPA) as the sole pre-service licensure examination for teachers entering the profession. Concomitantly, North Carolina uses the Educator Valued-Added Assessment System (EVAAS), as a value-added measure to quantify teacher effectiveness. This study used administrative data on first-year teachers and examined the association between pre-service teacher edTPA scores and standardized EVAAS scores that were standardized by test type for first-year teachers. Utilizing linear mixed models this study found for teachers who received a passing score based on North Carolina standards, the edTPA 1) total scores predict EVAAS scores, 2) construct scores inconsistently predict EVAAS scores, and 3) individual rubric scores do not predict EVAAS scores. Outcomes are discussed in terms of practical implications and suggestions for policy.

*Keywords:* Teacher Preparation, Teacher Effectiveness, Licensure Examinations

Since 2014, a select number of teacher preparation programs in North Carolina have been piloting the Educative Teacher Performance Assessment (edTPA) (Educational Testing Service [ETS], 2024) to assess pre-service teachers during their clinical experiences. The number of schools using this exam had steadily increased in anticipation of the licensure requirements outlined in state policy. Recently, the edTPA was mandated to be used by all Teacher Preparation Programs (TPPs)

that wish to prepare students for licensure in North Carolina (LICN-003, 2019). The edTPA portfolio, developed by Stanford University, is being used in approximately 17 states as a culminating project to demonstrate a teaching candidate's ability to teach in their subject area with 26 other states using the assessment on a limited basis (edTPA, 2022). However, its use is not without controversy (Gitomer et al., 2021). Some of the original states that adopted the assessment no longer require teacher candidates to pass the exam citing reasons such as the exam being a barrier to teachers entering the profession. The data from some studies suggest that the examination fails to measure what teaching is (Dover & Schultz, 2016) and may be technically flawed (Gitomer et al., 2021). Additionally, the edTPA might be biased as it has been found that candidates of color tend to score lower than their white peers (Williams et al., 2019). In addition, claims have been made that edTPA stifles programs focusing on social justice issues (Behizadeh & Neely, 2018; Petchauer et al., 2018). Despite this, there continues to be emphasis on the need for a standardized measure of teacher effectiveness (Peck et al., 2014; Darling-Hammond, 2014, 2020).

Recently, a policy brief was released reporting a link between the edPTA and the value-added measure used to assess teacher effectiveness (Bastian & Lys, 2016). The study found that the edPTA is a valid measure upon which to base candidate licensure and performance decisions. The researchers used a sample of students from four schools in the University of North Carolina (UNC) system that piloted edTPA in 2014-2015 with a total sample size of 202 teachers who had scores on both edTPA and the North Carolina teacher effectiveness measure, the Educator Valued-Added Assessment System (EVAAS). In addition, they found that edTPA scores predicted evaluation scores of first-year teachers (Bastian & Lys, 2016). This study extends this original research to include a more comprehensive sample of teachers from the 2014-2015 school years up to 2019. In addition, this study will investigate the degree to which candidate demographics such as race, sex, and school assignment are associated with edTPA scores in general and how these variables are associated with teacher effectiveness. The following research questions have guided this study:

1. Do edTPA scores predict North Carolina first-year teacher effectiveness scores (EVAAS)?
2. Do passing edTPA scores accurately classify teachers in terms of North Carolina teacher effectiveness measures?

The question herein is that given the decision to assess pre-service teachers, is that process one that can predict eventual teaching performance in terms of student achievement? The amount of human and financial capital that is dedicated to this process has not been quantified. However, considering that 1) Teacher Preparation Programs have redesigned their programs to the content of the edTPA (American Association of Colleges for Teacher Preparation, 2024), 2) states have established minimum passing scores for teacher licensure (ETS, 2024), and 3) the cost of taking the exam, we argue that the human and financial capital costs are substantial.

## WHAT IS THE EDUCATIVE TEACHER PERFORMANCE ASSESSMENT?

The edTPA, is a performance-based, subject specific assessment used to evaluate skills and knowledge of teacher candidates. The edTPA was developed by Stanford Center for Assessment, Learning, and Equity after much evaluation of other performance-based assessments to include

### 3 TEACHER EFFECTIVENESS AND edTPA

National Board Certification, Teacher Assessment and Support Consortium, and Performance Assessment for California Teachers (edTPA, 2023). The development of the edTPA can be traced to Performance Assessment for California Teachers (PACT) and reflects the research from other portfolio-based assessments in other states (Pecheone, 2007; Wilson, et al., 2014). The performance assessment is specific to the teaching candidate's respective area of study. There are 28 versions, across all grade bands and content areas.

The edTPA portfolio consists of three constructs - planning, instruction, and assessment. Each task requires students to submit artifacts and a commentary. Task one, planning, requires teaching candidates to complete the Context for Learning, where they specify information about the classroom and school to give demographic and background information. Also, within the planning task, students design three to five learning segments to teach the essential learning target and provide a planning commentary. This commentary provides guiding prompts for students to describe their knowledge of their students and their academic skills as they prepare the designed lesson plans. For the instruction task, teacher candidates select two video clips totaling 20 minutes to demonstrate their modeling and student engagement with the learning skill. The instruction commentary also uses guiding prompts to allow the student to describe their knowledge and application in their instruction. Finally, task three is focused on assessments used by the teaching candidate. Within the assessment task, students must submit student work samples, effective feedback, and data surrounding results on their designed final assessment. Like the other tasks, an assessment commentary is used for students to describe their assessment, results, and implications for future instruction. Together these tasks are expected to effectively and efficiently determine teacher preparation and readiness (Gitomer et al., 2021, Russell & Devall, 2016).

The edTPA portfolios are scored by experienced educators who received extensive training (edTPA, 2023). Each of the 28 teaching field handbooks includes fifteen rubrics used to score the portfolio. The first five rubrics are used to evaluate the planning component, while rubrics six through ten are for instruction, and eleven through fifteen are for the assessment component. The total score from all fifteen rubrics is used to determine if the teaching candidate passes. Each individual state sets the pass/fail score for each subject. In North Carolina the passing score for world languages is 32 and for all other subjects is 38 (edTPA, 2023).

#### Assessing Pre-Service Teachers – edTPA

The purpose of assessing pre-service teachers is multifaceted and ranges from 1) determining if the pre-service teacher has mastered the skills and attributes necessary to begin a career in teaching; 2) to support teacher preparation providers information on the effectiveness of the curriculum and practices utilized in each of their programs; and more recently 3) to establish the validity of assessing the teaching candidate's internship experience and linking those evaluation scores to K-12 student learning outcomes (Henry et al., 2011). This study focused on the final question as student achievement outcomes continue to be a critical, if not the sole, measure of value-added educational policies, especially in North Carolina (NCDPI, n.d.).

Examining the effectiveness of teacher preparation programs is not new or without discussion as an effort to improve overall teacher quality here and abroad (Akiba, et al., 2007; Kim et al., 2011, Leigh and Ryan, 2008). The dialogue on assessing teacher preparation programs was highlighted by Henry et al. (2011) as a way to support program improvement. The authors argued that it is important to assess teacher preparation programs consistent with how they are associated

with K-12 student outcomes. Goldhaber et al. (2013) looked at TPPs and student outcomes in Washington State and found that the type of preparation matters because it influences student learning outcomes. The dialogue on evaluating teacher preparation programs continued, and in their 2016 paper, Brabeck and colleagues suggested three approaches to establish the effectiveness of preparation. Two of their suggested approaches are germane to this study: (a) utilize student standardized test data from PreK–12 as a measure of program effectiveness; and (b) evaluate in-service teacher performance using valid and reliable observational instruments.

Defining what is an effective teacher preparation program is difficult and not without methodological assumptions and generalizations that obviously affect the interpretation and accuracy of the work (see Fuller, 2014). Nevertheless, given the complexities and nuances facing the task (Henry et al., 2013), North Carolina has adopted the idea that teacher preparation can be assessed and that ensuring a level of mastery is warranted for teachers to receive initial teaching credentials.

The edTPA portfolio was initially piloted in North Carolina in 2010 and is currently codified into use in state policy. Passing score thresholds have been established for North Carolina licensure, with a 38 considered qualified and a 48 considered highly qualified. These levels have been criticized as barriers which may not accurately foretell if someone will be an effective teacher in the classroom (Dover & Schultz, 2016). Previous studies have shown that the fifteen rubrics adequately map onto the three main constructs of planning, instruction, and assessment; thus, providing evidence of the validity of the overall design (Bastain et al., 2016). Also, it was found that under pilot conditions, there appeared to be some predictive ability when measured by a value-added score (Bastain et al., 2016).

*Education Value Added Assessment System.* Developed in Tennessee by Dr. William L. Sanders in the 1980s and originally known as the Tennessee Value-Added Assessment System (TVAAS), the Education Value-Added Assessment System (EVAAS) is now widely used to measure the association between individual teachers and student achievement (Amrein-Beardsley, 2019). In 2011-12, the State Board of Education selected to use EVAAS to measure student growth across North Carolina. EVAAS is meant for educators to use as a reflection tool to improve their teaching practice. Additionally, this system enables one to view patterns in subgroup performance and view visuals that show the depiction of the impact of teachers, schools, and districts on academic gains (NCDPI, n.d.).

EVAAS purports to have a substantial research base to support the reliability and validity of the system as an appropriate tool to document and evaluate student and teacher performance (SAS, 2024). The platform links student test scores directly to teachers, which implies that teachers can be labeled as adding value or not adding value when compared to their colleagues based on student annual test score growth. EVAAS calculates the average growth students make in a state for a given year and uses that average as the expected growth students should make in a subsequent year. Then a teacher's students' growth scores are compared to this average and a rating is determined (NCDPI, 2022). North Carolina reports this change for schools as well as including it in the annual evaluation process for teachers (EducationNC, 2024). For example, teachers who show growth in their student test scores are adding value. In contrast, teachers who don't meet growth expectations are considered to be not adding value. It is also important to note that teachers can be labeled as "not detectibly different (NDD)" if their data shows that students grew at a rate that is not statistically different from the average (Amrein-Beardsley, 2019). EVAAS is publicized as a means for teachers to easily access valuable diagnostic data to improve their overall teaching practice, which in return should ensure that every student has the chance to succeed academically.

## 5 TEACHER EFFECTIVENESS AND edTPA

It is argued that the use of EVAAS will help teachers use more current and relevant resources in their classroom and make more sound choices when designing their lessons (Amerin-Beardsley, 2020).

### METHODOLOGY

This study used data from the North Carolina Department of Public Instruction records for edTPA performance, licensing, teacher demographics, and annual teacher evaluation outcomes. Of particular interest in this study was the predictive power of edTPA rubric scores on first-year teacher value-added measures. The analysis was divided into three phases. First, to support the regression analysis, an exploratory factor analysis was done to examine the factor structure of the edTPA. Next a confirmatory factor analysis was conducted to ascertain the feasibility of using the mean scores from the constructs as the focal variables. Second, descriptive data were generated to better understand the nature of the sample. A series of mixed model regression analyses were conducted to determine if the three constructs had predictive power on first-year teacher value-added scores. Further, this analysis was extended to examine the predictive power for three measures of student performance by grade span K-2: Text Reading and Comprehension, 3-8: End of Grade Reading/Math/Science, and 9-12 End of Course Tests and Final Exams.

The utilization of a linear mixed model regression was selected as it was hypothesized that there may be a variance component attributable to clustering teachers within schools. One consideration that was of upmost concern was the unbalanced clustering of teachers in schools. While increasing cluster size has benefits to improve the power to estimate random effects (Austin & Leckie, 2018). In fact, small cluster sizes were found not to substantially bias the estimation results (Clarke & Wheaton, 2007; Maas & Hox, 2005). A simulation study found that small sample sizes per cluster unit are not problematic (Bell et al, 2008) if the number of clusters was sufficiently large, approximately 500 or larger. Thus, having cluster sample sizes equaling 1, is not associated with estimation and bias issues and therefore linear mixed modeling was used.

### Results

The results from the analysis are presented as follows. The exploratory factor analysis communicates the establishment of, and decisions made concerning the latent constructs (*Planning, Instructions, and Assessment*). Second, the descriptive data are shared and discussed. Finally, the results from the series of regressions of the three edTPA constructs onto the teacher value-added measure and the fifteen rubrics onto standardized EVAAS scores are presented.

***Exploratory Factor Analysis.*** The exploratory factor analysis was conducted to address the following: (1) to establish that the latent constructs of *Planning, Assessment, and Instruction* were consistent with the data in this study; (2) to contrast with previous factor analyses (Bastain & Lys, 2016; SCALE, 2013) of edTPA; and (3) determine if the results herein were consistent with these previous findings. Following the recommendations of Watkins (2022) the EFA was conducted using 2308 cases. The cases were limited to first year teachers with valid edTPA and EVAAS scores. To establish the suitability of the data for analysis the determinant of a singular matrix was conducted and a value of .005 found exceeding the value of .0001 suggested by Field et al., (2012). The Barlett (1950) test of sphericity showed that the matrix is not random with Chi square = 17589.13, with 105 degrees of freedom,  $p < 0.001$ . A Kaiser-Meyer-Olkin (1974)

evaluation was done revealing a KMO = 0.936 far exceeding recommended values. Given these values the correlation matrix was deemed appropriate for EFA analysis (Hair et al., 2019). A Promax rotation was used to allow for correlation among the factors, and a correlation lower limit of 0.3 was used to establish the loadings. The results, as presented in Table 1, were found to be inconsistent with previous analyses (Bastain & Lys, 2016; SCALE, 2013) as we found that all five *Planning* rubrics loaded onto a single factor, five *Assessment* rubrics (plus one from *Instruction*) loading onto a factor, and finally the remaining four rubrics loaded onto the *Instruction* factor.

TABLE 1  
Factor Analysis of edTPA on Regression Sample of First Year Teachers (n= 2308)

edTPA Task	edTPA Rubric	Factor 1	Factor 2	Factor 3
Planning	Planning for Content Understanding	<b>0.705</b>	-0.039	0.114
	Planning to Support Varied Student Learning Needs	<b>0.689</b>	0.033	-0.001
	Using Knowledge of Students to Inform Teaching	<b>0.603</b>	0.162	-0.001
	Identifying and Supporting Language Demands	<b>0.65</b>	0.089	0.025
	Planning Assessment to Monitor and Support Student Learning	<b>0.761</b>	0.011	0.032
Instruction	Learning Environment	0.057	-0.063	<b>0.796</b>
	Engaging Students in Learning	0.074	0.017	<b>0.835</b>
	Deepening Student Learning	0.018	0.049	<b>0.782</b>
	Subject-Specific Pedagogy	-0.017	0.118	<b>0.702</b>
Assessment	Analyzing Teacher Effectiveness	0.201	<b>0.343</b>	0.213
	Analysis of Student Learning	0.157	<b>0.608</b>	0.069
	Providing Feedback to Guide Further Learning	-0.072	<b>0.738</b>	-0.009
	Student Use of Feedback	-0.053	<b>0.815</b>	0.002
	Analyzing Students' Language Use	0.183	<b>0.525</b>	0.06
	Using Assessment to Inform Instruction	0.183	<b>0.655</b>	-0.001

According to the edTPA literature, the focus of this rubric is to have candidates develop a reflective narrative that includes positive & negative aspects of each of unit components (rationale, objectives, assessment, lesson plans, resources & references). In essence this rubric is a form of assessment; the analysis of instruction and the related materials including a reflection on these

## 7 TEACHER EFFECTIVENESS AND edTPA

materials and quantifying if they were appropriate for the learning tasks. Although this rubric had the weakest loading onto assessment it is not beyond reason that it should. Therefore, we did move this from the instruction construct to assessment and believe that this in fact is the most appropriate outcome. The retention of three factors was supported in the analysis as the eigenvalues (proportion of variance) for the three factors was 6.065 (0.199), 1.325 (0.16), and 1.115 (0.107) respectively. Regarding model fit, the Comparative Fit Index = 0.97, Tucker-Lewis = 0.964, Root mean square error of approximation = 0.043, and the standardized root mean square residual = 0.028; in total, suggesting an adequate model fit.

*Confirmatory Factor Analysis.* The purpose of the CFA was to examine the overall linkage between the measured elements (rubrics) and the three proposed latent constructs. The CFA was run using a maximum likelihood estimation with robust standard errors as ML estimation has been shown to be robust to violations of normality (Roos & Bauldry, 2022). The factor loadings are reported below in Table 2. The fit statistics reported  $\chi^2 = 529.541$ ,  $p < 0.001$ , Comparative Fit Index = 0.97, Tucker-Lewis Index = 0.964, Root mean squared error of approximation = 0.043,  $p = 1.0$ , and a standardized root mean square residual = 0.028 with a coefficient of determination of 0.983. This model was compared to the model with rubric 10 being included in the Planning construct, and this model had a slightly numerically smaller BIC of 67722.59 to 67923.59. Given the above, the model that is presented in Table 2 was deemed as an acceptable representation of the constructs and indicator loadings.

*Sample Descriptive Data.* The sample was taken from teachers who were employed in a North Carolina public school between 2016-2020, meeting the following criteria: 1) have an edTPA score from a portfolio with 15 rubrics; 2) have an EVASS (Education Valued Added Assessment System Score); and 3) who were first year teachers when the performance data was generated. This final number is smaller than the total population of teachers in North Carolina. North Carolina uses a variety of pathways for teacher certification, including a residency licensure pathway in which individuals with bachelor's degrees may get hired into teaching positions while completing coursework to obtain a teaching license. These individuals were removed from the analyses as they could have 2-3 years of teaching experience before completing the edTPA (NCDPI, n.d.). Additionally, it was decided that selecting first-year teachers reduced the possibility of other influences or confounders on EVAAS scores, such as experience or long-term exposure to district and school differences.

Concern has been expressed that the edTPA, like many standardized assessments, may be subject to bias. That edTPA scores may be either inflated or deflated and these effects might be a result of candidate characteristics. The preceding Table 2 clearly shows that in the case of comparing the ethnic/racial groups, there appears to be a very consistent systematic difference in scores derived from edTPA. A Bonferroni correction was used to address the potential family wise error of conducting multiple t-tests such that the adjusted  $p = 0.002$ . The Welch t-test revealed that when compared to White teachers, in all but one instance (Learning Environment), Black teachers had significantly different outcomes as compared to their White peers. Overwhelmingly Black teachers have lower scores, with an approximate total score difference of 3.4. It is important to note that no other ethnic/racial group had a significant difference when compared to White teachers. These outcomes appear to support, in part, the arguments presented by Williams et al., (2019) and contrast with the results of Goldhaber et al. (2017), who found the scores of Hispanic candidates were significantly lower while score differences were demonstrated in only a few instances for Black candidates.



Table 2  
Factor Loading for Confirmatory Factor Analysis for a Three Latent Construct Solution

Factor	Indicator	Estimate	S.E.	z-value	95% Confidence Interval	
					Lower	Upper
Planning	Rubric 1	0.372	0.013	29.161	0.347	0.397
	Rubric 2	0.486	0.015	32.915	0.457	0.515
	Rubric 3	0.401	0.013	31.179	0.376	0.426
	Rubric 4	0.422	0.012	36.11	0.399	0.445
	Rubric 5	0.469	0.014	34.479	0.442	0.496
Instruction	Rubric 6	0.224	0.011	20.125	0.202	0.246
	Rubric 7	0.453	0.012	38.646	0.43	0.476
	Rubric 8	0.453	0.012	37.574	0.429	0.476
	Rubric 9	0.484	0.014	35.796	0.457	0.51
Assessment	Rubric 10	0.378	0.013	30.03	0.353	0.403
	Rubric 11	0.518	0.013	38.747	0.492	0.544
	Rubric 12	0.507	0.016	31.894	0.476	0.538
	Rubric 13	0.509	0.013	38.657	0.483	0.535
	Rubric 14	0.446	0.012	35.949	0.422	0.47
	Rubric 15	0.519	0.014	37.822	0.492	0.546

*Note:* All p values were <0.001

*Regression Descriptive Statistics.* Teacher covariates included sex and race/ethnicity and all teachers in the sample were first-year educators. School-level data proposed

## 9 TEACHER EFFECTIVENESS AND edTPA

to be used as covariates in the regression included the school mean National School Lunch participation (EDS) and percent of students of color. Table 3 displays the sample statistics for the covariates. In previous research, the percentage of students of color in a school did not account for much of the variance in value-added scores, while the percentage of economically disadvantaged students was found to have a small association with such scores (Chen et al., 2010).

Table 3

### Regression Analysis Sample Statistics for First Year Teachers (n = 2801) and School Level Covariates

Variable	Percents (Mean)	Standard Deviation
Ethnicity/Race		
Asian	1.46	0.12
Black	10.39	0.3
Hispanic	4.71	0.21
Indigenous American	0.86	0.09
Other	0.18	0.04
White	82.4	0.38
Sex		
Females	81.51	0.38
School Covariates		
Mean School Percent Minority	52.99	24.78
edTPA Constructs		
Planning	(3.04)	.048
Learning	(2.94)	0.49
Assessment	(2.94)	0.56
Dependent (unstandardized)		
EVAAS	-0.03	1.02

The sample demographics show a high percentage of teachers who are ethnically White and predominately female. Regarding the three constructs, the mean scores in Table 4 are quite similar with little variance in the mean scores for each of the three constructs. It was found that

there is little difference in these scores both within the individual test area and between them. In fact, an effect size analysis to assess the between differences yielded the largest difference between the means for Assessment across K-2 and 9-12 of 0.3, a small effect.

Table 4

Mean edTPA Construct Mean Scores and (Standard Deviations) Across Grade Spans: Test Areas

Construct	K-2: Text Reading and Comprehension	3-8: End of Grade	9-12: End of Course and Final Exam
Planning	3.05 (0.46)	3.05 (0.48)	2.98 (0.47)
Learning	2.99 (0.38)	3.03 (0.47)	2.93 (0.49)
Assessment	2.96 (0.51)	2.92 (0.56)	2.80 (0.54)

*Regressions.* The regression analysis was run to determine if there were any associations between edTPA performance and teacher EVAAS scores. First, the three constructs were regressed onto the standardized value-added measure. When the EVAAS measure was standardized, it was done so by the grade and test subject area for each teacher to reflect the variance in these scores across these assessments. The first regression included only teachers that received a passing score on the edTPA in the analysis, and the second grouped the teachers by the test areas/grade spans used by North Carolina in its accountability system. The utilization of interaction terms was considered however, there were collinearity violations which could not be addressed and the base or comparison group estimates were not provided. Therefore, the group analysis was seen as the most parsimonious approach to providing the results. Finally, the third phase regressed the fifteen individual rubrics onto the value-added measure by state test area using the same criteria for the selection of teachers.

The first step in the linear mixed model regression analysis was to regress the total score first-year teachers received on the edTPA onto the standardized EVAAS scores. The dependent variable was standardized to better interpret the estimated coefficients of the edTPA constructs, as the EVAAS scale was felt to have limited value on its original scale. Total scores are simply the sum of the individual rubric scores, in this case the total number of rubrics was 15. We regressed passing scores of 38 and above. School level covariates were originally considered, however a pairwise correlation with the EVAAS score showed little to no correlation and when included in the regression the variance estimates were zero, therefore it was decided not to include these second level covariates in the final models. The three models had Wald Chi-squares values that were highly significant. It was also seen that the between school variance was high as the interclass correlations for all three test groups exceeded 0.1 and were significant. The results in Table 5 show that the edTPA total score does have some predictive association with the EVAAS scores.

Table 5

## 11 TEACHER EFFECTIVENESS AND edTPA

## Regression Estimates and (SE) of edTPA Total Rubric Score on Standardized EVAAS scores by Test Type and Passing Rubric Scores 38 and Greater

Variable		Text Reading and Comprehension	End of Grade	End of Course and Final Exam
	Total Rubric Score	0.018* (0.006)	0.019** (0.005)	0.015* (0.006)
Ethnicity	Asian	-0.31 (0.43)	0.312 (0.206)	-0.383 (0.244)
	Black	-0.245 (0.148)	0.239* (0.093)	-0.234* (0.109)
	Hispanic	0.013 (0.143)	0.199 (0.135)	0.103 (0.16)
	Indigenous American	-0.442(0.441)	0.138 (0.338)	-0.538(0.048)
	Other	-1.163*** (0.042)	0.827 (0.124)	-0.538 (0.043)
Gender	Male	-0.271(0.253)	-0.11 (0.109)	-0.166(0.048)
Intercept		-0.286 (0.608)	-0.719 (0.341)	-0.259 (0.067)
Wald $\chi^2$ (p value)		821.47 (<0.001)	68.85 (<0.001)	192.35 (<0.001)
Observations (Groups)		688 (460)	1062 (659)	1056 (466)

## Random Effects Parameters

	School	Estimate (Robust SE)		
Variance	Intercept	0.132 (0.081)	0.205 (0.052)	0.261 (0.048)
	Residual	0.849 (0.092)	0.797 (0.049)	0.747 (0.05)
Interclass Correlation		0.138 (0.08)	0.204 (0.048)	0.259 (0.042)

*Note* †p <.1, \*p <.05, \*\* p <.01, \*\*\*p<.001, White teachers are the base group for ethnicity, edTPA score of 38 is North Carolina minimum passing, all variance estimates for the random effects were significant.

For example, a one-unit change in the edTPA scores is predicted to increase a teacher's EVAAS scores by 0.015 standard deviations for teachers whose students take either high school End-of-

Course or a final exam. Similar results were found for the other two assessment groups. While it was seen in Table 2 that ethnicity for Black teachers as compared to white did show significant differences in rubric scores, in this model we found that Black teachers were predicted to have lower EVAAS scores when compared to their white peers. Of note there also was a significant predictor for teachers who identified as other when compared to white teachers for the text Reading and Comprehension group.

The second regression was run to determine whether the three edTPA constructs of *Planning, Instruction, and Assessment* had predictive power onto the value-added score (EVAAS) that North Carolina uses to assess teacher effectiveness. As supported by our two factor analyses, we believed that examining how the constructs, which measure different teacher behaviors and skills, may be associated differently with the EVAAS scores. We again included the same teacher level covariates and no covariates for the school level. Consistent with the first regression it was found that all three models were significant and showed large between school variances. There were subtle differences in the variance estimates but these differences were minor and expected. As seen in Table 6, in two instances the constructs were found to significantly predict EVAAS scores. For teachers being assessed by the End-of-Grade test (grades 3-8) a one-unit increase in overall mean score for the Instruction construct was predicted to increase EVAAS scores by 0.188 standard deviations. For teachers being assessed by End-of-Grade and Final exams (high school), it was found that a one-unit increase in the mean scores for the Assessment construct increased EVAAS scores by 0.224 standard deviations.

The third and final regression analysis was to examine if and how the fifteen individual rubric scores predict teacher EVAAS scores. A model was run that used rubrics from teachers who had received the passing score of 38 or above. The first step was to transform the original rubric scores to a binary variable with 1 indicating the teacher passed (scores 2.5 or higher) and 0 not. While there is no required passing score for each rubric, a total score of 38 is required for all portfolios (edTPA, 2023). As there are 15 rubrics, we decided that, on average, a score of approximately 2.5 is a reasonable cut score to classify as meeting a minimum or passing score. Results reported in Table 7 indicate only one instance where the estimated coefficient was significant: *Analyzing Teaching Effectiveness*. This association seems quite appropriate with the rubric asking, “How does the candidate use evidence to evaluate and change teaching practice to meet students’ varied learning needs?” Additionally, the passing criterion for this rubric examines if the candidate proposes changes to teaching that are aligned with or meet the collective learning needs of students.

Table 6

Regression Estimates and (SE) of edTPA Planning, Instruction, and Assessment Constructs on Standardized EVAAS scores by Test Type and Passing Rubric Scores Greater than 38

Variable		Text Reading and Comprehension	End of Grade	End of Course and Final Exam
	Planning	0.16 (0.188)	0.085 (0.082)	-0.055 (0.1)
	Instruction	-0.033 (0.121)	0.188* (0.088)	0.038 (0.089)
	Assessment	0.11 (0.097)	0.059 (0.085)	0.241** (0.079)
Ethnicity	Asian	-0.341 (0.42)	0.303 (0.205)	-0.354 (0.267)
	Black	-0.235 (0.15)	0.24* (0.095)	-0.225* (0.108)
	Hispanic	0.011 (0.146)	0.22 (0.14)	0.113 (0.185)
	Indigenous American	-0.461(0.437)	0.128 (0.326)	-0.113 (0.281)
	Other	-1.189*** (0.061)	0.865*** (0.124)	-0.548*** (0.08)
Gender	Male	-0.268(0.273)	-0.11 (0.109)	-0.168* (0.068)
	Intercept	-0.216 (0.67)	-0.719 (0.145)	-0.239 (0.306)
	Wald $\chi^2$ (p value)	811.56 (<0.001)	64.1 (<0.001)	194.4 (<0.001)
	Observations (Groups)	676 (456)	1049 (659)	1042 (464)

Table 6 cont.

Regression Estimates and (SE) of edTPA Planning, Instruction, and Assessment Constructs on Standardized EVAAS scores by Test Type and Passing Rubric Scores Greater than 38

Random Effects Parameters				
School		Estimate (Robust SE)		
Variance	Intercept	0.131 (0.083)	0.203 (0.052)	0.261 (0.048)
	Residual	0.861 (0.095)	0.798 (0.05)	0.744 (0.05)
Interclass Correlation		0.132 (0.082)	0.203 (0.048)	0.26 (0.043)

Note †p <.1, \*p <.05, \*\* p <.01, \*\*\*p<.001, White teachers are the base group for ethnicity, edTPA score of 38 is North Carolina minimum passing

## Discussion

This study examined the relationship between the edTPA and teacher effectiveness by establishing how well it predicts first-year teacher performance as measured by a value-added metric- EVAAS. Evidence from this effort appears to suggest that there are instances when candidate performance is related to race. Black teachers were seen to score significantly lower than their White peers when evaluated by the pre-service edTPA. These lower scores were not random, as there was a difference in every measure presented herein. Although the mean score of Black teachers was above the minimum score required by North Carolina, it was some 3.4 points lower than the scores white teachers achieved. While a complete understanding of these results was beyond the scope of this current study, a deeper examination of this outcome is required to determine if these differences are attributable to the assessment or other factors.

As to the question of whether the edTPA could predict teacher effectiveness as defined in North Carolina, it was found that, from a practical perspective, it does under certain restrictions. The edTPA total score, was found to have significant predictive estimates for all three test types, such that higher total scores did predict higher EVAAS scores. The second regression analysis using the three constructs of *Planning*, *Learning*, and *Assessment*, did not consistently predict EVAAS scores. In fact, only 2 of the 9 possible estimates resulted in significant estimates. Final estimates for the 15 individual rubrics revealed that only one rubric had a significant association with EVAAS scores. These results do have ramifications for both teacher preparation programs and school-based leadership. Teacher Preparation Programs need a mechanism to accurately and reliably assess the effectiveness of their programs in terms of long-term outcomes. It appears that for North Carolina Teacher Preparation Programs, the edTPA may not be a viable indicator to assess program curricula and the effectiveness of their graduates. Alternative measures may be

## 15 TEACHER EFFECTIVENESS AND edTPA

warranted if colleges and universities, along with other teacher preparation programs, are compelled to demonstrate that their approaches benefit novice teachers and the students they teach. However, assessing effective teaching and agreeing upon what defines an effective teacher is debatable. The content of the edTPA and what it purportedly measures may be the essential qualities and characteristics of what an effective professional educator should and must possess. To use the results herein to call into question the constructs or the measures used to assess the presence of these is beyond the scope of this study. However, our findings clearly demonstrated that alignment between how effective teachers are conceptualized and their EVAAS scores is absent.

Framing this study into a practical perspective, we examine the edTPA as a screener. In a manner similar to what Goldhaber et al. (2017) conducted, we also estimate the conditional probability of a teacher to have an EVAAS score in the lower quintile or 20% based on whether they failed or passed the edTPA and similarly the conditional probability for the upper 20% range. If the edTPA was not an effective screener or signal (having no predictive power), we expect that 20% of the teachers who failed the edTPA would be distributed equally across all five levels. We used a proportion test incorporating clustering and intraclass correlation to analyze the data. As seen in Table 8, we found that only for the End-of-Grade test were the observed proportions different from what we expected if no predictive power was present. Therefore, if the edTPA was used as a screening mechanism, about 13.4% of those teachers who were classified as having failed to meet the minimum cut score were found to be among the highest-performing with respect to the End-of-Grade test types. Additionally, it was found that 26.6% of the teachers that missed the minimum cut score did perform in the lowest quintile, slightly above what we would have expected.

Table 8

Quintile Proportions of Teacher Effectiveness for Failing and Passing the edTPA Minimum and Highly Qualified by Test Areas

Quintile	Reading and text Comprehension		End-of-Grade		End-of-Course and Final Exam	
	Fail	Pass	Fail	Pass	Fail	Pass
Bottom	0.266 (0.051)	0.199 (0.017)	0.268* (0.037)	0.209 (0.014)	0.237 (0.031)	0.209 (0.016)
Top	0.189 (0.045)	0.179 (0.016)	0.134* (0.029)	0.206 (0.014)	0.169 (0.027)	0.206 (0.016)
HQ only						
Bottom		0.183 (0.019)		0.207 (0.017)		0.189 (0.019)
Top		0.205 (0.02)		0.221 (0.017)		0.237 <sup>†</sup> (0.021)

*Note:* Each cell shows the proportion of teachers failing/passing the edTPA in each bottom and top quintiles for the EVAAS. HQ score has no failing edTPA total scores. Standard errors in parentheses. Test of significance is against the null hypothesis that the proportion = 0.2. <sup>†</sup> p < 0.1, \* p < 0.5. \*\* p < 0.01, \*\*\* p < 0.001



We extended the analysis to include those edTPA participants who were granted highly qualified status (HQ only in Table 8) as a result of their edTPA results. In North Carolina, if a teacher scores 48 or above, they are considered highly qualified and as seen in Table 8, the actual portions are not statistically different from the hypothesized value. Interestingly, this analysis shows that being rated as HQ does not necessarily translate into those teachers performing differently than expected. Reflecting on this analysis, the edTPA was found to have very limited predictive power and when it did, some teachers who failed to meet the minimum North Carolina score were among some of the highest-performing teachers.

Previous studies on professional examinations for the field of education have shown some associations between licensure tests and student outcomes. (Darling-Hammond et al., 2013; Goldhaber, 2007; Goldhaber et al., 2017). What is different between this study and others is multifold. The sample size in the Bastain and Lys study (2016), which is the most closely aligned with this one, was significantly smaller and the breadth of Teacher Preparation Programs used in the report was much more limited. By utilizing a much more comprehensive sample, the findings of this study illustrate the limited predictive power of the edTPA as a signaling mechanism for teacher effectiveness. It also has implications for Teacher Preparation Programs considering this pre-service measure as a possible benchmark for their performance outcomes.

## Policy Implications

Assessing preservice teachers has been and will continue to be a central component of teacher preparation; however, finding a way to do so has proven to be quite complex (see Regan et al., 2018). Firstly, what defines a good teacher and hence a good school, and who can tell? Secondly, what does it mean to have an effective teacher preparation program? Since No Child Left Behind (NCLB, 2004), the definition of good teaching has been narrowed down to positive student learning outcomes as measured by some type of assessment. It stands to reason that in this age of assessment-driven education that this should apply to Teacher Preparation Programs. Assessing teacher quality has relied heavily on principal evaluations with the research on this suggesting that principals have demonstrated to be as effective as some valued added measures (Harris & Sass, 2007). In contrast, the dialogue on what preparation programs can do to effectively prepare and support new teachers continues to be a significant focus of research (Cochran-Smith et al., 2015; Goldhaber & Ronfelt, 2020; Korthagen, 2004; Scheeler, 2008). The rapid and seemingly enthusiastic adoption of the edTPA by states and policymakers may reflect the urgency for accountability in teacher education programs. There is a strong public demand for highly effective teachers, and accountability programs appear to be solidly entrenched in education. The recent decline in states using the edTPA as a required licensing examination may reflect several concerns: the cost, passing rates as barriers to the profession (Swanson, 2021), pedagogical and ethical bias (Gilbert & Kuo, 2019), and possibly the usefulness of the assessment (Chung & Zou, 2022; Dover & Schultz, 2016; Regan, et al., 2017). The question of racial bias is ever present and, as shown herein, remains a valid concern as a significant obstacle in diversifying the teaching profession. Nearly ten years after the rapid and arguably unprecedented adoption of the edTPA, evidence appears to be growing that the assessment is underperforming.

Accountability and evaluation are ingrained in education, and utilizing assessments to determine the effectiveness of Teacher Preparation Programs is an essential aspect of this effort. Educators have readily supported using portfolio-based assessments as they enhance the collection of a wide range of information, are highly personalized, and can be adapted to various goals. The

## 17 TEACHER EFFECTIVENESS AND edTPA

main issue policymakers and professionals face is that of usefulness and pragmatics. As demonstrated in this study, the edTPA fails to predict first-year teacher valued-added performance, appears biased, and does not effectively screen or signal who will be a high or low-performing first-year teacher. So where do we go from here? Carnine (2000) lamented the lack of scientific rigor when educational reforms were adopted, and some 20 years later, the edTPA may be a contemporary example of this. We do not suggest abandoning the edTPA is in the best interest of educational reform, but we are enhancing our understanding of what value it can add and role it can fulfill. It is important to remember that all assessments have limitations, and policymakers and state legislators need to heed this as they develop and implement accountability measures.

### Suggestions for Future Research

The findings for differences in the edTPA rubric scores across race was consistent with previous work i.e., race-based differences (see Goldhaber et al., 2016). It is suggested that this outcome be explored by using student level data from Teacher Preparation Programs and investigate how student demographics such as race, sex, high school GPA, college GPA, major and other preparation related data explain the edTPA outcomes. Using these data as covariates to explain the edTPA scores might point to the factors which contribute to any potential sources of bias. This study limited the range of edTPA rubrics to the first fifteen, which excluded specialty areas. If the data support this we suggest examining the relationships between such areas a special education, mathematics, and science to determine if the relationships found herein are present with these areas. Finally, an examination of how the edTPA scores and teacher annual evaluation results are related may yield data to determine how teachers are evaluated both before and after entering the profession. The edTPA is widely used and a gatekeeper for new teachers entering the profession. A deeper understanding of how well this exam provides reliable and relevant data to predict the future success of novice teachers is warranted. Not only would this serve to establish the usefulness of the edTPA, but might lead to better teacher effectiveness measures.

## REFERENCES

- Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational researcher*, 36(7), 369-387.
- American Association of Colleges for Teacher Education (2024). edTPA. Retrieved from: <https://aacte.org/faq/edtpa/>
- Amrein-Beardsley (2019). The Education Value-Added Assessment System (EVAAS on Trial: A Precedent-Setting Lawsuit With Implications for Policy and Practice. *Journal of Education Policy*. <https://files.eric.ed.gov/fulltext/EJ1234497.pdf>
- Amrein-Beardsley, A., & Geiger, T. (2020). Methodological Concerns About the Education Value-Added Assessment System (EVAAS): Validity, Reliability, and Bias. *SAGE Open*, 10(2). <https://doi.org/10.1177/2158244020922224>.
- Austin, P. C., & Leckie, G. (2018). The effect of number of clusters and cluster size on statistical power and Type I error rates when testing random effects variance components in multilevel linear and logistic regression models. *Journal of statistical computation and simulation*, 88(16), 3151-3163.
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British journal of psychology*.
- Bastian, K. C., & Lys, D. (2016). Initial findings from edTPA implementation in North Carolina. *Education Policy Initiative at Carolina*. Retrieved from <https://publicpolicy.unc.edu/files/2017/03/edTPABrief-Revised.pdf>.
- Behizadeh, N., & Neely, A. (2018). Testing injustice: Examining the consequential validity of edTPA. *Equity & Excellence in Education*, 51(3-4), 242-264.
- Bell, B. S., Kanar, A. M., & Kozlowski, S. W. (2008). Current issues and future directions in simulation-based training in North America. *The International Journal of Human Resource Management*, 19(8), 1416-1434.
- Brabeck, M. M., Dwyer, C. A., Geisinger, K. F., Marx, R. W., Noell, G. H., Pianta, R. C., ... & Worrell, F. C. (2016). Assessing the assessments of teacher preparation. *Theory into practice*, 55(2), 160-167.
- Brown, C. S. (2017). Aligning a Performance-based Observation Rubric to Support a
- Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—simulation and Computation*, 39(4), 860-864.
- Clarke, P., & Wheaton, B. (2007). Addressing data sparseness in contextual population research: Using cluster analysis to create synthetic neighborhoods. *Sociological methods & research*, 35(3), 311-351.
- Cochran-Smith, M., Villegas, A. M., Abrams, L., Chavez-Moreno, L., Mills, T., & Stern, R. (2015). Critiquing teacher preparation research: An overview of the field, part II. Retrieved from: [Critiquing Teacher Preparation Research \(sagepub.com\)](https://www.sagepub.com/journalsPermissions.nav)
- Darling-Hammond, L. (2014). Strengthening clinical preparation: The holy grail of teacher education. *Peabody Journal of Education*, 89(4), 547-561.
- Darling-Hammond, L. (2020). Accountability in teacher education. *Action in Teacher Education*, 42(1), 60-71.
- Darling-Hammond, L., Newton, S. P., & Wei, R. C. (2013). Developing and assessing beginning teacher effectiveness: The potential of performance assessments. *Educational Assessment, Evaluation and Accountability*, 25, 179-204.
- Dover, A. G., & Schultz, B. D. (2016, January). Troubling the edTPA: Illusions of objectivity and rigor. In *The Educational Forum* (Vol. 80, No. 1, pp. 95-106). Routledge.
- EducationNC (2024). The EdExplainer: The ins and outs of EVAAS. Retrieved from: <https://www.ednc.org/edexplainer-ins-outs-evaas/>
- edTPA. (2023). *edTPA for North Carolina*. [https://www.edtpa.com/PageView.aspx?f=GEN\\_NorthCarolina.html](https://www.edtpa.com/PageView.aspx?f=GEN_NorthCarolina.html)
- Educational Testing Service (2024). edTPA. Princeton Education Inc. Retrieved from: <https://www.edtpa.com/>
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage publications.
- Fuller, E. J. (2014). Shaky Methods, Shaky Motives: A Critique of the National Council of Teacher Quality's Review of Teacher Preparation Programs. *Journal of Teacher Education*, 65(1), 63-77. <https://doi.org/10.1177/0022487113503872>.
- Gitomer, D. H., Martínez, J. F., Battley, D., & Hyland, N. E. (2021). Assessing the assessment: Evidence of reliability and validity in the edTPA. *American Educational Research Journal*, 58(1), 3-31.
- Gilbert, K. A., & Kuo, N-C. (2019). Ethical, legal, and pedagogical issues in edTPA. *Critical Education*, 10(8), 1-17. Retrieved from: <http://ojs.library.ubc.ca/index.php/criticaled/article/view/186418>
- 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., Cowan, J., & Theobald, R. (2016). Evaluating prospective teachers: Testing the predictive validity of the edTPA. Calder Working Paper 157. Available from: <http://www.caldercenter.org/sites/default/files/WP%20157.pdf>
- Goldhaber, D., Liddle, S., & Theobald, R. (2013). The gateway to the profession: Assessing teacher preparation programs based on student achievement. *Economics of Education Review*, 34, 29-44.
- Goldhaber, D., & Ronfeldt, M. (2020). Toward causal evidence on effective teacher preparation. Linking teacher preparation program design and implementation to outcomes for teachers and students, 211-236. Retrieved from: [Linking Teacher Preparation Program Design and Implementation to Outcomes ... - Google Books](#)
- Hair, J. F., Black, W. C., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th, illustr ed.). Haitovsky, Y. (1969). Multicollinearity in regression analysis: Comment. *The Review of economics and statistics*, 486-489.
- Harris, D. N., & Sass, T. R. (2009). *What makes for a good teacher and who can tell?* (Vol. 30). Washington, DC: Urban Institute.
- Henry, G. T., Campbell, S. L., Thompson, C. L., Patriarca, L. A., Luterbach, K. J., Lys, D. B., & Covington, V. M. (2013). The predictive validity of measures of teacher candidate programs and performance: Toward an evidence-based approach to teacher preparation. *Journal of Teacher Education*, 64(5), 439-453.
- Henry, G. T., Kershaw, D. C., Zulli, R. A., & Smith, A. A. (2012). Incorporating teacher effectiveness into teacher preparation program evaluation. *Journal of Teacher Education*, 63(5), 335-355.
- Kaiser, H. F. (1974). An index of factorial simplicity. *psychometrika*, 39(1), 31-36.
- Kim, R. Y., Ham, S. H., & Paine, L. W. (2011). Knowledge expectations in mathematics teacher preparation programs in South Korea and the United States: Towards international dialogue. *Journal of teacher education*, 62(1), 48-61.
- Korthagen, F. A. (2004). In search of the essence of a good teacher: Towards a more holistic approach in teacher education. *Teaching and teacher education*, 20(1), 77-97.
- Leigh, A., & Ryan, C. (2011). Long-run trends in school productivity: Evidence from Australia. *Education Finance and Policy*, 6(1), 105-135.
- Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86-92.
- No Child Left Behind (2004). No child left behind. Retrieved from: <https://files.eric.ed.gov/fulltext/ED521514.pdf>
- North Carolina Department of Public Instruction. (2022). Understand Student Growth Reports in 2022. Retrieved from: <https://ncdpi.sas.com/support/EVAAS-NC-Gap-Year-Resource.pdf>
- North Carolina Department of Public Instruction. (n.d.). *Residency licensure*. <https://www.dpi.nc.gov/educators/educators-licensure/residency-licensure>
- North Carolina Department of Public Instruction. (n.d.). EVAAS. <https://www.dpi.nc.gov>
- Peck, C.A., Singer-Gabella, M., Sloan, T., & Lin, S. (2014). Driving blind: Why we need standardized performance assessment in teacher education. *Journal of Curriculum and Instruction*, 8(1), 8-30.
- Petchauer, E., Bowe, A. G., & Wilson, J. (2018). Winter is coming: Forecasting the impact of edTPA on Black teachers and teachers of color. *The Urban Review*, 50(2), 323-343.
- Primo, D. M., Jacobsmeier, M. L., & Milyo, J. (2007). Estimating the impact of state policies and institutions with mixed-level data. *State Politics & Policy Quarterly*, 7(4), 446-459.
- Russell, V., & Davidson Devall, K. F. (2016). An examination of the edTPA portfolio assessment and other measures of teacher preparation and readiness. *Foreign Language Annals*, 49(3), 479-501.
- SAS (2024). EVAAS North Carolina. Retrieved from: <https://ncdpi.sas.com/support/KeyResearchFindings.pdf>
- SCALE. (2013). 2103 edTPA filed test: Summary report. Available from: [https://secure.aacte.org/apps/rl/res\\_get.php?fid=827&ref](https://secure.aacte.org/apps/rl/res_get.php?fid=827&ref).
- Scheeler, M. C. (2008). Generalizing effective teaching skills: The missing link in teacher preparation. Retrieved from: <https://link.springer.com/article/10.1007/s10864-007-9051-0>
- Swanson, Peter, "The 4 r's of edTPA: Rationale, roadblocks, remediation, and recommendations" (2021). World Languages and Cultures Faculty Publications. 85. Retrieved from: [https://scholarworks.gsu.edu/mcl\\_facpub/85](https://scholarworks.gsu.edu/mcl_facpub/85)
- Teacher Performance Assessment. *Journal Of Interdisciplinary Studies In Education*, 5(2), 11-25.
- Watkins, M. (2022). *A Step-By-Step Guide to Exploratory Factor Analysis with Stata*. Routledge, New York, New York. DOI: 10.4324/9781003149286
- Williams III, J. A., Hart, L. C., & Algozzine, B. (2019). Perception vs. reality: edTPA perceptions and performance for teacher candidates of color and White candidates. Retrieved from: [Perception vs. reality: edTPA perceptions and performance for teacher candidates of color and White candidates - ScienceDirect](#)

Wilson, M., Hallam, P. J., Pecheone, R., & Moss, P. A. (2014). Evaluating the validity of portfolio assessments for licensure decisions. *Education Policy Analysis Archives*, 22(6), n6.