UNIVERSITY of WISCONSIN UWMILWAUKEE

UNIVERSITY OF WISCONSIN SYSTEM

# The Utility and Cost of Requiring Wisconsin Teachers Pass the FoRT and edTPA assessments

SREED™ Socially Responsible Evaluation in Education

*Curtis J Jones University of Wisconsin Milwaukee, Barbara Bales University of Wisconsin System* March 2020

In this study, conducted by The Office of Socially Responsible Evaluation in Education at the University of Wisconsin Milwaukee and the University of Wisconsin System, we explore the utility and impact of the requirement that preservice teachers meet a minimum score on the both the Educative Teacher Performance Assessment (edTPA) and the Wisconsin Foundations of Reading Test (FoRT). Although Wisconsin required prospective teachers to take the edTPA in the 2015-16 school year, teachers have only been required to reach a minimum score since the fall of 2016. The FORT requirement was enacted in the winter of 2014. Both the FoRT and EdTPA are Pearson assessments and self-purported to ensure candidates are prepared before they become a teacher. Besides expensive and time consuming, these mandated assessments may be contributing to Wisconsin's teacher shortage (Umhoeher & Hauer, 2016) and impeding efforts to diversify the state's teaching corps (Jones, 2019). Further, while no research was found on the ability of the FoRT to predict future teacher effectiveness, prior national research on the edTPA has shown it to be modestly predictive of future teacher effectiveness (Bastian et al., 2016; Bastian & Lys, 2016; Goldhaber, Cowan, & Theobald 2016). Further, a recent analysis of the edTPA suggests it has questionable psychometric properties. (Gitomer, Martinez, & Battey, 2020). Finally, little is known regarding the unique "value added" of the edTPA for predicting teacher effectiveness, i.e. how much better are we at predicting the future effectiveness of a teacher when we know their edTPA score than we are by just knowing how well they performed in their preparation program and where they received their certification.

## STUDY QUESTIONS

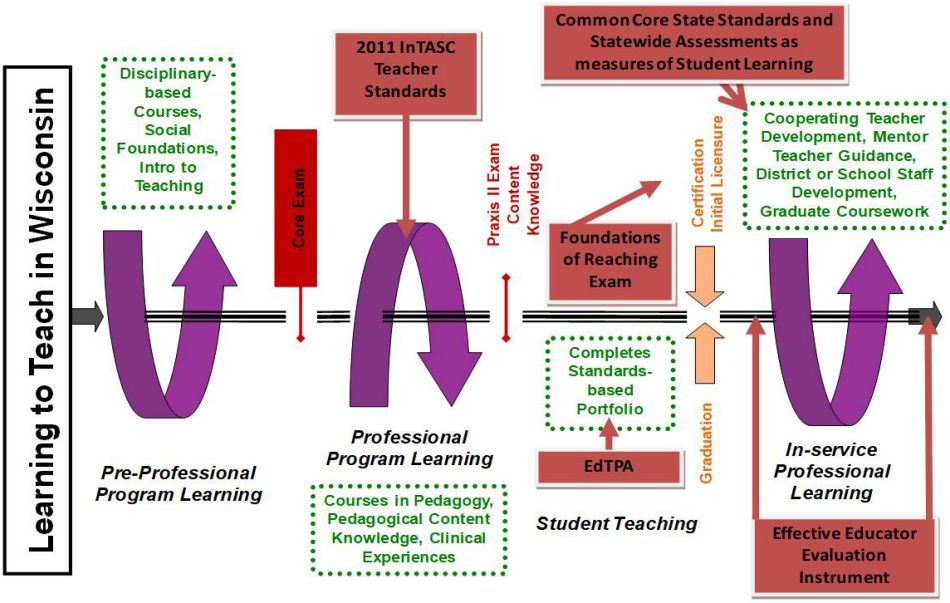
Within this cloudy milieu of Wisconsin's teacher shortage, extensive series of assessments of candidate quality, and the state's EE system, our analysis of these assessments examines two questions:

1. **How do the FoRT and edTPA impact the number of students who become teachers?**
2. **How well do the FoRT and edTPA predict the effectiveness of new teachers?**

The answers should prove useful to Wisconsin education agencies struggling to find ways to increase the number of teachers completing preparation programs. The findings also offer insights for policy makers as to whether the assessments' value in helping ensure new teachers are effective is worth the associated time, financial costs, and risk of graduating fewer teachers.

## What is the assessment process for Wisconsin teachers in training?

The figure below illustrates the learning and assessment experiences in a typical, undergraduate, teacher preparation program. As shown, the edTPA and FoRT are two of four assessments designed to measure a candidate's readiness for the classroom. Before admission into a professional licensure program, students must pass the Praxis Core Academic Skills for Educators (Core). After completing their coursework, students must pass Praxis Subject Assessments for each area they are seeking certification. Later, teacher candidates in elementary, early childhood, special education, and middle-childhood must pass the FoRT. Finally, during their student-teaching experience, candidates must pass the edTPA. Upon hiring, teachers then enter the statewide, teacher evaluation system – Educator Effectiveness (EE), where teachers receive performance ratings on either the 22 components of the Framework for Teaching (FfT) (Danielson, 2013) or six domains of the Stronge Framework (2002).




## Who participated in the study?

Seven University of Wisconsin teacher preparation programs provided student data (assessment scores, completion status, GPA, etc.) from the 2015-16, 2016-17, and 2017-18 school years. Although there are 42 teacher preparation programs across Wisconsin, these seven programs enrolled 7,742 students and graduated 3,808 teachers during this time; numbers that represents roughly 50% of all teachers in training attending Wisconsin teacher education programs and 45% of all graduates.

Teachers who graduated these seven UW preparation programs were matched to state education data. These data provided information about which district employed these teachers and the effectiveness ratings assigned to them as part of the state's EE process. Two thousand six hundred seventy six (2,676) of the 3,808 graduates matched to state data. The other graduates may have taken a position outside of Wisconsin public education. Of these, effectiveness ratings were recorded for 1,740. About 58% of teachers with effectiveness ratings worked within a school district that used the FfT while 42% of teachers were in districts using the Stronge Framework.

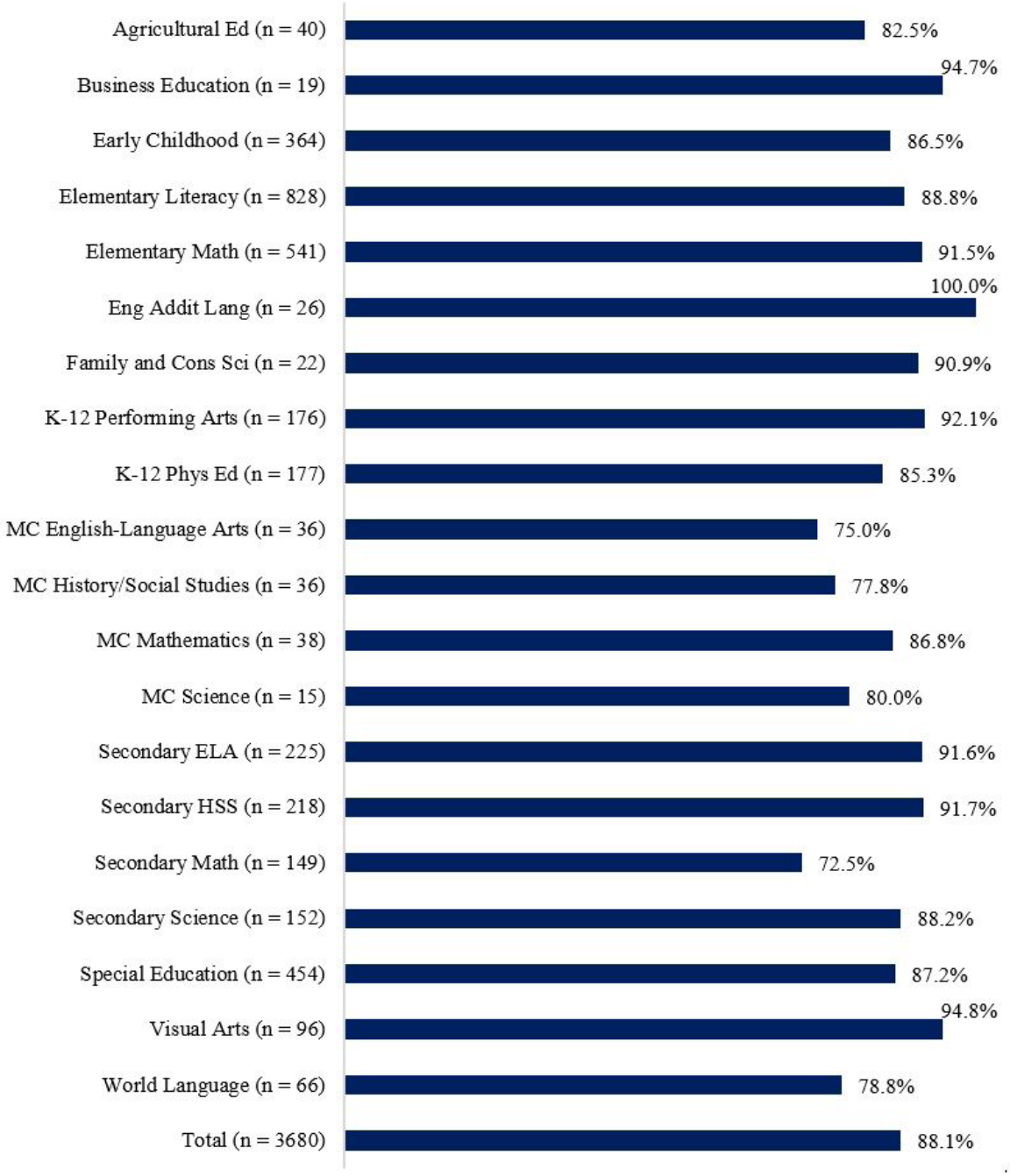
## What methods did we use to answer the two study questions?

Assessment pass rates were first analyzed to determine the number of students who did not meet the requirements for each assessment. This was done overall and by demographic subgroups to explore the possibility the assessments may be disproportionally preventing persons of color from becoming teachers. Bivariate correlations were then used to determine if FORT and edTPA scores were correlated with new teacher effectiveness ratings. Next, we used statistical modeling to isolate the unique relationship between assessment results and teacher effectiveness. This analysis tells us the "value-added" of the FORT and edTPA for predicting teacher effectiveness in addition to the information we already have about teacher candidates, such as their GPA and which program they completed. Finally, we capitalized on the DPI policy that required all 2015-16 teacher candidates to take the edTPA but did not attach any stakes to scores. This created a natural experiment, with teachers who scored below the required threshold on the assessment permitted to become certified teachers. Comparing the effectiveness of these teachers with those who scored above the threshold that year allows for a relatively clean analysis of the predictive validity of the edTPA.

## Most students "passed" the edTPA

Within the seven UW campuses participating in the study during the 2015-16, 206-17, and 2017-18 school years, 88.1% of all edTPA administrations met the required minimum score. However, there were some education areas where teacher candidates were more likely to pass than others suggesting that some of its subject area tests may represent more of a roadblock.

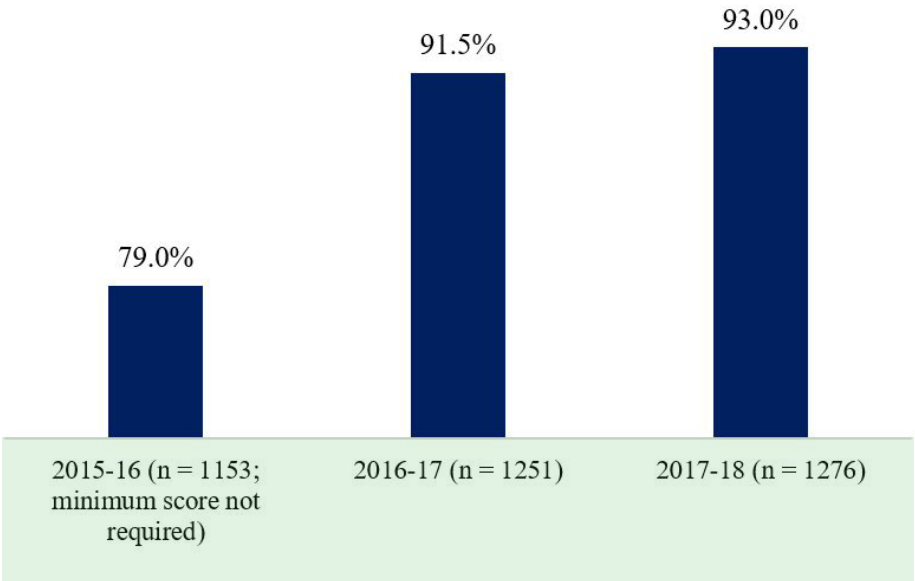
edTPA pass rates per subject area assessment



## Nearly all students passed the edTPA after DPI required students to meet a minimum score.

While only 79% of teacher candidates scored at or above the required score in 2015-16, in 2016-17 and 2017-18, 91.5% and 93%, respectively, of assessment results were at or above the required threshold. Further, teachers who failed to meet the minimum score these years often retook the edTPA and met the requirement. By the end of 2016-17 and 2017-18 school years, 96% and 97% of teacher candidates met the certification requirement, respectively.

edTPA pass rates before and after teacher candidates were required to meet a minimum score




## Only about half of FoRT administrations were "passed".

Although 70% of teacher candidates only took the FoRT one time, the other 30% were likely to take it multiple times. 39% of all teacher candidates who had to retake the FoRT, retook it three or more times. As a result of students having to take the FoRT multiple times, the overall FoRT pass rate was only 54.4%.

## The FoRT is a greater impediment to certification than the edTPA.

By the end of 2017-18, only 27 out of 3,542 teacher candidates withdrew from their program after taking and not passing the edTPA. Regarding FoRT, 98 of the 1,117 students who took the FoRT during the 2015-16 or 2016-17 program years did not meet the requirement and left their program before completion. A number of students who never passed the FoRT were still able to become a teacher.

## The FoRT represents a greater impediment to certification for students of color.

While there is no evidence of any racial differences in edTPA performance, there may be regarding the FoRT. The average White teacher candidate took the FoRT 1.5 times, while persons of color took it 1.8 times. Specifically, Latinx students took the assessment 2.0 times on average. However, 88 of the 98 students who took the FoRT during the 2015-16 or 2016-17 program years did not meet the requirement and left their program before completion were White, suggesting that the assessment did not ultimately prevent many students of color from becoming a certified teacher. Still, that students of color typically have to take the assessment more times to pass suggests the requirement is more of a roadblock for students of color.

## EdTPA and FoRT scores were associated with teacher effectiveness ratings.

Consistent with previous research, as a group, teachers with higher edTPA scores were rated as slightly more effective teachers if they were in a district using the FfT ($r = .112$, $p < .001$) but not if they were in a district using the Stronge Framework ($r = .025$, $p > .05$). Conversely, teachers with higher FoRT scores were rated as more effective if they were in a district using the Stronge Framework ($r = .218$, $p < .05$) but not if they were in a district using the FfT ($r = -.019$, $p > .05$). The relationships between edTPA and FfT effectiveness ratings were mostly consistent across the four effectiveness domains.

One possibility for the differences in the relationships between FoRT and edTPA with the different frameworks producing effectiveness ratings is that the FoRT is only required for a subset of students who take edTPA. Thus, the samples are not equivalent. To account for this, we correlated edTPA and FoRT assessment results with effectiveness ratings, only including teacher candidates who took both assessments. 754 teacher candidates with effectiveness ratings took both assessments. The results did not change from the full sample. Teachers with higher edTPA scores were rated as more effective teachers if they were in a district using the FfT ($r = .142$, $p = .004$) but not if they were in a district using the Stronge Framework ($r = -.027$, $p > .05$). Conversely, teachers with higher FoRT scores were rated as more effective if they were in a district using the Stronge Framework ($r = .104$, $p = .055$) but not if they were in a district using the FfT ($r = .002$, $p > .05$). These confusing results suggest the need for further study.

## After accounting for performance in their education program, EdTPA scores add slightly to our ability to predict the future effectiveness of teacher candidates. FoRT does not.

To estimate the unique impact of edTPA scores and teacher effectiveness ratings, we accounted for the impact of the preparation program, the teacher's cumulative GPA, their race, and gender. Since different districts may have different practices for the types of ratings typically assigned to teachers, we also accounted for the district the teacher was working in. This was ultimately done by accounting for the economic conditions, racial composition, and size of districts teachers worked in. After accounting for all of these factors, edTPA scores were found to be uniquely predictive of FfT effectiveness ratings ($B = 0.081$, $p = .011$) but not Stronge ratings ($B = -0.025$, $p = .53$). FoRT scores did not uniquely predict FfT effectiveness ratings ($B = -0.029$, $p = .607$) or Stronge ratings ($B = 0.053$, $p = .297$).

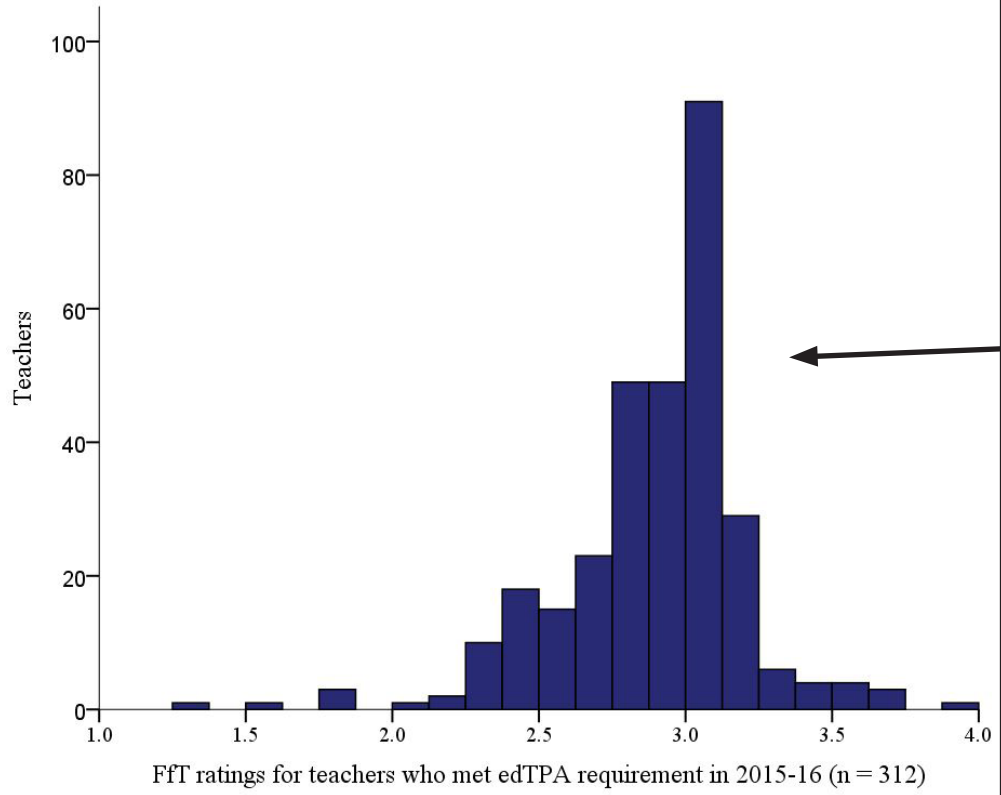
## Teachers who scored below the minimum edTPA requirement during the 2015-16 program year were just as effective as teachers who met the required score.

Effectiveness ratings were available for 131 teachers whose 2015-16 edTPA score was below what was to become the minimum score requirement in 2016-17. Their effectiveness ratings were compared to the 643 teachers from that year who scored above the minimum requirement. The average rating of a teacher who met the edTPA requirement was 2.90 as compared to 2.88 for those below the requirement score. Statistical modelling further demonstrated that these two groups did not perform differently once they were teachers ($B = 0.06$, $p = .618$). The distribution of ratings between the two groups were very similar, regardless of evaluation framework used (FfT or Stronge).

# KEY FINDINGS

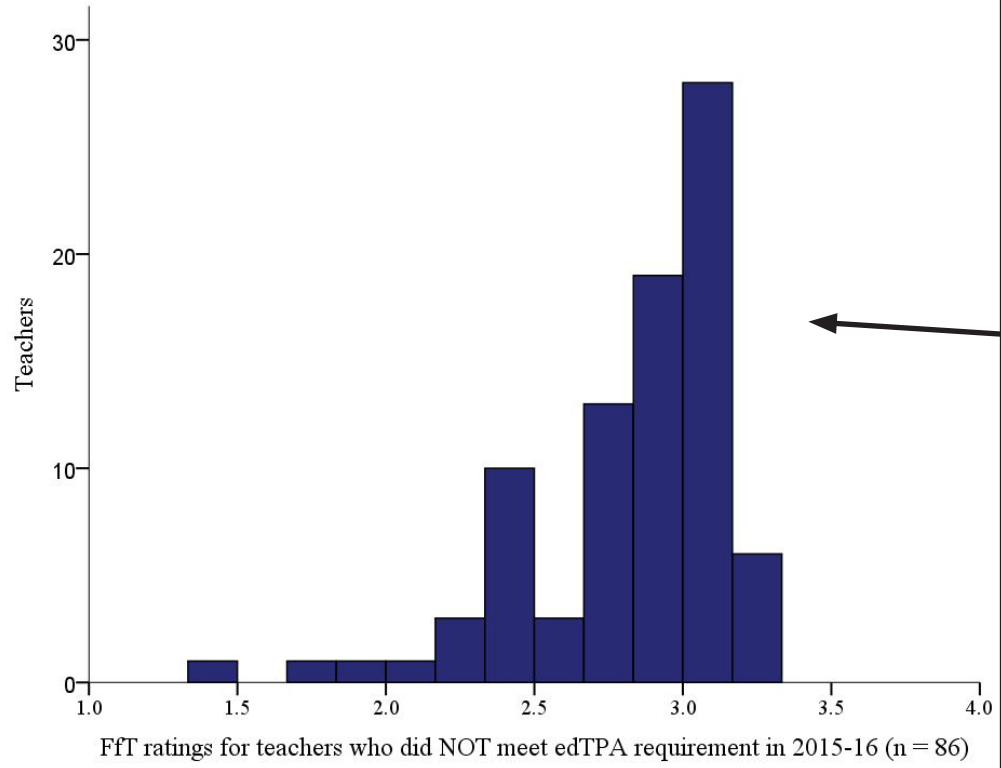
## How well do the FoRT and edTPA predict the effectiveness of new teachers?

FfT effectiveness ratings



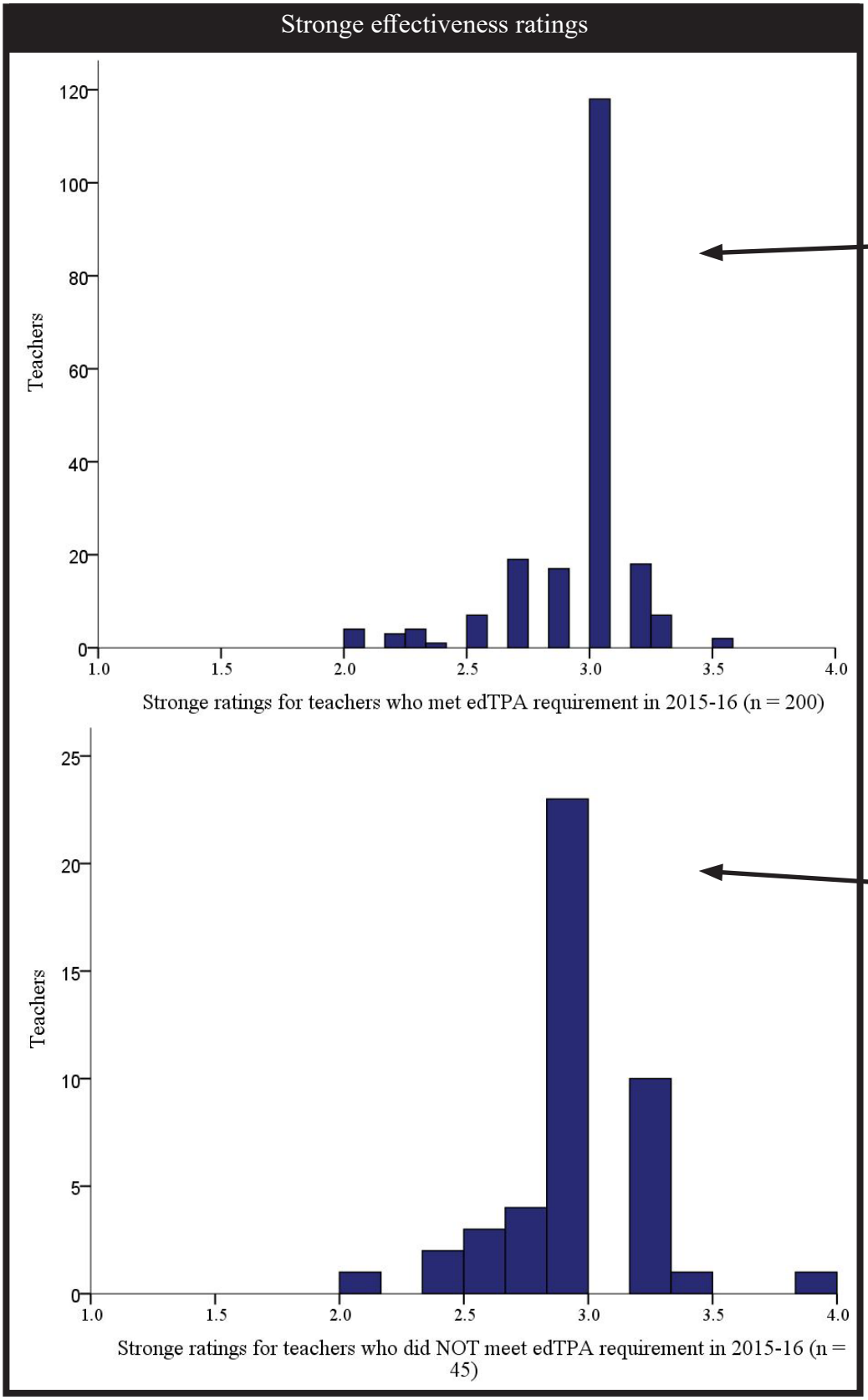

Teachers who "passed" edTPA.

The FfT ratings of 86 teachers who failed to meet the minimum edtpa score were similar to the 312 who met the requirement.




Teachers who did not "pass" edTPA.

# How well do the FoRT and edTPA predict the effectiveness of new teachers?




Teachers who "passed" edtpa.

Teachers who did not "pass" edTPA.

# CONCLUSIONS

- Since nearly all students passed the edTPA, there is little evidence it represents a significant logistical barrier to students becoming teachers. However, our analysis does not speak to the possibility that some potential teachers would decide not to pursue a teaching career because of the cost of taking the edTPA. Future research could explore this possibility by documenting the factors students weigh when choosing to enter into teaching.

- The FoRT assessment represents a greater barrier for students to become certified teachers. Many students had to take the FoRT multiple times before they passed it. Although ultimately, the FoRT prevented few from becoming teachers, the cost of retaking the FoRT and delay in program completion potentially represents a significant barrier to becoming a Wisconsin teacher.

- There is evidence that students of color score lower on the FoRT than White students. White students were more likely to pass the FoRT on their first attempt than students of color. As such, the FoRT is not only a significant barrier preventing students from becoming teachers generally, it represents an even greater barrier to potential teachers of color. Given the small numbers of students in color included in our sample, future research could attempt to replicate this finding including more preparation programs or more years of data.

- Both edTPA and FoRT had small correlations with the effectiveness ratings assigned to teachers. Students who scored higher on these assessments tended to be rated as slightly more effective teachers.

- FoRT scores were not associated with teacher effectiveness ratings when also accounting for their college GPA, where teachers completed their certification program, and other demographics characteristics of the teacher and district. Cumulative college GPA was a better predictor of effectiveness ratings than FoRT scores. Thus, the FoRT does not add to our ability to predict the effectiveness of future teachers.

- edTPA scores were still associated with FfT effectiveness ratings after accounting for cumulative college GPA and where teachers completed their certification program. This finding may suggest it is advisable to continue the requirement that teachers in training score above a required score. However, the relationship was small, suggesting there is not much of a difference in the expected effectiveness of teacher candidates based on their edTPA scores.

- Our analysis of teachers who scored below the threshold in 2015-16, before the minimum score requirement was implemented, suggests they were rated just as effective as those who scored above the threshold that year. This finding challenges the assessment's intended utility and begs the question of whether the Wisconsin education system would be better off had these students never become teachers. Considering the teacher shortage across Wisconsin, the answer to this question would seem to be no. Still, if DPI ultimately chooses to discontinue the edTPA requirement, it will be prudent to track the effectiveness ratings of new teachers after the change.

# REFERENCES


Bastian, K. C., Henry, G. T., Yi, P., & Lys, D. (2016). Teacher candidate performance assessments: Local scoring and implications for teacher preparation program improvement. Teaching and Teacher Education, 59, 1-12.

Bastian, K. C., & Lys, D. (2016). Initial Findings from edTPA Implementation in North Carolina. Retrieved from https://publicpolicy.unc.edu/files/2017/03/edTPABrief-Revised.pdf

Danielson, C. (2013). The Framework for Teaching Evaluation Instrument: 2013 Instructionally Focused Edition. Princeton, NJ: The Danielson Group.

Gitomer, D. H., Martínez, J. F., Battey, D., & Hyland, N. E. (2019). Assessing the Assessment: Evidence of Reliability and Validity in the edTPA. American Educational Research Journal. https://doi.org/10.3102/0002831219890608

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

Jones, C. (2019). Race, Relational Trust, and Teacher Retention in Wisconsin Schools. Available from: https://uwm.edu/sreed/wp-content/uploads/sites/502/2019/11/WEERP-Brief-Nov-2019-Race-Relational-Trust-and-Teacher-Retention.pdf

Stronge, J. H. (2002). Qualities of Effective Teachers. Alexandria, VA: Association for Supervision and Curriculum Development.

# ACKNOWLEDGMENTS


Thanks to the following for helping make this study happen:

Ben Passmore
Carleen Vande Zande

And thanks to the following for reviewing drafts of this report:

Bradley Carl
Steve Kimball
Katharine Rainey

A special thanks to David Blough for organizing the data collection and cleaning process.

*Curtis Jones* is the Director of the Office of Socially Responsible Evaluation in Education at the University of Wisconsin Milwaukee

*Barbara Bales* is the Director of Strategic Initiatives and Educational Innovation - PK-20 Education Programs at the University of Wisconsin System Administration

Funding for this study was provided by the University of Wisconsin System.

**For more information about this report please contact Curtis Jones at jones554@uwm.edu or visit www.uwm.edu/sreed**


## Data provided by UW System

We worked with seven UW system campuses to provide information on students in teacher preparation programs across the 2015-16, 2016-17, and 2017-18 program years. We received data for 7,742 teachers in training, which represents roughly 50% of all teachers in training in Wisconsin teacher education programs during that period. Demographic (race, gender, name, and birthdate) and performance data (GPA and assessment results) were provided for 7,413 of these students. Similar to statewide data, the students in the seven participating teacher education program were mostly White (between 91% and 92%). The student data provided by the seven campuses included 3,808 graduates, which represents 44.6% of all graduates statewide during the period.

Student demographics in the seven teacher preparation programs

| | 2015-16 | | 2016-17 | | 2017-18 | |
|---|---|---|---|---|---|---|
| | Students | % | Students | % | Students | % |
| African American | 30 | 0.7% | 33 | 0.8% | 42 | 1.0% |
| Latinx | 142 | 3.3% | 131 | 3.2% | 157 | 3.7% |
| Native American | 11 | 0.3% | 5 | 0.1% | 7 | 0.2% |
| Southeast Asian | 42 | 1.0% | 38 | 0.9% | 34 | 0.8% |
| Other Asian American | 22 | 0.5% | 28 | 0.7% | 29 | 0.7% |
| White | 3922 | 92.3% | 3756 | 91.7% | 3837 | 90.6% |
| Other | 4 | 0.1% | 6 | 0.1% | 13 | 0.4% |
| Two or more races | 74 | 1.7% | 98 | 2.4% | 116 | 2.7% |
| Total | 4247 | | 4095 | | 4235 | |
| Female | 3301 | 77.7% | 3177 | 77.6% | 3295 | 77.8% |
| Male | 950 | 22.3% | 919 | 22.4% | 940 | 22.2% |
| Total | 4251 | | 4096 | | 4235 | |

## Data provided by the Wisconsin Department of Public Instruction (DPI)

Wisestaff – UW Sytem graduates were then matched to WiseStaff data managed by the DPI. Graduate birthdate and name were used to match with Wisestaff records from the 2016-17, 2017-18, and 2018-19 school years. Wisestaff data includes a greater deal of information about the educator but we were mostly interested in if they were included in the file, indicating they were working as a teacher, and the district they were working in. 70% (2,676) were successfully matched to state data, confirming their employment as a public education teacher. While fewer teachers where matched in 2017-18 and 2018-19 than in 2016-17, this is likely due to the reduced time these teachers have been on the job market, rather than fewer teachers entering public education. District demographic characteristics were also pulled from data available from DPI.

Teachers from the seven UW campuses matched to state employment data

| | All Teachers | % |
|---|---|---|
| Asian | 37 | 1.4% |
| Black or African American | 19 | 0.7% |
| Hispanic/Latino | 49 | 1.8% |
| Other | 23 | 1.4% |
| White | 2548 | 95.2% |
| Female | 2040 | 76.2% |
| Male | 636 | 23.8% |
| Total | 2676 | 100% |

My Learning Plan – Districts use the My Learning Plan data system to document their Educator Effectiveness processes, including the effectiveness ratings assigned to teachers. Of the 2,676 teacher candidates matched to state data, effectiveness ratings were recorded for 1,740 teachers. At the end of the year, new teachers receive performance ratings from their evaluator on either the 22 components of the Framework for Teaching (FfT (Danielson, 2013) or 6 domains of the Stronge Framework (2002). About 58% of teachers with effectiveness ratings worked within a school district that used the FfT. 42% of teachers were in districts using the Stronge Framework.

| | | Teachers | Percent |
|---|---|---|---|
| Educator Effectiveness Rubric | CESA 6 Stronge Framework | 687 | 42.4% |
| | Framework for Teaching | 934 | 57.6% |
| Gender | Missing | 13 | 0.8% |
| | Female | 1226 | 75.6% |
| | Male | 382 | 23.6% |
| Race/Ethnicity | Missing | 13 | 0.8% |
| | Asian | 16 | 1.0% |
| | African American | 10 | 0.6% |
| | Latinx | 29 | 1.8% |
| | Other | 15 | 0.9% |
| | White | 1538 | 94.9% |
| Year graduated | 2015-16 | 670 | 38.5% |
| | 2017-17 | 583 | 33.5% |
| | 2017-18 | 487 | 28.0% |
| First year evaluated | 2016-17 | 483 | 27.8% |
| | 2017-18 | 623 | 35.8% |
| | 2018-19 | 634 | 36.4% |

Teachers are assigned either a (1) Unsatisfactory, (2) Basic, (3) Proficient or (4) Distinguished rating on each of the 22 components (FfT) or six domains (Stronge). FfT component ratings are averaged to calculate domain ratings and domain ratings are averaged to calculate an overall effectiveness rating. Stronge Model ratings are averaged to calculate an overall rating. Overall ratings for new teachers were similar between the two frameworks, averaging 2.94 (Stronge) and 2.88 (FfT).

Teacher performance results in their preparation program and in their teaching position

| | Teachers | Min | Max | Mean | Std. Deviation |
|---|---|---|---|---|---|
| CESA6 Strong Model ratings | 760 | 1.50 | 4 | 2.94 | 0.24 |
| FfT ratings | 980 | 1.27 | 4 | 2.88 | 0.31 |
| FfT Domain 1 – Planning and Preparation | 979 | 1.00 | 4 | 2.84 | 0.36 |
| FfT Domain 2 – The Classroom Environment | 980 | 1.20 | 4 | 2.91 | 0.41 |
| FfT Domain 3 – Instruction | 980 | 1.20 | 4 | 2.85 | 0.35 |
| FfT Domain 4 – Professional Responsibilities | 979 | 1.00 | 4 | 2.91 | 0.30 |
| Cumulative Grade Point Average | 1693 | 2.73 | 4 | 3.60 | 0.27 |
| edTPA score | 1631 | 19 | 74 | 45.24 | 6.18 |

Characteristics of districts where teachers were working

| | Districts | Min | Max | Mean | Std. Deviation |
|---|---|---|---|---|---|
| District Size | 270 | 115 | 75431 | 2561 | 5542 |
| District % eligible f/r lunch | 270 | 1.6% | 95.6% | 39.0% | 16.0% |
| District % White | 270 | 0% | 98.3% | 81.2% | 17.1% |
| District % African American | 270 | 0% | 98.0% | 3.1% | 10.9% |
| District % Latinx | 270 | 0% | 97.8% | 8.7% | 10.3% |

# CORRELATIONS

Correlations between teacher performance in their preparation program, their effectiveness ratings, and characteristics of the district they worked in

| | edTPA score | FORT score | Cumulative Grade Point Average | CESA6 Stronge Model ratings | FfT ratings |
|---|---|---|---|---|---|
| edTPA score | 1 | | | | |
| FORT score | .159** | 1 | | | |
| Cumulative Grade Point Average | .202** | .337** | 1 | | |
| CESA6 Stronge Model ratings | .025 | .218* | .155** | 1 | . |
| FfT ratings | .112** | -.019 | .135** | .c | 1 |
| FfT Domain 1 – Planning and Preparation | .103** | .021 | .121** | .c | .905** |
| FfT Domain 2 – The Classroom Environment | .067* | -.028 | .100** | .c | .872** |
| FfT Domain 3 – Instruction | .123** | -.008 | .118** | .c | .902** |
| FfT Domain 4 – Professional Responsibilities | .110** | -.054 | .149** | .c | .855** |
| District Size | .001 | .044 | -0.008 | -.006 | -.240** |
| District % eligible f/r lunch | -.077** | -.002 | -.073** | -.023 | -.270** |
| District % White | .007 | .015 | .033 | .036 | .246** |
| District % AA | -.016 | .053 | -.058* | -.007 | -.223** |
| District % Latinx | .03 | -.109* | -.017 | -.043 | -.143** |

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).
c Cannot be computed because at least one of the variables is constant.

## Statistical Models Predicting Effectiveness Ratings

Generalized linear models (GLM) with linear error terms, an identity link function, and robust standard errors were used to estimate the relationships of edTPA scores and cumulative GPA with effectiveness ratings. Effectiveness ratings for both FfT and Strong Framework were standardized. edTPA sores, FoRT scores, and GPA were also standardized. Three separate modeling strategies were used to isolate the unique relationships of edTPA, FoRT, and GPA with effectiveness ratings. Each included teacher demographics and teacher preparation program as covariates. Each model treated school district differently though. The first model included the fixed district effect. This approach was rejected however, as many districts only had one or two teachers in the analysis. Thus, the inclusion of district results in many teachers being excluded from the analysis. The second approach did not include district. This was rejected since much of a teacher's effectiveness rating is related to where they work as a teacher. The third method controlled for fixed district characteristics but did not control for the district itself. This was ultimately chosen as the preferred strategy since it controlled for district differences without excluding any teachers from the analysis.

Model 1 (fixed district effects):

$$Y_{ij} = \beta_0 + \beta_1(GPA_{ij}) + \beta_2(edTPA_{ij}) + \sum_{s=1}^{S-1} \beta_{3.s} X_{sij} + \sum_{m=1}^{M} \beta_{4.m} X_{mij} + \sum_{j=1}^{J-1} \beta_{5.j} district_j + \varepsilon_i$$

Model 2 (no district effects):

$$Y_{ij} = \beta_0 + \beta_1(GPA_{ij}) + \beta_2(edTPA_{ij}) + \sum_{s=1}^{S-1} \beta_{3.s} X_{sij} + \sum_{m=1}^{M} \beta_{4.m} X_{mij} + \varepsilon_i$$

Model 3 (fixed district characteristics):

$$Y_{ij} = \beta_0 + \beta_1(GPA_{ij}) + \beta_2(edTPA_{ij}) + \sum_{s=1}^{S-1} \beta_{3.s} X_{sij} + \sum_{m=1}^{M} \beta_{4.m} X_{mij} + \sum_{n=1}^{N} \beta_{5.n} X_{nj} + \varepsilon_i$$

$Y_{ij}$ is the Effectiveness Rating for the ith teacher in the j$^{th}$ district; $\beta_0$ is the intercept; $\beta_1$ is the relationship of Cumulative College GPA with Effectiveness Ratings; $GPA_{ij}$ is the Cumulative GPA for teacher i in district j; $\beta_2$ is the relationship of edTPA or FoRT score with Effectiveness Ratings; $edTPA_{ij}$ or $FoRT_{ij}$ is the edTPA score or FoRT score for teacher i in district j; $\beta_{3.s}$ is the effect of the teacher preparation program; $X_{sij}$ = the s$^{th}$ of S teacher preparation programs of teacher i in district j; $\beta_{4.m}$ is the effects of teacher demographics; $X_{mij}$ = the m$^{th}$ of M demographic characteristics of teacher i in district j (gender and race); $\beta_{5.j}$ = the effect of district (i.e., the difference in the intercept between district j and the reference district); $\beta_{5.n}$ = the effects of district characteristics; $X_{nj}$ = the n$^{th}$ of N district characteristics (EE model - FfT or Stronge, district size, % eligible for F/R lunch, % White, % African American, and % Latinx); and $\varepsilon_i$ is the error term.

Teacher included in statistical models predictive effectiveness ratings

| | Teachers with edTPA scores | Teachers with FoRT scores |
|---|---|---|
| Models predicting Stronge ratings | 723 | 320 |
| Models predicting FfT ratings | 985 | 413 |
| Models predicting combined ratings | 1708 | 733 |

## Modeling Results - FoRT Predicting Effectiveness Ratings

| Outcome | Predictor | *B* | Robust Std. Error | Sig. |
|---|---|---|---|---|
| Model 1: Summary model results with college, school performance, teacher demographics, and fixed district effects included in models | | | | |
| Combined ratings | FoRT | -0.002 | 0.042 | 0.962 |
| | GPA | 0.154 | 0.051 | 0.003 |
| FfT ratings | FoRT | 0.003 | 0.059 | 0.964 |
| | GPA | 0.174 | 0.079 | 0.028 |
| Stronge ratings | FoRT | 0.065 | 0.051 | 0.206 |
| | GPA | 0.087 | 0.052 | 0.091 |
| Model 2: Summary model results with college, school performance, and teacher demographics included in models | | | | |
| Combined ratings | FoRT | 0 | 0.039 | 0.995 |
| | GPA | 0.095 | 0.045 | 0.034 |
| FfT ratings | FoRT | -0.036 | 0.056 | 0.522 |
| | GPA | 0.123 | 0.069 | 0.075 |
| Stronge ratings | FoRT | 0.065 | 0.051 | 0.206 |
| | GPA | 0.087 | 0.052 | 0.091 |
| Model 3: Summary model results with college, school performance, teacher demographics, and district demographics included in models | | | | |
| Combined ratings | FoRT | 0.001 | 0.039 | 0.979 |
| | GPA | 0.085 | 0.045 | 0.058 |
| FfT ratings | FoRT | -0.029 | 0.056 | 0.607 |
| | GPA | 0.097 | 0.069 | 0.158 |
| Stronge ratings | FoRT | 0.053 | 0.051 | 0.297 |
| | GPA | 0.096 | 0.053 | 0.070 |

## Modeling Results - edTPA Predicting Effectiveness Ratings

| Outcome | Predictor | *B* | Robust Std. Error | Sig. |
|---|---|---|---|---|
| Model 1: Summary model results with college, school performance, teacher demographics, and fixed district effects included in models | | | | |
| Combined ratings | edTPA | 0.066 | 0.025 | 0.009 |
| | GPA | 0.120 | 0.029 | < .001 |
| FfT ratings | edTPA | 0.090 | 0.033 | 0.006 |
| | GPA | 0.087 | 0.039 | 0.027 |
| Stronge ratings | edTPA | 0.008 | 0.039 | 0.833 |
| | GPA | 0.175 | 0.043 | < .001 |
| Model 2: Summary model results with college, school performance, and teacher demographics included in models | | | | |
| Combined ratings | edTPA | 0.058 | 0.025 | 0.020 |
| | GPA | 0.129 | 0.028 | < .001 |
| FfT ratings | edTPA | 0.102 | 0.032 | 0.002 |
| | GPA | 0.101 | 0.038 | 0.007 |
| Stronge ratings | edTPA | -0.018 | 0.037 | 0.633 |
| | GPA | 0.180 | 0.041 | < .001 |
| Model 3: Summary model results with college, school performance, teacher demographics, and district demographics included in models | | | | |
| Combined ratings | edTPA | 0.046 | 0.025 | 0.059 |
| | GPA | 0.128 | 0.028 | < .001 |
| FfT ratings | edTPA | 0.081 | 0.032 | 0.011 |
| | GPA | 0.093 | 0.037 | 0.012 |
| Stronge ratings | edTPA | -0.025 | 0.037 | 0.503 |
| | GPA | 0.184 | 0.042 | < .001 |

## Modeling the effectiveness of teachers who scored above or below the required edTPA score in 2015-16

In this analysis, we take advantage of a natural experiment. In 2015-16 DPI required all teacher candidates to take the edTPA. However, there were no stakes attached to the results, i.e. teachers did not have to meet any minimum score to become certified. Teachers who scored below the threshold would have been prevented from becoming a teacher in subsequent years. In this analysis we compare the effectiveness ratings of teachers who scored below the threshold to those who scored above it.

There were 131 teachers who scored below the threshold who later became teachers. 86 were rated using the FfT and 45 with the Stronge Framework. 200 teachers who took edTPA that year scored above the required minimum score and were later rated with the Stronge Framework, while 312 were rated with the FfT.

Charateristics of teachers who took edTPA in 15-16, became teachers, and received EE ratings

| | Teachers | Teachers |
|---|---|---|
| edTPA | Met EdTPA | 512 |
| | Below edTPA | 131 |
| District model | Stronge | 245 |
| | FfT | 398 |
| Sex | Female | 499 |
| | Male | 144 |

The effectiveness ratings of teachers who scored below the required edTPA score were compared to that of teachers who scored above the score. The model used for this is a derivation of Model 3. The only difference was that edTPA scores were modified to measure the difference in effectiveness ratings for teachers who scored above and below the score required in subsequent years.

Summary model results with college, school performance, teacher demographics, and district demographics included in models

| Outcome | Predictor | *B* | Robust Std. Error | Sig. |
|---|---|---|---|---|
| Combined ratings | Passed edTPA | 0.053 | 0.1198 | 0.661 |