



Estimating the Impact of Teacher Preparation Programs and Staff Characteristics on Outcome Measures

Technical Report

Clint Sattler
Supervisor, Research and Evaluation
Knox County Schools
Department of Research, Evaluation, and Assessment

June 2017

Overview

Interim Knox County schools superintendent Buzz Thomas set a goal in the Fall of 2016 to become the best school system in the South. Working towards this goal continues to be a focus for the Great Schools Partnership and other strategic partners of the school system. Current efforts are focused on supporting innovative and promising programs that are expected to positively impact the school system.

The Knox County schools' strategic plan states, "Our students can only achieve at their highest levels when our teachers, leaders and staff also excel in their work" (Knox County Schools, 2013). Activities that support achieving this goal include continual professional development of our current staff and ensuring the district is hiring the highest quality candidates. This analysis seeks to quantify differences in teacher outcomes based on the teacher preparation program (TPP) in which they were trained. An understanding of the sources of variation in teacher performance may impact targeted recruitment practices in the future.

There are growing movements across the nation to hold TPPs accountable for the quality of their programs, but there is considerable debate about how this can be measured (Cochran-Smith, 2016). The bulk of the available research focuses on rating TPPs by their selectivity (both among applicants and faculty), quality of the curriculum, pass rates of certification exams, and hiring rates (Feuer, 2013). There is an emerging body of research that attempts to estimate TPP effectiveness with impacts on students (Cochran-Smith, 2016). Several states are investigating the use of teacher effectiveness (as estimated by student growth) in evaluating TPP quality (Boyd, 2009; Henry, 2012). A recent study in Tennessee used classroom observation scores to detect significant differences in teacher ratings based on the TPP a teacher attended (Ronfeldt, 2016). However, none of these output-focused methods are ideal as they rely on single point-estimates of teacher effectiveness.

This study takes a purely quantitative approach to estimating TPP effectiveness using multiple measures. Teacher value-added estimates were combined with classroom observation data and other teacher characteristics in order to estimate teacher impacts in the classroom. The population studied in this research was limited to a subset of teachers that generated value-added data from state assessments. The results of this study is not intended to be generalized beyond this population of teachers.

The results of the analysis identified factors that significantly impacted the probability that a teacher would have higher than average classroom observation and value-added scores. In most cases, the TPP at which a teacher was trained had less impact on performance when compared to characteristics such as undergraduate grade point average and previous teaching experience. Other findings of this study can refine the district's recruitment efforts and identify critical strategic partners for new-teacher development.

Methodology

The data used in this analysis was obtained from several systems. Undergraduate institution, undergraduate grade point average, and highest degrees obtained were obtained from personnel files maintained by the Human Resources department. Only one cohort of teachers was used in this study because this data collection was a very labor-intensive process. The initial sample of teachers used in this study was limited to any teacher who generated an individual value-added score in 2014-2015 (SY1415). Each teacher in the initial pool was then linked to their years of experience (from their salary step in the KCS human resources database, NextGen), current position (from NextGen), value-added index (archived data from the TVAAS site, <https://tvaas.sas.com>), and classroom observation data (from the RANDA Tower database).

Analysis of variance (ANOVA) testing and graphical inspection were used for exploratory analysis of the dependent variables. Graphical inspection was used to determine the optimum method to partition the data in order to calculate unbiased parameter estimates. TVAAS indices exhibited significant variation in distributions by the subject area tested, therefore all TVAAS indices were normalized within subjects. The classroom observation data were standardized within schools by subtracting the average apprentice teacher observation score from each teacher-level observation score. This procedure accounted for school-to-school variation in rating scores. Cross tabulation was used on the raw data to determine which TPPs had sufficient number of alumni for statistical modeling.

The modeling used both classroom observation data and value-added estimates as dependent variables. The standardized observation score and the normalized TVAAS index were used to classify teachers into two categories. Teachers were classified as having high standardized observation scores and high normalized valued-added scores if both scores were greater than the median value of the sample population. Teachers were classified as having low standardized observation scores and low normalized value-added if both scores were less than the median value of the sample population. The independent variables used in the study were the undergraduate institution the teachers attended (their TPP), the teachers' years of experience, undergraduate GPA, subject area taught, grade-level taught, and if the teacher was ever awarded an advanced degree. The analysis did not consider whether an advanced degree was awarded in a teacher's content area. The available data only indicated if a teacher was awarded any kind of advanced degree. The dependent variables and independent variables were linked using the logit function. The final models were constructed for parsimony (through monitoring AIC), and verified for precision using receiver operator characteristic (ROC) curves and residuals analysis.

The software used for this analysis were Tableau version 9.1.0 (graphical exploration), IBM SPSS Statistics version 24 (ANOVA and ROC curves), and R version 3.4.0 (on RStudio 1.0.143 for logistic regression).

Results

Results: Schools for analysis

The raw teacher data set contained 1,924 teachers with a value-added score from 347 TPPs. However, there were only 14 TPPs with 10 or more alumni earning a value-added composite score in SY1415. Effective retention rates from SY1415 to SY1617 were determined for each TPP. Retirees and deceased teachers were considered effectively retained in the district. The retention rates for each TPP with 10 or more teachers can be found in the Table 1 and Figure 1.

Table 1: TPP Population Statistics and Effective Retention Rates

TPP	n	% of Sample	Effective Retention Rate
TPP 1	790	41.1%	82.2%
TPP 2	104	5.4%	84.6%
TPP 3	102	5.3%	71.6%
TPP 4	98	5.1%	82.7%
TPP 5	64	3.3%	70.3%
TPP 6	41	2.1%	58.5%
TPP 7	29	1.5%	65.5%
TPP 8	28	1.5%	78.6%
TPP 9	21	1.1%	81.0%
TPP 10	21	1.1%	81.0%
TPP 11	14	0.7%	85.7%
TPP 12	13	0.7%	76.9%
TPP 13	12	0.6%	75.0%
TPP 14	10	0.5%	50.0%
All Other TPPs	577	30.0%	71.8%
All TPPs	1924	100.0%	77.2%

The data were aggregated by TPP for the 14 institutions with at least 10 alumni. The mean school-level retention rate for these 14 TPPs was 74.5%. The upper bound of the 95% confidence interval was 80.6% and the lower bound of the 95% confidence interval was 68.5%. TPPs 1, 2, 4, 9, and 10 had effective retention rates that were greater than the upper bound of the 95% confidence interval. TPPs 6, 7, and 14 had effective retention rates that were less than the lower bound of the 95% confidence interval.

The data in table 1 also indicates that the majority of the teachers in the SY1415 cohort who had individual TVAAS scores were coming from a small number of TPPs. More than 50% of the cohort attended 3 TPPs.

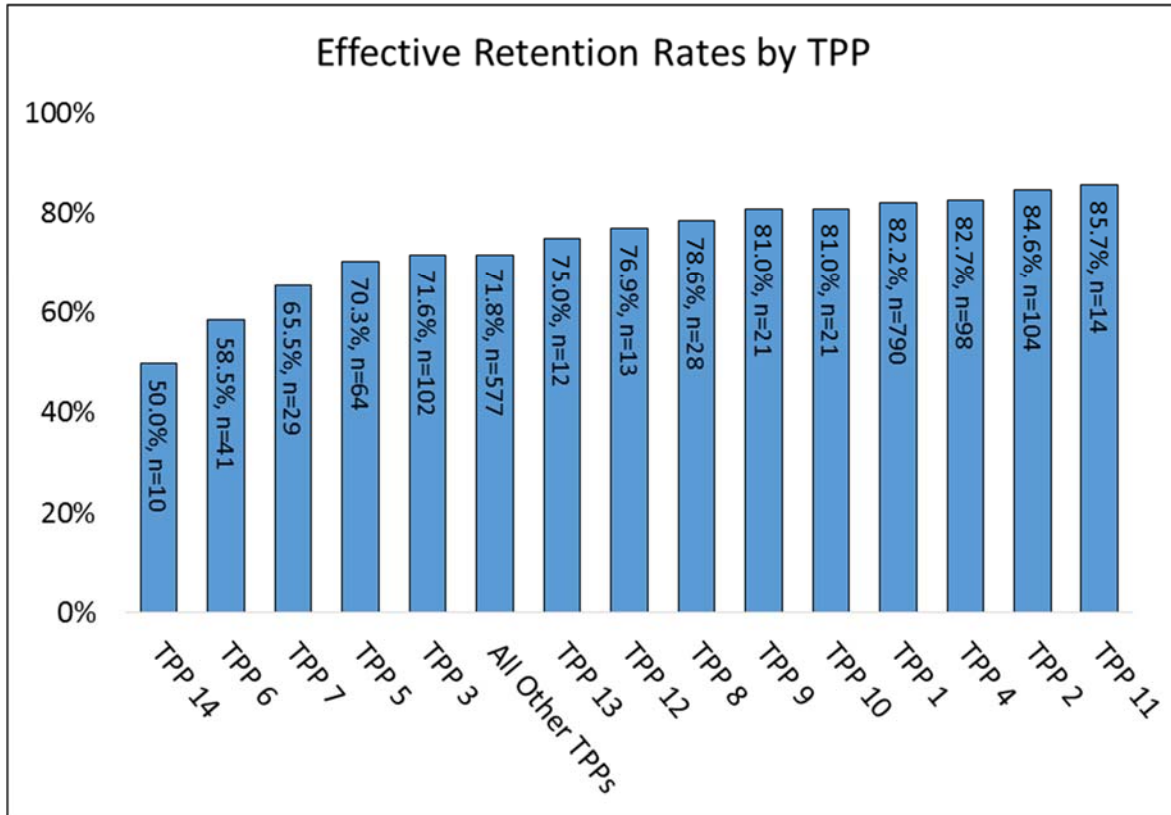


Figure 1: Effective Retention Rates by TPP

Results: Logistic Modeling

There was attrition in the sample due to teachers without complete observation scores (likely due to approved leave during the SY1415 school year) and teachers who did not have undergraduate GPAs in their personnel files. Additionally, the model was constructed using TPPs with at least 10 alumni in the cohort (TPPs 10-14 were included with “All Other TPPs” in further analysis). Eight hundred forty eight teachers were included in the sample used for logistic modeling.

Results: Logistic Modeling: High Observation Scores and High TVAAS Scores

Teachers were coded as a success (dependent variable=1) in this model if their standardized observation score and normalized TVAAS index were greater than the median value of the sample. The most parsimonious model estimates (in log odds units) can be found in Table 2. TPP 1 was used as the TPP reference factor in the analysis because the majority of the cohort attended TPP 1. Four or more years of experience was used as the experience level

reference factor in the analysis because most of the teachers in the sample were veteran teachers.

Table 2: Logistic Model Parameters (High Observation and High TVAAS)

Parameter	Estimate	exp(B)	Std. Error	z value	p value
Intercept	-1.490	0.225	0.601	-2.48	0.013*
Years of Experience (1 or Fewer Years)	-2.134	0.118	0.440	-4.85	0.000*
Years of Experience (2 or 3 years)	-0.891	0.410	0.251	-3.54	0.000*
TPP 2	0.281	1.324	0.323	0.87	0.385
TPP 3	-0.592	0.553	0.419	-1.41	0.158
TPP 4	-0.012	0.988	0.371	-0.03	0.974
TPP 5	-2.664	0.070	1.040	-2.56	0.010*
TPP 6	-0.956	0.384	0.658	-1.45	0.146
TPP 7	-0.876	0.417	0.686	-1.28	0.202
TPP 8	-0.824	0.439	0.808	-1.02	0.308
TPP 9	-0.433	0.649	0.708	-0.61	0.541
All Other TPPs	-0.120	0.887	0.178	-0.67	0.501
Undergraduate GPA	0.434	1.544	0.182	2.39	0.017*
Advanced Degree = Yes	-0.3839	0.681	0.1726	-2.224	0.026*

Chi squared statistics indicate that the model is significant ($p=6.2e-11$). The Nagalkerke R^2 for the fit is 0.119. This relatively low R^2 indicates that the independent variables only describe a small portion of the overall variance in teacher performance. Years of experience, undergraduate GPA, and attainment of an advanced degree were significant predictors influencing the probability of having an observation score and a TVAAS index greater than the sample median. Attending TPP 5 was associated with negatively impacting the probability of a teacher having both an observation score and TVAAS index greater than the median of the sample. The list below summarizes the significant results from the model.

- Teachers with 1 or fewer years of experience were 8 (1/0.118) times less likely to have a standardized observation score greater than the median and a normalized TVAAS index greater than the median when compared to a veteran teacher.
- Teachers with 2 or 3 years of experience were 2.5 (1/0.410) times less likely to have a standardized observation score more than the median and a normalized TVAAS index more than the median when compared to a veteran teacher.

- Graduates from TPP 5 were 14 (1/0.070) times less likely to have a standardized observation score more than the median and a normalized TVAAS index more than the median when compared to a teacher from TPP 1.
- A one point increase in undergraduate GPA means that a teacher is 1.5 times more likely to have a standardized observation score more than the median and a normalized TVAAS index more than the median.
- Obtaining an advanced degree makes a teacher 1.5 (1/0.681) times less likely to have a standardized observation score greater than the median and a normalized TVAAS index greater than the median.

The output from the logistic model provides the probability that a specific teacher would have a standardized observation score greater than the median of the sample and a normalized TVAAS index greater than the median of the sample. A receiver operator characteristic (ROC) curve was plotted using the model output versus the observed data. This was done to determine if the model was a more accurate classifier of teacher performance than the overall frequency of success in the sample. The area under the curve was estimated at 66.2%. The 95% confidence interval had an upper bound of 69.9% and a lower bound of 62.4%. The ROC curve provides evidence that the model is a more accurate classifier of teacher performance because this interval does not include 50%. The cut-off value that maximizes the difference between the true positive identification rate and the false positive identification rate was 33.3%. This cut-off value results in a true positive rate of 77.7% and a false positive rate of 52.6%. Approximately 24% of the high performing teachers had a probability of success less than 34.6%. The data regarding the accuracy of the fit can be found in Table 3. The ROC curve is available in Figure 2.

Table 3: Classification Table (High observation and High TVAAS)

TPP	Observed Success Rate	Success Rate as Modeled	% of Teacher Correctly Classified
TPP 1	35.4%	74.0%	52.8%
TPP 2	44.0%	82.0%	50.0%
TPP 3	22.5%	20.0%	72.5%
TPP 4	33.3%	56.4%	51.3%
TPP 5	4.2%	0.0%	95.8%
TPP 6	16.7%	0.0%	83.3%
TPP 7	21.4%	0.0%	78.6%
TPP 8	14.3%	0.0%	85.7%
TPP 9	27.3%	36.4%	72.7%
All Other TPPs	34.0%	60.8%	53.6%
All TPP	32.8%	60.4%	56.8%

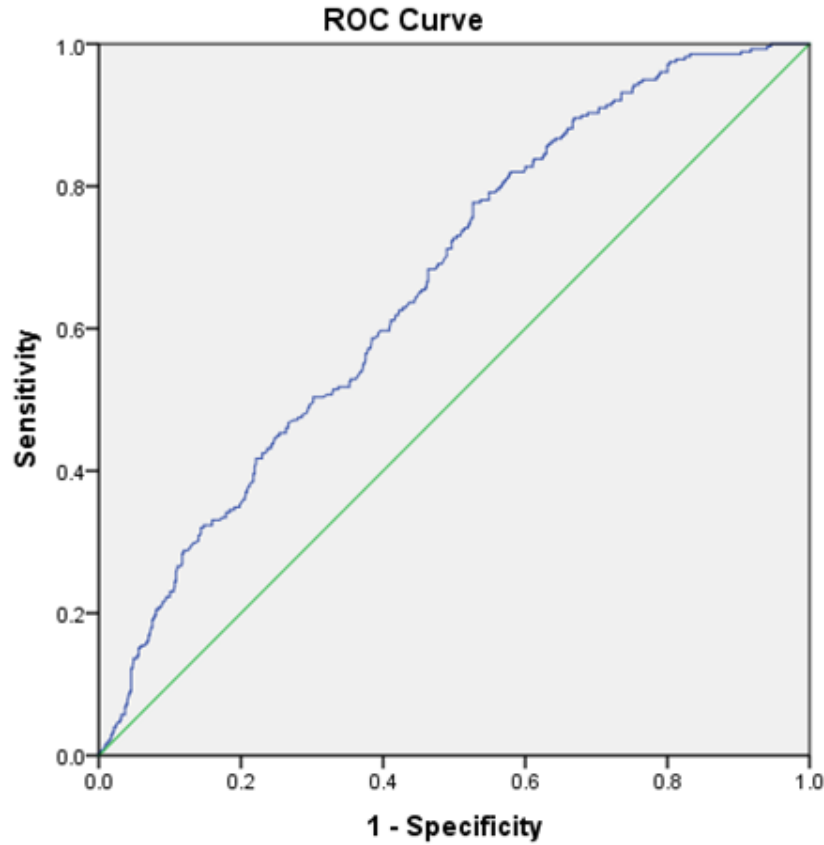


Figure 2: ROC Curve (High observation and High TVAAS)

Results: Logistic Modeling: Low Observation Scores and Low TVAAS Scores

The analysis of the residuals from this regression indicates that there may be some non-linearities in the log transformed data, and therefore the model estimates may be biased. However, the similarities in the results from this analysis, and the high observation scores and high TVAAS scores regression (in which residuals do not show patterns indicative of nonlinearities) provides some evidence that the factors labeled as significant have an impact on teacher performance even though all statistical estimates are likely biased. The results of this modeling effort are included in this report as a historical record.

Teachers were coded with the dependent variable=1 if their standardized observation score was less than the median value of the sample and if their normalized TVAAS index was less than the median value of the sample. The most parsimonious model estimates (in log odds units) can be found in the Table 4. TPP 1 was used as the TPP reference factor in the analysis because the majority of teachers attended TPP 1. Including the attainment of an advanced degree in the model increased the AIC so it was not included in the final model. Four or more years of experience was used as the experience level reference factor in the analysis because most of the teachers were veteran teachers.

Table 4: Logistic Model Parameters (Low Observation and Low TVAAS)

Parameter	Estimate	Std. Error	z value	p value	Significant
Intercept	1.408	0.603	2.33	0.020	*
Years of Experience (1 or Fewer Years)	1.833	0.262	6.99	0.000	*
Years of Experience (2 or 3 years)	0.824	0.224	3.68	0.000	*
TPP 2	0.190	0.343	0.55	0.579	
TPP 3	-0.248	0.401	-0.62	0.536	
TPP 4	0.485	0.360	1.35	0.178	
TPP 5	-0.051	0.487	-0.11	0.916	
TPP 6	-0.210	0.568	-0.37	0.712	
TPP 7	0.453	0.617	0.73	0.463	
TPP 8	-0.208	0.638	-0.33	0.744	
TPP 9	0.814	0.644	1.26	0.206	
All Other TPPs	-0.120	0.189	-0.63	0.527	
Undergraduate GPA	-0.825	0.191	-4.31	0.000	*

Chi squared statistics indicate that the model is significant ($p=3.0e-10$). The Nagalkerke R^2 for the fit is 0.114. This relatively low R^2 indicates that the independent variables only describe a small portion of the overall variance in teacher performance. Years of experience and undergraduate GPA were significant predictors influencing the probability of having an observation score and a TVAAS index less than the sample median. The TPP attended did not significantly contributed to the prediction. The list below summarizes the significant results from the model.

- Teachers with 1 or fewer years of experience were 6.4 times more likely to have a standardized observation score less than the median and a normalized TVAAS index less than the median when compared to a veteran teacher.
- Teachers with 2 or 3 years of experience were 2.5 times more likely to have a standardized observation score less than the median and a normalized TVAAS index less than the median when compared to a veteran teacher.
- A one point increase in undergraduate GPA means that a teacher is 2 times less likely to have a standardized observation score less than the median and a normalized TVAAS index less than the median.

The output from the logistic model provides the probability that a specific teacher would have a standardized observation score less than the median and a normalized TVAAS index less than the median. A receiver operator characteristic (ROC) curve was plotted using the model output versus the observed data. This was done to determine if the model was a more accurate classifier of teacher performance than the overall frequency of “success” in the sample. The area under the curve was estimated at 66.8%. The 95% confidence interval had an upper bound of 71.8% and a lower bound of 61.8%. The ROC curve provides evidence

that the model is a more accurate classifier of teacher performance because this interval does not include 50%. The cut-off value that maximizes the difference between the true positive identification rate and the false positive identification rate is 27.9%. This cut-off value results in a true positive rate of 60.9% and a false positive rate of 30.8%. Approximately 39% of teachers with low standardized observation scores and low normalized TVAAS scores would not be classified as potential low performing teachers in this model. The data regarding the accuracy of the fit can be found in Table 5. The ROC curve is available in Figure 3.

Table 5: Classification Table (Low Observation and Low TVAAS)

TPP	Observed Success Rate	Success Rate as Modeled	% of Teacher Correctly Classified
TPP 1	30.0%	31.9%	67.0%
TPP 2	28.0%	42.0%	70.0%
TPP 3	25.0%	25.0%	72.5%
TPP 4	33.3%	79.5%	56.4%
TPP 5	16.7%	37.5%	66.7%
TPP 6	33.3%	33.3%	61.1%
TPP 7	28.6%	42.9%	78.6%
TPP 8	28.6%	42.9%	78.6%
TPP 9	45.5%	90.9%	54.5%
All Other TPPs	27.2%	26.0%	71.7%
All TPP	28.8%	33.8%	68.5%

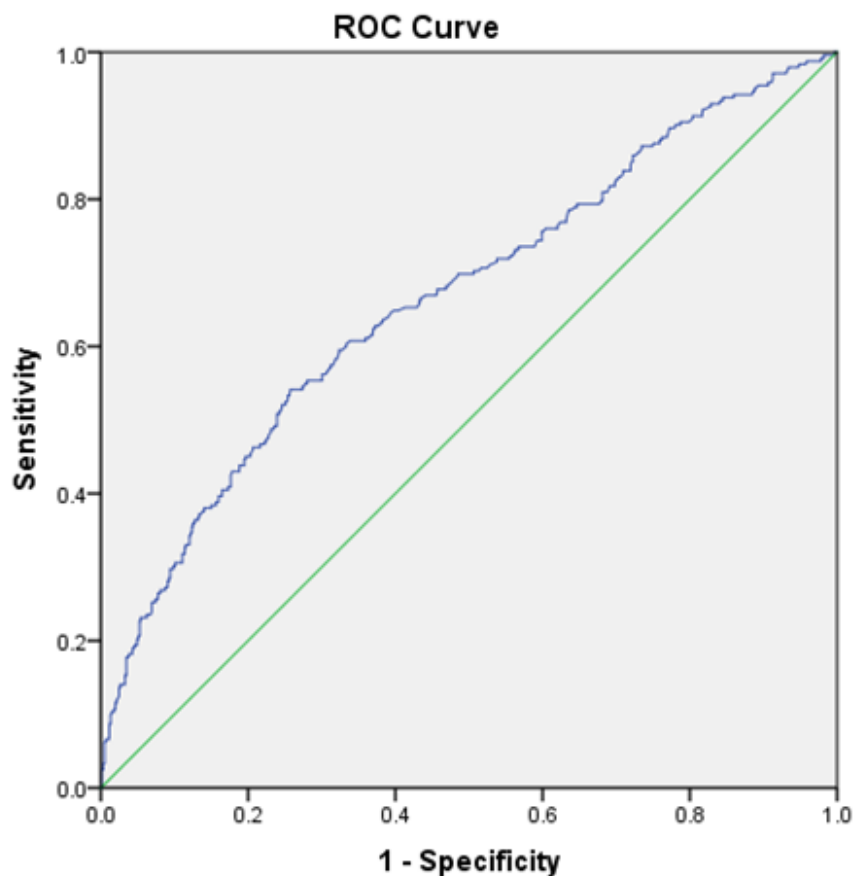


Figure 3: ROC Curve (Low Observation and Low TVAAS)

Conclusions & Considerations

Hiring high-quality candidates to fill teaching vacancies in the Knox County Schools (KCS) will continue to be a key component in achieving the district’s strategic goals. This analysis provides some information that can be leveraged towards better meeting that goal.

Data indicate that the majority of teachers in this analysis came from a relatively small number of teacher preparation programs (TPPs). Slightly more than 50% of the teachers in this sample attended one of three teacher preparation programs, with over 40% of the teachers in the district coming from a single TPP. KCS should continue to find ways to build partnerships with these TPPs to ensure that new hires have the teaching skills that are most valuable to the district. The KCS Human Resource department can also monitor the impact of their recruitment activities. Longitudinal trends in the distribution of TPPs providing teachers to Knox County can provide evidence of the impact of targeted recruitment at each TPP. Additionally, the district may wish to limit recruiting activities at TPPs with low historical retention rates. The district provides significant training and professional development to new employees. It is not cost effective for the district to invest in staff that leave the district shortly after they are hired. Further analysis can be done to determine if low retention rates among certain TPPs occur in more cohorts of teachers. Statistical

modeling uncovered evidence that, among this cohort of teachers, staff hired from TPP 5 were significantly less likely to have both high observation scores and TVAAS index.

There was some evidence in the outcome data that the TPP in which a teacher was trained impacted the probability of a teacher having both high (standardized) observation scores and high (normalized) TVAAS index. The models may be too coarse for general hiring, but may have utility in candidate screening. Typically, large numbers of potential teachers apply for openings in the district. Hiring managers should be relatively confident in screening their initial candidate pools by undergraduate GPA or prior teaching experience. Hiring managers do not need to concern themselves with whether a teacher has attained an advanced degree.

References

Boyd, D.J., Grossman, P.L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher Preparation and Student Achievement. *Education Evaluation and Policy Analysis*, 31(4), 416-440.

Cochran-Smith, M., Stern, R., Sanchez, J.G., Miller, A., Keefe, E.S., Fernandez, M.B., Chang, W. Carney, M. C., Burton, S., & Baker, M., (2016). Holding Teacher Preparation Accountable: A Review of Claims and Evidence. Boulder, CO: National Education Policy Center.

Feuer, M. J., Floden, R. E., Chudowsky, N., Ahn, J. (2013). Evaluation of Teacher Preparation Programs: Purposes, Methods and Policy Options. Washington, D.C.: National Academy of Education.

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.

Knox County Schools (2013), Deepening Our Work: Excellence for Every Child: Five-Year Strategic Plan 2014-2019 (2013), Knoxville, TN: Knox County Office of the Superintendent.

Ronfeldt, M., & Campbell, S.L. (2016). Evaluating Teacher Preparation Using Graduates' Observational Ratings. *Education Evaluation and Policy Analysis*, 38(4), 603-625.

Sattler, C. (2014), An Analysis of Gallup TeacherInsight Scores, Knoxville TN: Knox County Department of Research, Evaluation, and Assessment.