



## Measuring Up: The National Council on Teacher Quality's Ratings of Teacher Preparation Programs and Measures of Teacher Performance

The National Council on Teacher Quality (NCTQ) regularly prepares reports rating teacher preparation programs (TPPs) across the United States on a set of input and process-based standards. The purpose of the research described in this brief is to use data drawn from North Carolina to examine the associations between the NCTQ ratings of TPPs and two measures of teacher performance—teacher value-added scores and teacher evaluation ratings.

In our analysis of NCTQ's overall TPP ratings, we find that in one out of 42 comparisons the graduates of TPPs with higher NCTQ ratings have higher value-added scores than graduates of TPPs with lower ratings; in eight out of 30 comparisons graduates of TPPs with higher NCTQ ratings receive higher evaluation ratings than graduates of TPPs with lower NCTQ ratings. There are no significant negative associations between NCTQ's overall TPP ratings and teacher performance. In our analysis of NCTQ's TPP standards, out of 124 value-added comparisons, 15 of the associations are positive and significant and five are negative and significant; out of 140 teacher evaluation rating comparisons, 31 associations are positive and significant and 23 are negative and significant.

**With our data and analyses, we do not find strong relationships between the performance of TPP graduates and NCTQ's overall program ratings or meeting NCTQ's standards.** Results for several NCTQ standards suggest directions for improving TPP graduates' performance:

1. *Ensuring that TPPs systematically and routinely obtain data on teacher candidate and graduate performance;*
2. *Setting higher standards for admission into TPPs;*
3. *When input and process standards are used to evaluate teacher candidates' performance and the delivery of teacher training, adding measures of quality to those of quantity.*

# Introduction

Disappointment with educational performance in the United States, coupled with research evidence highlighting the importance of teachers to student achievement (Bill and Melinda Gates Foundation, 2013), has given rise to numerous efforts to improve teacher quality. As part of these initiatives, teacher preparation programs (TPPs) have come under increased scrutiny, with researchers identifying the value-added effectiveness of program graduates (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Gansle, Noell, & Burns, 2012; Goldhaber, Liddle, & Theobald, 2013; Koedel, Parsons, Podgursky, & Ehlert, 2012), policymakers promoting higher levels of program accountability, and teacher education accreditation bodies strengthening program requirements (CAEP, 2013). Included in these efforts to improve teacher preparation is the National Council on Teacher Quality (NCTQ) and its *Teacher Prep Review*. For the *Teacher Prep Review*, NCTQ has promulgated 19 standards for TPPs and has undertaken an effort to obtain data from and rate as many TPPs in the United States as possible (NCTQ, 2014). Unlike federal initiatives, such as the Race to the Top funding competition or the U.S. Department of Education's recently announced TPP regulations, which place more of a focus on the effectiveness of TPP graduates as teachers-of-record, NCTQ's standards for TPPs focus on inputs and processes used by TPPs, such as selection criteria, course requirements, content of courses as described in syllabi, supervision of student teaching, and the collection of outcome data for program graduates. To the extent possible, given the nascent research base examining the associations between teacher preparation practices and teacher performance, these NCTQ standards are grounded in available research and are completely transparent (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; NCTQ, 2013; National Research Council, 2010).

NCTQ's standards for TPPs focus on inputs and processes used by TPPs, such as selection criteria, course requirements, content of courses as described in syllabi, supervision of student teaching, and the collection of outcome data for program graduates.

The underlying rationale for the NCTQ standards can be stated as a simple, straightforward, and compelling *if-then* proposition: *if* TPPs meet the input and process standards and receive higher NCTQ ratings, *then* the graduates of

the TPPs will be more effective as classroom teachers than they would have been in a program that does not meet those standards. If this *if-then* proposition holds, a TPP will prepare more effective teachers than it would have otherwise by targeting reform efforts on the processes and inputs identified by NCTQ; conversely, if there is not a strong relationship between NCTQ ratings and teacher performance, TPP reform may be better focused on other practices linked to teacher effectiveness. To date, researchers have not thoroughly investigated the linkage—the predictive validity—between NCTQ overall program scores and NCTQ standard ratings and the performance of TPP graduates as classroom teachers. In this study, we evaluate the predictive validity of the NCTQ ratings with two measures of teacher performance: teachers' value-added to student achievement and teachers' evaluation ratings from their school principals. The evaluation ratings are the annual summative assessments of teachers' performance on the North Carolina Professional Teaching Standards: (1) demonstrating leadership; (2) classroom environment; (3) content knowledge; (4) facilitating student learning; and (5) reflecting on practice. Together, these performance measures present a more comprehensive view of TPP graduates than either outcome measure alone and allow us to determine whether specific NCTQ standards predict closely-related teacher outcome measures (e.g. the NCTQ Elementary Mathematics standard and elementary teacher value-added in mathematics).

## Background

Beginning with small-scale studies in a handful of states and then expanding to TPPs nationwide, NCTQ has undertaken efforts to gather data and rate TPPs based on their processes and inputs. In its most recent report, the *2014 Teacher Prep Review*, NCTQ examined six types of teacher preparation programs—elementary undergraduate, elementary graduate degree, secondary undergraduate, secondary graduate degree, special education undergraduate and special education graduate degree—and provided those TPPs with an overall program score and ratings on the extent to which the TPPs meet up to 19 standards. The NCTQ overall program score ranges from 0 to 125 points and is broken into four categories: (1) Level 1 programs (0-50 points); (2) Level 2 programs (51-66 points); (3) Level 3 programs (67-82 points); and (4) Level 4 programs (83 points or higher). The 19 standards on which NCTQ rates TPPs are as follows: Selection Criteria, Early Reading, English Language Learners, Struggling Readers, Elementary Mathematics, Elementary Content, Middle School Content, High School Content, Content for Special Education, Classroom Management, Lesson Planning, Assessment and Data, Equity, Student Teaching,

Secondary Methods, Instructional Design for Special Education, Outcomes, Evidence of Effectiveness, and Rigor<sup>1</sup>. Because most TPPs declined to turn over all of the necessary course materials that would have allowed NCTQ to rate programs against all of its standards, a TPP's overall rating is determined primarily (80 percent of the score) on the basis of a subset of the standards NCTQ considers most essential to program quality. If data are missing on any of these “essential” standards<sup>2</sup>, NCTQ does not provide the TPP with an overall rating. A TPP can add to its overall score (up to 20 percent) by performing well on non-essential standards. Overall, the *2014 Teacher Prep Review* reported that a majority of TPPs, particularly elementary undergraduate and elementary graduate degree programs, are in the lowest program score category (Level 1) and that less than seven percent of programs earn a program rating in the highest category<sup>3</sup>.

Overall, the *2014 Teacher Prep Review* reported that a majority of TPPs, particularly elementary undergraduate and elementary graduate degree programs, are in the lowest program score category (Level 1) and that less than seven percent of programs earn a program rating in the highest category.

In this study we use data provided by NCTQ, the University of North Carolina General Administration (UNCGA), and the North Carolina Department of Public Instruction to examine the predictive validity of NCTQ's TPP ratings for two teacher performance measures—value-added to student achievement and teacher evaluation ratings. For both of these outcomes we focus on North Carolina public school teachers in the 2011–12 and 2012–13 school years who are in their first or second-year of teaching and who we can link to a TPP rated by NCTQ.

We focus on only the first four program types rated by NCTQ: elementary undergraduate and graduate degree and secondary undergraduate and graduate degree. Our full sample includes 4,513 unique teachers, of whom 66 percent are linked to an elementary (undergraduate or graduate degree) TPP and nearly 80 percent are linked to an in-state (public or private) university.

We estimate value-added models in elementary grades mathematics and reading, middle grades mathematics and reading, and for North Carolina's three End-of-Course exams—algebra I, biology, and English I/II. We estimate evaluation rating models for the five North Carolina Professional Teaching Standards directly assessed by a school administrator. In all of these analyses we enter the NCTQ standard ratings into models individually, control for a rich set of covariates, and cluster standard errors at the teacher level (Koedel, Parsons, Podgursky, & Ehlert, 2012). We estimate models with first-year teachers only and with first and second-year teachers combined (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009).

## Limitations to our analysis

There are several limitations to the available data that should be kept in mind when interpreting our findings. While we do not believe that these limitations are sufficiently serious to undermine our analysis, they do mean that our findings should be interpreted carefully.

First, the data are limited to novice North Carolina public school teachers who graduated from NCTQ-rated TPPs in North Carolina or from NCTQ-rated programs in other states. Teacher performance data from other states may find different associations with NCTQ ratings.

Second, because NCTQ was unable to collect data from TPPs to rate all standards, there are a substantial number of missing NCTQ ratings<sup>4</sup>. Our preferred method to address missing data is multiple imputation. We also examined

<sup>1</sup> The Early Reading, English Language Learners, Struggling Readers, Elementary Mathematics, and Elementary Content standards are only available for elementary-level TPPs; the Middle School Content, High School Content, and Secondary Methods standards are only available for secondary-level TPPs. This report does not include the following standards: Content for Special Education, Equity, Instructional Design for Special Education, Evidence of Effectiveness, and Rigor.

<sup>2</sup> Standards whose ratings are included in the overall score are: (elementary) Selection Criteria, Early Reading, Elementary Mathematics, Elementary Content, and Student Teaching; (secondary) Selection Criteria, Middle School Content, High School Content, and Student Teaching. Standards whose ratings provide additional points that are added to the overall program score are: (elementary) English Language Learners, Struggling Readers, Classroom Management, and Outcomes; (secondary) Secondary Methods, Classroom Management, and Outcomes. Standards whose ratings are not included in the overall program score are: Lesson Planning, Assessment and Data, Evidence of Effectiveness, and Rigor. Equity is not rated.

<sup>3</sup> To view the most recent report, please see [http://www.nctq.org/dmsStage/Teacher\\_Prep\\_Review\\_2014\\_Report](http://www.nctq.org/dmsStage/Teacher_Prep_Review_2014_Report).

<sup>4</sup> For more on NCTQ's data collection methods, timeline, and participation by TPPs please see [http://www.nctq.org/dmsView/TPR\\_2014\\_GM\\_Appendix](http://www.nctq.org/dmsView/TPR_2014_GM_Appendix)

the default option of deleting TPPs from an analysis if the program has any missing data for the variables included in that analysis (case-wise deletion) and replacing missing observations with a zero and adding an indicator variable for observations with any missing data (dummy variable with zero replacement). Neither of these alternative methods for addressing missing data returns findings substantively different from the multiple imputation models except in the case of NCTQ's Classroom Management standard. Therefore, we conclude that using multiple imputation to generate correlation estimates is sound, though we acknowledge that using imputation could raise issues of interpretation. Please see our accompanying technical appendix (Bastian & Henry, 2015) for results from all of these models and further detail on our data sources, study sample, missing data procedures, measures (outcomes and covariates), and analysis methods.

Please see our accompanying technical appendix (Bastian & Henry, 2015) for results from all of these models and further detail on our data sources, study sample, missing data procedures, measures (outcomes and covariates), and analysis methods.

Finally, NCTQ's indicators for some standards may assess aspects of teacher training that are not well-aligned with North Carolina's existing measures for evaluating teacher performance—value-added and evaluation ratings. For example, NCTQ's three standards related to early reading instruction (general knowledge of early reading instruction, teaching reading to English Language Learners, and early reading instruction of struggling readers) are most relevant for teachers assigned to teach grades K-2. However, for our analyses here, we are only able to assess impact on teacher value-added by examining test scores in grades 4 and 5, past the point when most early reading instruction is delivered. Likewise, the NCTQ Classroom Management standard and the North Carolina Classroom Environment evaluation rating standard have similar labels, however, the indicators used by NCTQ to rate TPPs on Classroom Management and by school principals to rate teachers on Classroom Environment have limited overlap.

## What are the Associations Between NCTQ Ratings and Teacher Value-Added?

For the value-added analyses, the left panel of Table 1 displays counts of the total number of statistical tests performed and the number of statistically significant positive and statistically significant negative associations. Our interpretation of the significant associations are as follows: positive (negative) and statistically significant associations indicate that, on average, graduates of TPPs with higher NCTQ ratings are more (less) effective than graduates of TPPs with lower ratings.

We find one positive and significant association out of the 42 tests of the relationship between the NCTQ overall program score—whether graduates of Level 2, Level 3, or Level 4 programs were more or less effective than graduates of Level 1 programs—and teacher value-added; in the remaining 41 tests, the association between the NCTQ overall program score and teacher value-added is non-significant.

For the 14 NCTQ standards for which data were available, we present 124 tests between the NCTQ standard and teacher value-added. Fifteen associations are positive and significant, five are negative and significant, and 104 are non-significant. While not all of the NCTQ standards may be conceptually well-aligned with teaching practices that may promote student achievement gains in tested grades, we note that the NCTQ Elementary Mathematics standard is not significantly associated with elementary grades mathematics value-added and that the NCTQ Secondary Methods standard is not significantly associated with value-added in middle grades or for secondary grades End-of-Course exams. Likewise, there are no significant associations between NCTQ's reading standards—Early Reading, English Language Learners, and Struggling Readers—and elementary grades reading value-added.

For the 14 NCTQ standards for which data were available, we present 124 tests between the NCTQ standard and teacher value-added. Fifteen associations are positive and significant, five are negative and significant, and 104 are non-significant.

## What are the Associations Between NCTQ Ratings and Teacher Evaluation Ratings?

For teacher evaluation rating analyses, the right panel of Table 1 displays counts of the total number of statistical tests performed and the number of statistically significant positive and statistically significant negative associations.

Positive (negative) and statistically significant associations indicate that, on average, graduates of TPPs with higher NCTQ ratings earned higher (lower) evaluation ratings than graduates of TPPs with lower ratings.

In 8 of 30 tests of the association between the NCTQ overall program score—whether graduates of Level 2, Level 3, or Level 4 programs earned higher evaluation ratings than graduates of Level 1 programs—and teacher

Table 1: Summary of NCTQ Value-Added and Evaluation Rating Results

NCTQ Ratings	Associations with Teacher Performance Measures					
	Teacher Value-Added			Teacher Evaluation Ratings		
	Total Tests	Significant Positive Associations	Significant Negative Associations	Total Tests	Significant Positive Associations	Significant Negative Associations
NCTQ Overall Program Rating	42	1	0	30	8	0
Totals with NCTQ Standards	124	15	5	140	31	23
Selection Criteria	14	3	0	10	8	0
Early Reading	4	0	0	10	0	0
English Language Learners	4	0	0	10	8	0
Struggling Readers	4	0	0	10	0	1
Elementary Mathematics	4	0	0	10	0	0
Elementary Content	4	0	1	10	9	0
Middle School Content	4	0	1	10	0	8
High School Content	6	0	0	10	0	1
Classroom Management	14	0	0	10	0	10
Lesson Planning	14	0	0	10	0	3
Assessment and Data	14	4	2	10	0	0
Student Teaching	14	2	1	10	0	0
Secondary Methods	10	0	0	10	6	0
Outcomes	14	6	0	10	0	0

Note: This table displays counts of the total number of statistical tests performed and counts of the statistically significant associations between the NCTQ ratings and measures of teacher performance in North Carolina public school classrooms. For the full set of value-added and evaluation rating coefficients please see the accompanying technical appendix.



evaluation ratings, there are positive and significant associations. The remaining 22 associations are non-significant.

For the predictive validity of NCTQ standards and teachers' evaluation ratings, we present 140 tests and there are 31 positive and significant associations, 23 negative and significant associations, and 86 non-significant results.

For the predictive validity of NCTQ standards and teachers' evaluation ratings, we present 140 tests and there are 31 positive and significant associations, 23 negative and significant associations, and 86 non-significant results. As noted above, not all of the NCTQ standards appear to be conceptually well-aligned with the teaching practices measured by the North Carolina teacher evaluation ratings and therefore, we do not always expect positive and significant associations. However, we note that (1) the NCTQ Elementary Content standard is positively associated with the evaluation rating for Content Knowledge; (2) the NCTQ Middle Content standard is negatively associated with the evaluation rating for Content Knowledge; and (3) the NCTQ Secondary Methods standard is positively associated with the evaluation rating for Facilitating Student Learning.

## An In-Depth Focus on Five NCTQ Standards

Based on conversations with NCTQ representatives, we chose five NCTQ standards for further, *indicator-level* value-added and evaluation rating analyses. These indicators (e.g. for the Outcomes standard, whether a TPP surveys its graduates) are the underlying measures collected by NCTQ to generate standard scores.<sup>5</sup> Results from these models may indicate particular sets of TPP practices that benefit or adversely impact teacher performance. Below, we briefly review the overall and indicator-level results for these five standards. The accompanying technical appendix provides further details on these indicators and the full-set of indicator-level results.

*Outcomes:* The NCTQ Outcomes standard returns the highest number of positive associations with teacher value-added—6 of the 14 tests return positive and significant associations—and no significant associations with teacher evaluation ratings. These results suggest that TPPs that systematically and routinely obtain outcomes data produce more effective graduates. The indicators for the NCTQ Outcomes standard focus on three types of information on TPP graduates (value-added scores, employer surveys, and graduate surveys), the use of teacher candidate performance assessments (e.g. edTPA) and the frequency with which TPPs collect these data. At the indicator level, collecting graduates' value-added data has five positive value-added associations and three positive evaluation rating associations. Curiously, surveys of TPP graduates' employers have negative associations with teachers' evaluation ratings. Surveys of graduates themselves and collecting teacher performance assessment data do not have strong relationships with teacher performance<sup>6</sup>.

Results suggest that TPPs that systematically and routinely obtain outcomes data produce more effective graduates.

*Selection Criteria:* The NCTQ Selection Criteria standard shows consistently positive associations with value-added (3 out of 14 tests) and teacher evaluation ratings (8 out of 10 tests). At the indicator level, a measure for education schools requiring SAT (ACT) scores of 1120 (24) or above is positively associated with teacher evaluation ratings on four out of five of the evaluation standards (excluding the Classroom Environment standard) and in middle grades reading value-added. The results for two other Selection Criteria indicators, TPP setting a minimum GPA requirement and having a minimum GPA requirement above 3.0, are curiously mixed, with four positive and nine negative associations with teacher value-added (out of 26 tests).

*Elementary Mathematics:* The NCTQ Elementary Mathematics standard is not significantly associated with teacher value-added or evaluation ratings in any of our analyses. At the indicator level, a measure for mathematics

<sup>5</sup> Please see [http://nctq.org/dmsView/NCTQ\\_Teacher\\_Prep\\_Review\\_1\\_0\\_Standards](http://nctq.org/dmsView/NCTQ_Teacher_Prep_Review_1_0_Standards) for a listing and description of all the NCTQ indicators.

<sup>6</sup> In another study, we find that locally-scored teacher candidate performance assessments are positively associated with measures of teacher performance (Bastian, Henry, Pan, & Lys, 2014).

methods credit hours is positively associated with teacher value-added in elementary grades reading; a measure for schools being sufficiently selective in admissions is positively associated with teacher evaluation ratings on four out of five of the evaluation standards (all but the Classroom Environment standard). These evaluation rating results mirror those for the Selection Criteria indicator of requiring SAT scores of 1120 or above and suggest that the cognitive skills of teacher candidates may impact their evaluation ratings when they become teachers.

*Classroom Management:* The NCTQ Classroom Management standard is not significantly associated with teacher value-added and is negatively associated with teacher evaluation ratings in all 10 tests<sup>7</sup>. The indicators for this standard measure whether or not TPPs assess their student teachers with regards to nine classroom management related tasks, such as an ability to manage time, materials, and student behaviors (minor or significant) and maintain engagement. Overall, the pattern of associations between the nine classroom management indicators and measures of teacher performance are not substantively revealing, with the possible exception of (1) assessing student teachers' ability to manage materials, which is negatively associated with value-added in three tests and negatively associated with the Classroom Environment evaluation rating and (2) assessing student teachers' ability to manage minor student misbehavior, which is positively associated with teachers' evaluation rating on the Classroom Environment standard. One prior study shows that the actual ratings of student teachers—rather than indicators regarding the particular aspects of student teacher performance TPPs assess—are also uncorrelated with value-added scores (Henry, Campbell, Thompson, Patriarca, Luterbach, Lys, & Covington, 2013). This may suggest that TPPs need to review their practices for assessing student teachers' classroom management.

*Student Teaching:* The NCTQ Student Teaching standard is positively associated with two value-added measures, negatively associated with one value-added measure (out of 14 tests) and not significantly associated with teachers' evaluation ratings. For this standard, NCTQ drew heavily on one of the best prior studies of the associations between TPP characteristics and teacher value-added (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009), however,

the pattern of associations between NCTQ student teaching indicators and teacher performance are mixed. For example: (1) requiring student teaching observations at regular intervals has two positive value-added associations, one negative value-added association, and eight negative evaluation rating associations; and (2) TPP gathering data on student teaching placements has three negative value-added associations and eight positive evaluation rating associations. Finally, there is not a strong relationship between requiring more student teaching observations and teacher performance.

## Discussion

Across value-added and evaluation rating analyses, we present 72 tests of the NCTQ overall program score and find 9 positive and significant associations and 63 non-significant associations. After conducting 264 tests of the NCTQ standards, we find 46 positive and significant associations, 28 negative and significant associations, and 190 non-significant associations. These results should not suggest that teacher preparation is unimportant or that the TPP requirements—selection, content and pedagogy courses, student teaching—rated by NCTQ are unimportant. Rather, these results simply indicate that in our analysis there is not a strong relationship between NCTQ's ratings and meeting their standards and the performance of TPP graduates. As we noted above, there are three important limitations to this overall conclusion: (1) this study is confined to North Carolina only and teacher performance data from other states may find different associations; (2) there are a substantial amount of missing NCTQ standard scores; and (3) not all of the

These results should not suggest that teacher preparation is unimportant or that the TPP requirements—selection, content and pedagogy courses, student teaching—rated by NCTQ are unimportant. Rather, these results simply indicate that in our analysis there is not a strong relationship between NCTQ's ratings and meeting their standards and the performance of TPP graduates.

<sup>7</sup> In case-wise deletion and dummy variable replacement models there is not a strong negative relationship between the NCTQ Classroom Management standard and teacher evaluation ratings. Please see the accompanying technical appendix for the full set of results. As noted in the limitations section, the indicators used by NCTQ to rate TPPs on the Classroom Management standard and the indicators used by school principals to rate teachers on the Classroom Environment standard have limited overlap.

NCTQ standards may be conceptually well-aligned with the teaching practices that may promote higher value-added in tested grades or higher evaluation ratings.

While the majority of NCTQ standards have few, if any, significant associations, results for several standards suggest directions for moving forward. First, results for the Outcomes standard suggest that TPPs that regularly obtain outcomes data may improve their performance relative to their peers. These efforts to obtain outcome data at the program level could be most efficiently facilitated by states, due to the economies of scale from centralized data collection and the ability to compare across TPPs with common data measures. Beyond obtaining such outcomes data, TPPs may also benefit from research that identifies effective methods for using outcomes data for evidence-based TPP reform.

Results for the Outcomes standard suggest that TPPs that regularly obtain outcomes data may improve their performance relative to their peers.

Second, based on the Selection Criteria results and the results for the Elementary Mathematics selectivity indicator, setting higher standards for admission into TPPs may be a promising practice. Specifically, the indicator-level analyses suggest that focusing on SAT/ACT scores will be more beneficial than admissions requirements

based on grade point averages. This finding is in line with the new CAEP requirement for higher TPP admissions standards; future research can examine the impact of these new requirements on entry into TPPs and the performance of TPP graduates. Beyond indicators of cognitive ability, emerging research suggests that TPPs may also benefit from further research on alternative admissions measures, such as personality traits, grittiness, or task persistence (Robertson-Kraft & Duckworth, 2014).

Finally, if TPP inputs and training processes are to be rated using proximal evidences of quality (e.g. course syllabi, course descriptions, student teaching handbooks), such ratings may need to be supplemented with more direct measures of quality. Together, these multiple measures of TPP inputs may have a stronger relationship with measures of teacher performance and may provide TPPs with richer feedback to inform evidence-based program reform.

## Acknowledgements

We wish to thank Alisa Chapman with the University of North Carolina General Administration (UNCGA) for her support and feedback; Arthur McKee, Julie Greenberg, and Rob Rickenbrode with NCTQ for providing access to the NCTQ standards and indicator data as well as providing detailed feedback throughout the development and implementation of the research; and Dan Goldhaber and Cory Koedel, members of NCTQ's technical panel, for technical advice and consultation. The authors acknowledge partial funding for this research was provided by the University of North Carolina General Administration Teacher Quality Research Initiative. No other external funds were used to support this research.



## References

- Bastian, K.C., & Henry, G.T. (2015). Technical appendix: The National Council on Teacher Quality's ratings of teacher preparation programs and measures of teacher performance. Education Policy Initiative at Carolina. Available from: <http://publicpolicy.unc.edu/epic-2/reportspolicy-briefs-working-papers/>
- Bastian, K.C., Henry, G.T. Yi, P. & Lys, D. (2014). Evaluating a pilot of the Teacher Performance Assessment: The construct validity, reliability, and predictive validity of local scores. Education Policy Initiative at Carolina. Available from: <http://publicpolicy.unc.edu/epic-2/reportspolicy-briefs-working-papers/>
- Bill and Melinda Gates Foundation. (2013). Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's Three-Year Study. Available from: [http://www.metproject.org/downloads/MET\\_Ensuring\\_Fair\\_and\\_Reliable\\_Measures\\_Practitioner\\_Brief.pdf](http://www.metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf)
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, 31(4), 416-440.
- Council for the Accreditation of Educator Preparation. (2013). CAEP Accreditation Standards. Available from: [http://caepnet.files.wordpress.com/2013/09/final\\_board\\_approved1.pdf](http://caepnet.files.wordpress.com/2013/09/final_board_approved1.pdf)
- Gansle, K.A., Noell, G.H., & Burns, J.M. (2012). Do student achievement outcomes differ across teacher preparation programs? An analysis of teacher education in Louisiana. *Journal of Teacher Education*, 63(5), 304-317.
- 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(2), 29-44.
- Henry, G.T., Campbell, S.L., Thompson, C.L., Patriarca, L.A., Luterbach, K.J., Lys, D.B., & Covington, V. (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.
- Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2012). Teacher preparation programs and teacher quality: Are there real differences across programs? Available from: [https://economics.missouri.edu/working-papers/2012/WP1204\\_koedel\\_et\\_al.pdf](https://economics.missouri.edu/working-papers/2012/WP1204_koedel_et_al.pdf)
- National Council on Teacher Quality. (2013). *Teacher prep review: A review of the nation's teacher preparation programs*. Available from: [http://www.nctq.org/dmsView/Teacher\\_Prep\\_Review\\_2013\\_Report](http://www.nctq.org/dmsView/Teacher_Prep_Review_2013_Report).
- National Council on Teacher Quality. (2014). *2014 Teacher Prep Review: A review of the nation's teacher preparation programs*. Available from: [http://www.nctq.org/dmsView/Teacher\\_Prep\\_Review\\_2014\\_Report](http://www.nctq.org/dmsView/Teacher_Prep_Review_2014_Report).
- National Research Council. (2010). *Preparing teachers: Building evidence for sound policy*. Washington, D.C.: National Academies Press.
- Robertson-Kraft, C. & Duckworth, A.L. (2014). True grit: Trait-level perseverance and passion for long-term goals predicts effectiveness and retention among novice teachers. *Teachers College Record*, 116(3), 1-27.

---

**Study Authors:** Gary T. Henry and Kevin C. Bastian (May 2015)

*EPIC is an interdisciplinary team that conducts rigorous research and evaluation to inform education policy and practice. We produce evidence to guide data-driven decision-making using qualitative and quantitative methodologies tailored to the target audience. By serving multiple stakeholders, including policy-makers, administrators in districts and institutions of higher education, and program implementers we strengthen the growing body of research on what works and in which context. Our work is ultimately driven by a vision of high quality and equitable education experiences for all students, and particularly students in North Carolina.*

<http://publicpolicy.unc.edu/epic-home/>

---



**UNC**  
COLLEGE OF  
ARTS & SCIENCES  
PUBLIC POLICY