



STATE REQUIREMENTS FOR TEACHER EVALUATION POLICIES PROMOTED BY RACE TO THE TOP

Congress appropriated approximately \$5.05 billion for the Race to the Top (RTT) program between 2009 and 2012, including approximately \$4.35 billion through the American Recovery and Reinvestment Act of 2009. The RTT initiative encouraged states to implement education policies in six core areas, including teacher evaluations. Evaluations are an important strategy to assess teacher quality, and there is growing consensus about the need for evaluation systems that could yield higher-quality information to improve teacher performance. An increasing body of evidence also suggests that some of the teacher evaluation policies promoted by RTT, such as using multiple measures and multiple rating categories, could help to produce more valid and reliable estimates of teacher quality.^{1,2} Given the substantial federal investment in RTT to promote certain teacher evaluation policies, it is important to learn about the policies that states are actually requiring. Many states have started to alter their policies, but knowledge of their progress remains limited.

This brief describes the extent to which states required teacher evaluation policies aligned with the RTT initiative as of spring 2012. Although teacher evaluation policies appear to be rapidly evolving, documenting policy requirements in the early years of RTT implementation can help inform policymakers about the pace of policy innovation nationally. This brief examines the presence of state-level requirements for certain practices but not the actual district- or school-level implementation of such practices. Key findings, based on interviews with administrators from 49 states and the District of Columbia (12 Round 1 and 2 RTT states, 7 Round 3 RTT states, and 31 non-RTT states), include the following:

- *States, on average, reported requiring less than half of eight teacher evaluation policies aligned with RTT priorities, although the number of policies required by RTT states was higher than non-RTT states (3.7 policies for Round 1 and 2 RTT states, 3.6 for Round 3 RTT states, and 2.2 for non-RTT states).*
- *States' reported teacher evaluation policies were most aligned with RTT priorities focused on using multiple measures to evaluate teacher performance (30 states); using multiple rating categories to classify teacher performance (31 states); and conducting annual evaluations (25 states).*
- *States' reported teacher evaluation policies were least aligned with RTT priorities focused on using evaluation results to inform decisions regarding career advancement (1 state) and compensation (6 states for annual salary increases, and 5 states for performance-based compensation).*

In recent years, federal and state agencies, foundations, and other education organizations have increased their focus on how teachers are evaluated. Federal initiatives such as Race to the Top (RTT) and the School Improvement Grants program, as well as conditions articulated by the U.S. Department of Education for state waivers from *No Child Left Behind* (NCLB), and privately-funded investments such as the Measures of Effective Teaching project funded by the Bill & Melinda Gates Foundation have all promoted similar changes to teacher evaluation policies.

Congress appropriated approximately \$5.05 billion for the RTT program between 2009 and 2012, making it the largest competitive grant program ever administered by the U.S. Department of Education. RTT competitively awarded grants to state education agencies, ranging from \$75 million to \$700 million per grantee in the first two rounds of awards and from \$17 million to \$43 million per grantee in the third round. The RTT grant application laid out guidance in six areas, including the criteria by which states' plans would be evaluated.³ For the teacher evaluation area, these criteria included whether states proposed to:⁴

1. Establish clear approaches to measuring student achievement growth for individual students.
2. Design and implement rigorous, transparent, and fair evaluation systems for teachers.
3. Differentiate effectiveness using multiple rating categories that take student achievement growth into account as a significant factor and are designed with teacher involvement.
4. Conduct annual evaluations that include timely and constructive feedback and provide teachers with data on student achievement growth for their students, classes, and schools.
5. Use evaluations to inform decisions about staff development, compensation, promotion, tenure, certification, and removal of ineffective teachers.

A growing body of evidence suggests that at least some of the teacher evaluation policies promoted by RTT could help to produce more valid and reliable estimates of teacher quality. For example, evidence suggests that evaluations using more than one measure can reliably capture teacher performance,^{5,6} and that using more than two rating categories can better differentiate teacher performance.^{1,7} The goal of these types of policies is to produce higher-quality information to help improve teacher performance and thus student achievement.

While many states have started to change their teacher evaluation policies, understanding the progress in this area remains limited. RTT Annual Performance Reports provide information about progress in individual states that received RTT grants,⁸ and a recent Government Accountability Office report examined the progress that Round 1 and 2 RTT states had made in implementing teacher and principal evaluations systems as of the 2012-2013 school year.⁹ An October 2013 National Council on Teacher Quality report focused on the progress made by all 50 states and the District of Columbia in adopting teacher evaluation policies and in using that information to inform teacher policy and classroom practice.¹⁰ This brief is the first to present data on state teacher evaluation policies aligned with RTT application criteria, for states that received RTT grants and those that did not.

This brief describes the extent to which RTT and non-RTT states reported requiring teacher evaluation policies aligned with RTT priorities as of spring 2012. After presenting results of an analysis of state requirements for eight teacher evaluation policies aligned with the RTT application criteria, it focuses on areas where states' reported teacher evaluation policies were most and least aligned with RTT priorities (where most and least aligned are defined as the areas in which the largest and smallest percentages of states reported requiring these policies). The findings do not prove that receiving RTT grants *caused* any differences, as RTT states might have initiated changes even without the grants. They simply summarize the teacher evaluation policies reported by all of the RTT states and all but one of the non-RTT states.

Data

The data in this brief come from structured interviews conducted in spring 2012 with administrators in 49 states and the District of Columbia.¹¹ We distinguish between RTT Rounds 1 or 2 and Round 3 to account for differences in the *timing, levels of funding, and scope* between these groups of RTT states.¹²

Round 1 awards were made in March 2010, Round 2 awards in August 2010, and Round 3 awards in December 2011. When the data for this brief were collected in spring 2012, Round 1 and 2 grantees had been implementing their RTT plans substantially longer than Round 3 grantees. Therefore, one might expect differing progress from Round 3 grantees, who had their grants for less than six months at the time of our spring 2012 interviews.

Funding also differed between rounds. The awards to the 12 grantees in the first two rounds ranged from \$75 million to \$700 million, while awards to the seven grantees in the third round ranged from \$17 million to \$43 million.¹³ These different funding levels are another reason one might expect differing progress from Round 3 grantees.

Finally, the scope of the Round 3 grants differed from that of the earlier rounds. In the Round 3 application, the U.S. Department of Education acknowledged the smaller grant size and asked applicants to focus on a subset of policy areas. In contrast, the grantees from Rounds 1 and 2 were asked to implement policies in all six areas listed in the application.¹⁴ As a result, grantees from Round 3 were not required to implement policies in the teacher evaluation area; nevertheless, five of the seven grantees planned to do so, according to their applications.

We refer to Round 1 and 2 grantees as *early RTT states* and Round 3 grantees as *later RTT states* in this brief. We refer to *states* for simplicity when discussing RTT grantees while recognizing that one early RTT grantee is the District of Columbia. Throughout this brief, we use the phrase “states reported” as a concise method of conveying what the state *administrators* reported during interviews.

Given that the grants were competitively awarded, differences between RTT (early and later) and non-RTT states might be expected in general. Table 1 shows that the groups did differ on several characteristics in 2010-2011. On average, early and later RTT states served higher percentages of black students and lower percentages of Hispanic students than non-RTT states. Early RTT states also had a higher average percentage of students eligible for free or reduced-price lunch than non-RTT states. Early and later RTT states had a higher average percentage of schools eligible for Title I funds than non-RTT states. Readers should keep this context in mind when interpreting the comparisons in this brief, as receipt of an RTT grant is not the only difference between the groups of states.

Table 1. State Characteristics by RTT Status, 2010-2011

| | Early RTT | Later RTT | Non-RTT |
|---|-----------|-----------|-----------|
| Percentage of Students, by Race/Ethnicity | | | |
| White, non-Hispanic | 53.2 | 57.5 | 55.2 |
| Black, non-Hispanic | 22.9 | 15.7 | 12.1 |
| Hispanic | 15.8 | 19.7 | 22.4 |
| Asian | 4.4 | 3.9 | 5.3 |
| Other | 3.6 | 3.0 | 4.8 |
| Percentage of Students Eligible for Free or Reduced-Price Lunch | 49.2 | 43.7 | 48.1 |
| Percentage of Schools Eligible for Title I | 77.5 | 75.1 | 68.2 |
| Number of States, by Region | | | |
| Northeast | 3 | 2 | 4 |
| Midwest | 1 | 1 | 10 |
| South | 7 | 2 | 7 |
| West | 1 | 2 | 10 |
| Number of States | 12 | 7 | 31 |

Source: 2010-2011 Common Core of Data for 49 states and the District of Columbia.

Note: For each group of states (early RTT, later RTT, and non-RTT), the cells in the table report the percentage of students by race/ethnicity and eligibility for free or reduced-price lunch, the percentage of schools eligible for Title I, and the number of states from each region of the United States. Student race/ethnicity and free or reduced-price lunch eligibility are weighted by average student enrollment, while school Title I eligibility is weighted by the total number of schools with a nonmissing value for Title I status. Early RTT states include Round 1 and 2 grantees; Later RTT states include Round 3 grantees. Texas was excluded from all analyses because it did not participate in the data collection for this brief.

Methods

In this section of the brief, we describe the method used to analyze the extent to which RTT and non-RTT states reported requiring teacher evaluation policies aligned with RTT priorities as of spring 2012. The data used for this analysis come from closed-ended questions asked as part of the structured state administrator interview in spring 2012. The protocols for the interviews were originally designed to cover all six RTT policy areas, not just teacher evaluation policies. When the instruments were designed in 2010, the interview questions were informed by the policies encouraged in the RTT application but did not necessarily cover all of the criteria in all six areas. There were two reasons for this. First, the study team found that it was difficult to create objective, easy-to-measure interview questions for some RTT application criteria. Second, the study team faced constraints on the interview length. To reduce the length, the study team asked staff from the RTT program office at the U.S. Department of Education to provide input about which questions were of greatest interest at that time, some of which were in areas other than teacher evaluation policies.

To select questions that aligned with the teacher evaluation policies promoted by RTT, the study team reviewed the explicit guidance about RTT to states in Section D(2) of the RTT application and in the RTT Program Guidance and Frequently Asked Questions document.¹⁵ The team then reviewed the spring 2012 state administrator interview protocol and identified questions about states' requirements for teacher evaluation policies that were aligned with Section D(2).¹⁶ Table A.1 in the Appendix lists the interview questions addressing each teacher evaluation policy aligned with the RTT application criteria. To avoid overweighting particular interview questions, each question was aligned with only one policy. Exhibit 1 lists the criteria in Section D(2) and the eight teacher evaluation policies from the state interview protocol that aligned with these criteria. The eight policies cover four of the five criteria listed in the RTT application and on page 2 of this brief.¹⁷

Exhibit 1. RTT Application Criteria (*in italics*), with Aligned Teacher Evaluation Policies Listed Below from State Interview Protocol

Design and implement rigorous, transparent, and fair evaluation systems for teachers.

1. State required multiple measures of performance to evaluate teachers in tested and nontested grades and subjects.

Differentiate effectiveness using multiple rating categories that take student achievement growth into account as a significant factor and are designed with teacher involvement.

2. State required a minimum number of rating levels (two) be used to classify teacher performance.
3. State required districts to use student achievement growth to evaluate some or all teachers and specified the extent to which it must factor into teacher evaluations.

Conduct annual evaluations that include timely and constructive feedback and provide teachers with data on student achievement growth for their students, classes, and schools.

4. State required that evaluations of all teachers (probationary and nonprobationary) take place at least annually.

Use evaluations to inform decisions about staff development, compensation, promotion, tenure, certification, and removal of ineffective teachers.

5. State required that teacher evaluation results be used to guide decisions about professional development.
6. State required that teacher evaluation results be used to guide decisions about dismissal.
7. State required that teacher evaluation results be used to guide decisions about compensation (including annual salary increases or performance-based compensation).
8. State required that teacher evaluation results be used to guide decisions about career advancement.

Next, the study team constructed eight variables, each corresponding to one of these eight policies and each ranging from 0 to 1. For each variable, a value of 1 indicated that the state reported requiring all of that teacher evaluation policy, a value of 0 indicated that the state reported requiring none of that teacher evaluation policy, and a value between those two limits indicated that the state reported requiring some of that teacher evaluation policy. For items one and seven in Exhibit 1, it was possible for a state to be categorized as having reported requiring some of these particular teacher evaluation policies.¹⁸

For four items in Exhibit 1, the study team was unable to find explicit guidance in the RTT application or corresponding materials to guide particular aspects of variable creation. The team addressed each case as follows:

1. RTT materials did not specify a minimum number of measures required for evaluating the performance of teachers of tested and nontested grades and subjects, so for item one in Exhibit 1 the study team allocated more points to states as the number of reported measures for evaluating teachers of tested and nontested grades and subjects increased.
2. RTT materials did not specify the number of recommended rating categories, so for item two in Exhibit 1 the study team defined multiple rating categories as two or more categories.¹⁹
3. RTT materials did not specify the weight at which student achievement growth would be considered a “significant” portion of the teacher evaluation, so for item three in Exhibit 1 the study team did not require a particular weight for student growth.²⁰
4. RTT materials did not specify whether annual evaluations should be conducted for all teachers or only for certain subgroups of teachers, so for item four in Exhibit 1 the study team interpreted the RTT guidance to mean that the requirement to evaluate teachers annually applied to *all* teachers.²¹

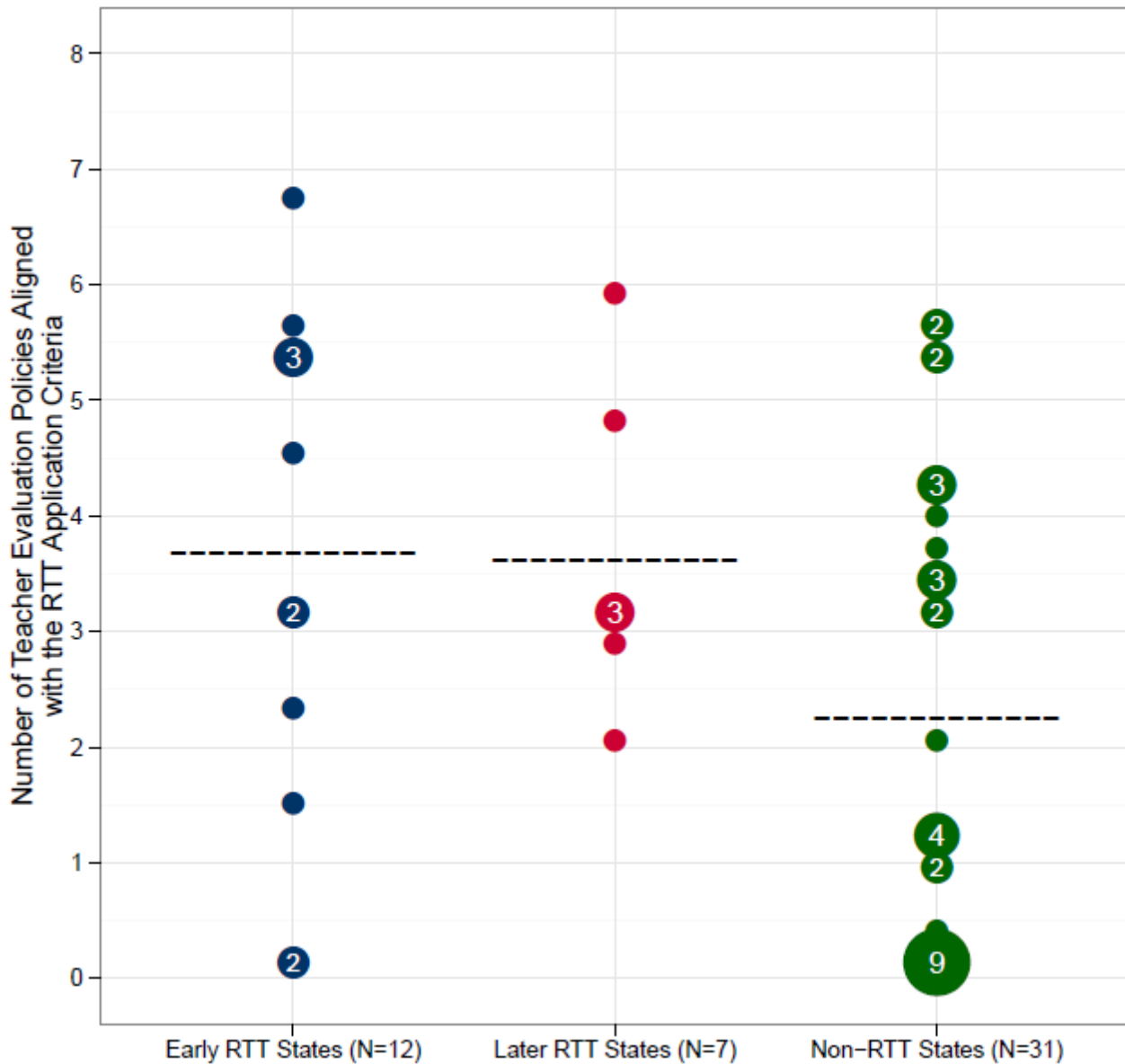
Finally, to determine each state's overall requirements for teacher evaluation policies aligned with the criteria in Section D(2) of the RTT application, the study team added up the number of points the state received across all eight teacher evaluation policies (where a state could receive between 0 and 1 point for each policy).²² Each state thus received a total of between 0 and 8 points, with the possibility of fractional points. Then, for each group of states (early RTT, later RTT, and non-RTT), we averaged their total points. The resulting number for each group is interpreted as the average number of teacher evaluation policies that group of states reported requiring, out of eight teacher evaluation policies aligned with the RTT application criteria in the area of effective teachers.

On Average, RTT and Non-RTT States Reported Requiring Less than Half of Eight Teacher Evaluation Policies Aligned with RTT Priorities, Although the Number of Policies Required by RTT States Was Higher Than Non-RTT States

Early RTT and later RTT states reported requiring an average of 3.7 and 3.6 teacher evaluation policies, compared to 2.2 policies for non-RTT states. However, the number of policies required within each group of states varied. Early RTT states reported requiring between 0.1 and 6.8 teacher evaluation policies aligned with the RTT application criteria, later RTT states reported requiring between 2.1 and 5.9, and non-RTT states reported requiring between 0 and 5.6 (Figure 1).²³

An examination of the extent to which states reported requiring the *individual* policies aligned with the RTT application criteria can suggest policy areas that states have and have not addressed. Table 2 displays the number of states that reported requiring each of the teacher evaluation policies aligned with the RTT application criteria. The next two sections of the brief discuss five RTT-promoted policies that states most and least frequently reported requiring.

Figure 1. Number of Teacher Evaluation Policies Aligned with the RTT Application Criteria in Spring 2012, by State



Source: Interviews with state administrators from 49 states and the District of Columbia in spring 2012.

Note: Each dot in this figure represents the number of states that reported requiring a particular number of teacher evaluation policies (of eight examined) that were aligned with the RTT application criteria. The number inside each dot is the number of states represented by the dot; dots that represent only one state have no number inside. The policies summarized in this figure are presented in Exhibit 1. In all but two cases, a yes response received one point. In the other two cases, it was possible for a state to receive a fraction of one point. See the Methods section for details on the way in which the number of policies was determined for each state. The dashed line denotes the average number of policies required by each group of states.

Table 2. Number of States Requiring Individual Teacher Evaluation Policies Aligned with the RTT Application, by RTT Criteria

| | Number of States that Reported Requiring | | | |
|---|--|----------------|-----------|----------------|
| | All States | Early RTT | Later RTT | Non-RTT |
| Design and Implement Rigorous, Transparent, and Fair Evaluation Systems for Teachers | | | | |
| Two or more performance measures to evaluate teachers in both tested and nontested grades and subjects ^a | 30 | 9 | 5 | 16 |
| Differentiate Effectiveness Using Multiple Rating Categories that Take Student Achievement Growth Into Account as a Significant Factor and Are Designed With Teacher Involvement | | | | |
| A minimum number of rating levels (two) be used to classify teacher performance | 31 | 8 | 6 | 17 |
| Districts to use student achievement growth to evaluate some or all teachers and specifying the extent to which it must factor into teacher evaluations ^b | | | | |
| Districts to use student achievement growth to evaluate some or all teachers | 29 | 8 | 5 | 16 |
| Specifying the extent to which it must factor into teacher evaluations | 21 | 7 | 4 | 10 |
| Conduct Annual Evaluations That Include Timely and Constructive Feedback and Provide Teachers With Data on Student Achievement Growth for their Students, Classes, and Schools | | | | |
| Evaluations of all teachers (probationary and nonprobationary) take place at least annually | 25 | 8 | 2 | 15 |
| Use Evaluations to Inform Decisions About Staff Development, Compensation, Promotion, Tenure, Certification, and Removal of Ineffective Teachers | | | | |
| Teacher evaluation results be used to guide decisions about: | | | | |
| Professional development | 20 | 7 | 4 | 9 |
| Dismissal | 21 | 7 | 6 | 8 |
| Compensation ^c | | | | |
| Annual salary increases | 6 | 2 | 1 | 3 |
| Performance-based compensation | 5 | 1 | 1 | 3 |
| Career advancement opportunities | 1 | 1 | 0 | 0 |
| Number of States^d | 47 - 50 | 11 - 12 | 7 | 29 - 31 |

Source: Interviews with state administrators from 49 states and the District of Columbia in spring 2012.

^a This construct differs from item one in Exhibit 1. Because RTT guidance did not specify a minimum number of measures required for evaluating the performance of teachers of tested and nontested grades and subjects, the study team allocated more points to states for item one in Exhibit 1 as the number of reported measures for evaluating teachers of tested and nontested grades and subjects increased. For ease of presentation of this individual policy in this table, we present the number of states that required two or more performance measures to evaluate teachers in both tested and nontested grades and subjects.

^b As described in the Methods section, for this policy, states received one point if they responded yes to two questions about: (1) requiring student achievement growth as an evaluation measure for some or all teachers, and (2) whether state regulations specified the extent to which student achievement growth must factor into evaluations. States that required student achievement growth as an evaluation measure for some or all teachers, but did not specify the extent to which it must factor into evaluations, received zero points. Table 2 presents the number of states that responded yes to each question separately.

^c As described in the Methods section, this construct was based on states' responses to two questions (one about annual salary increases and the other about performance-based compensation). Table 2 presents the number of states that responded yes to each question separately.

^d Sample sizes refer to the number of states used in the analysis. A range is provided when the sample sizes varied across items in the table due to nonresponse.

States’ Reported Teacher Evaluation Policies Were Most Aligned with RTT Priorities Focused on Using Multiple Measures to Evaluate Teacher Performance, Using Multiple Rating Categories, and Conducting Annual Evaluations

About two-thirds of all states (30 states) reported requiring two or more measures to evaluate teachers in both tested and nontested grades and subjects. Across the three groups of states, 9 early RTT states (75 percent), 5 later RTT states (71 percent), and 16 non-RTT states (52 percent) reported requiring two or more measures to evaluate teachers in both tested and nontested grades and subjects (Table 2). As of spring 2012, states reported requiring that an average of 2.2 measures be used to evaluate teachers in *tested* grades and subjects. For this group of teachers, early RTT states reported, on average, requiring 2.8 measures, later RTT states reported requiring 2.6 measures, and non-RTT states reported requiring 1.9 measures. Similar requirements were reported for teachers in *nontested* grades and subjects (an average of 2.0 measures across all states, and 2.2, 2.6, and 1.8 measures for the three groups of states, not shown).

Information about the combination of measures that states reported requiring for teachers in tested and nontested grades and subjects provides more detail on the way in which states are implementing multiple teacher evaluation measures. Table 3 presents the combinations of measures that states reported requiring.

Table 3. Combination of Measures Required to Evaluate Teachers, Spring 2012

| | Number of States that Reported Requiring | | | |
|--|--|-----------|-----------|-----------|
| | All States | Early RTT | Later RTT | Non-RTT |
| For Teachers in <i>Tested</i> Grades and Subjects | | | | |
| Student achievement growth, classroom observations, and at least one other measure | 13 | 5 | 1 | 7 |
| Student achievement growth and classroom observations only | 13 | 3 | 4 | 6 |
| Classroom observations and another measure besides student achievement growth | 4 | 1 | 0 | 3 |
| Only one measure | 13 | 3 | 2 | 8 |
| No required measures | 7 | 0 | 0 | 7 |
| For Teachers in <i>Nontested</i> Grades and Subjects | | | | |
| Student achievement growth, classroom observations, and at least one other measure | 9 | 2 | 1 | 6 |
| Student achievement growth and classroom observations only | 15 | 5 | 4 | 6 |
| Classroom observations and another measure besides student achievement growth | 6 | 2 | 0 | 4 |
| Only one measure | 13 | 3 | 2 | 8 |
| No required measures | 7 | 0 | 0 | 7 |
| Number of States | 50 | 12 | 7 | 31 |

Source: Interviews with state administrators from 49 states and the District of Columbia in spring 2012.

Note: For teachers in *tested* and *nontested* grades and subjects, interviewers asked state administrators about whether any of the following other measures besides student achievement growth and classroom observations were required by the state for teacher evaluations: self-assessments, portfolios or other artifacts of teacher practice, peer assessments other than classroom observations, student work samples, student surveys or other feedback, and parent surveys or other feedback.

Three patterns for combining multiple measures to evaluate teachers emerged:

1. ***States did not require the same combination of multiple measures.*** States that required two or more measures to evaluate teachers employed one of three distinct combinations: (1) student achievement growth, classroom observations, and at least one other measure, such as self-assessments, teaching portfolios, peer assessments, student work samples, and student or parent surveys; (2) student achievement growth and classroom observations; or (3) classroom observations and a measure besides student achievement growth. Even within groups of states, the combinations of measures varied. For example, among the nine early RTT states that required two or more measures to evaluate teachers in tested grades and subjects, five (56 percent) required a combination of student achievement growth, classroom observations, and at least one other measure, while three (33 percent) required student achievement growth and classroom observations only, and one (11 percent) required classroom observations and a measure besides student achievement growth.
2. ***Classroom observations were a key measure required by many states for teacher evaluations.*** All the combinations of two or more measures included classroom observations. To a somewhat lesser extent, the combinations of measures also included student achievement growth. Other measures, such as student or parent surveys or other feedback or student work samples, were less frequently required by states in combination with student achievement growth or classroom observations.
3. ***Many states required similar combinations of measures for evaluating teachers in tested and nontested grades and subjects.*** For example, among the 30 states that required two or more measures to evaluate teachers, 26 required student achievement growth to be one of the measures for teachers in *tested* grades and subjects, while 24 similarly required it for teachers in *nontested* grades and subjects. States typically use state-wide assessments required by NCLB in math and reading for grades 3 to 8 to generate student achievement growth scores for individual teachers in *tested* grades and subjects, but employing this measure to directly attribute student achievement growth to individual teachers in *nontested* grades and subjects is not possible in many states. Although we did not ask directly in our interviews about the student achievement measures states might be using for teachers in *nontested* grades and subjects, some state administrators volunteered that they required Student Learning Objectives or school-wide measures (including state- or district-wide assessments) for evaluating teachers in *nontested* grades and subjects. These reports are consistent with findings from a recent literature review on alternative measures for these types of teachers.²⁴

About two-thirds of all states (31 states) reported requiring two or more rating categories to classify teacher performance. Across the three groups of states, 8 early RTT states (67 percent), 6 later RTT states (86 percent), and 17 non-RTT states (55 percent) reported requiring two or more rating categories (Table 2). We also examined how many states reported requiring at least three rating categories. As of spring 2012, 8 early RTT states (67 percent), 3 later RTT states (43 percent), and 12 non-RTT states (39 percent) reported requiring a minimum of three or more rating categories (not shown).

About half of all states (25 states) reported requiring that evaluations of all teachers (probationary and nonprobationary) take place at least annually. Across the three groups of states, 8 early RTT states (67 percent), 2 later RTT states (29 percent), and 15 non-RTT states (52 percent) reported requiring evaluations of all teachers (probationary and nonprobationary) to occur at least annually (Table 2).

State requirements for frequency of evaluations could differ depending on the probationary status of the teacher. For example, states might require teachers in their probationary period (i.e., early in their

career) to be evaluated more frequently as a way to ensure that those teachers receive regular feedback about their performance. Evidence suggests that this was the case. As of spring 2012, among the 44 states that reported having a probationary period for teachers, 41 states reported requiring evaluations of *probationary* teachers at least annually.²⁵ Across all 49 states and the District of Columbia,²⁶ 26 states reported requiring at least annual evaluations for *nonprobationary* teachers (not shown).

States' Reported Teacher Evaluation Policies Were Least Aligned with RTT Priorities Focused on Using Evaluation Results to Inform Decisions About Compensation and Career Advancement

Across all states, few states reported requiring that teacher evaluation results be used to guide decisions about annual salary increases (6 states) or performance-based compensation (5 states). Of the six states that reported requiring that teacher evaluation results be used to guide decisions about annual salary increases, RTT states comprised half of these states (Table 2). Of the five states that reported requiring teacher evaluation results be used to inform performance-based compensation, two were RTT states.

Only one state reported requiring that teacher evaluation results be used to guide decisions about career advancement. The state with this requirement was an early RTT state (Table 2). However, RTT states reported using teacher evaluation results in other ways. For example, 7 early RTT states (58 percent) and 4 later RTT states (57 percent) reported requiring the use of evaluation results to guide professional development and support for individual teachers, compared to 9 non-RTT states (29 percent, Table 2). In addition, 7 early RTT states (58 percent) and 6 later RTT states (86 percent) reported requiring the use of evaluation results to guide dismissal decisions, compared to 8 non-RTT states (26 percent).

Discussion

On average, RTT and non-RTT states reported requiring less than half of eight teacher evaluation policies aligned with RTT priorities, although the number of policies required by states differed by RTT status, with RTT states requiring about half of the policies and non-RTT states requiring about a quarter. States' reported teacher evaluation policies were most aligned with RTT priorities focused on using multiple measures to evaluate teacher performance, using multiple rating categories to classify teacher performance, and conducting annual evaluations. States' reported teacher evaluation policies were least aligned with RTT priorities focused on using evaluation results to inform decisions regarding compensation and career advancement.

As noted in Exhibit 1, this brief only addresses four of the five teacher evaluation criteria laid out in Section D(2) of the RTT application. While the policies examined in this brief represent a subset of the teacher evaluation policies encouraged by RTT, our findings provide information to help gauge states' progress during the early years of RTT with implementing important teacher evaluation policies and to identify areas that states have not yet addressed.

Why hadn't all of the RTT states required all eight policies aligned with the RTT application criteria? There are several potential explanations. These findings could reflect the fact that policies encouraged by RTT are time-consuming to implement. For example, states often had to first engage in discussions with teachers' unions before they could implement changes to teacher evaluation systems.^{27,28} These discussions can take time due to the contentious nature of the proposed changes. A 2013 GAO report and evidence from data collected for this study is consistent with this explanation.²⁹ The GAO report found that 11 of 12 early RTT states reported that it was a challenge to address teacher concerns about changes to teacher evaluation systems. Our study found that on average RTT states with collective bargaining agreements reported requiring a lower number of teacher evaluation policies aligned with the

RTT application criteria than RTT states without collective bargaining agreements (3.0 policies vs. 5.0 policies).³⁰ If this explanation is correct and RTT states are eventually able to overcome any challenges with their teachers, we might expect to observe RTT states requiring more of the policies over time.

Alternatively, it is possible that states chose not to require a specific set of teacher evaluation policies statewide, but worked directly with participating local education agencies to require the policies promoted by RTT. Evidence from data collected for this study provides some support for this explanation. For example, five early RTT states reported requiring that districts administer teacher evaluations that comply with state statutes (as opposed to requiring that districts use a particular evaluation model prescribed by the state).

When interviews for this study were conducted in spring 2012, RTT states were no more than halfway through their four-year grant period, so the findings also could reflect RTT states' partial progress earlier in the grant. In particular, 7 of the 19 RTT states received their grants in the third round of awards (in December 2011), so may still have been in the very early stages of funding and requiring policies promoted by RTT. Further, because of the smaller grant sizes for Round 3, later RTT states focused their applications on a subset of policy areas, with two not including plans to implement policies in the teacher evaluation area.

Why did RTT states require more policies aligned with the RTT application criteria than non-RTT states? There are several potential explanations. This finding could reflect the effect of the RTT grants on state policies. RTT aimed to increase the extent to which states required these policies, so this pattern of findings could reflect RTT states' progress toward that goal. Alternatively, this finding could reflect differences between the groups of states in the number of these policies already required *prior* to the awarding of RTT grants. For example, before applying for the third round of RTT grants, some later RTT states may have changed some teacher evaluation policies in an effort to strengthen their Round 3 applications. More broadly, it may not be surprising to find such pre-RTT differences in some cases because the RTT application criteria explicitly awarded additional points to states that had already required some policies prior to applying. However, data from our interviews indicate that there were no differences between the groups of states prior to the awarding of RTT grants for four of the eight teacher evaluation policies that we examined (data from the period prior to RTT awards were not collected for the other four policies). This finding could also reflect underlying differences between the groups of states. As shown in Table 1, RTT states differed from non-RTT states in the geographic regions, types of schools, and students served. These differences could potentially signal differences between the groups of states in their propensity to require these policies. If so, we might expect the RTT states to require more policies aligned with the RTT application criteria than non-RTT states even if the RTT states had not received RTT grants.

Although we cannot definitively reject or accept these potential explanations, we offer them as starting points for future investigations on this topic. Because the process of changing teacher evaluation policies can be complex and require substantial time to implement, we conducted a second round of interviews with states in spring 2013. Future reports will present the results from those interviews, which will reflect one additional year of RTT implementation.

APPENDIX

Table A.1 presents the state interview questions that address each of the eight teacher evaluation policies aligned with the RTT application criteria. Table A.2 presents sensitivity checks for the findings presented in the brief, examining how findings would change if two constructs for policies aligned with RTT application criteria were coded differently.

Table A.1. State Interview Questions Addressing the Teacher Evaluation (TL) Policies Aligned with the RTT Application Criteria

| Policy | Interview Questions Addressing the Teacher Evaluation Policy |
|--|---|
| <p>Requiring multiple measures of performance to evaluate teachers in tested and nontested grades and subjects</p> | <p>TL36. Currently, does the state require that districts use student growth as one measure when evaluating all teachers, some teachers, or no teachers?</p> <p>TL42. Other than the student growth performance measure we just discussed, which of the following other measures of teacher performance are currently required by the state for teacher evaluations for teachers in tested grades and/or subjects? a. classroom observations conducted by the principal; b. classroom observations conducted by a school administrator other than the principal; c. classroom observations conducted by someone other than a school administrator, such as a peer or mentor teacher; d. self-assessment; e. portfolios or other artifacts of teacher practice; f. peer assessments other than classroom observations; g. student work samples; h. student surveys or other feedback; i. parent surveys or other feedback</p> <p>TL43. Other than the student growth performance measure we just discussed, which of the following other measures of teacher performance are currently required by the state for teacher evaluations for teachers in nontested grades and/or subjects? a. classroom observations conducted by the principal; b. classroom observations conducted by a school administrator other than the principal; c. classroom observations conducted by someone other than a school administrator, such as a peer or mentor teacher; d. self-assessment; e. portfolios or other artifacts of teacher practice; f. peer assessments other than classroom observations; g. student work samples; h. student surveys or other feedback; i. parent surveys or other feedback</p> |
| <p>Requiring a minimum number of rating levels (two) be used to classify teacher performance</p> | <p>TL44. Do state regulations specify a required minimum number of rating levels to be used when evaluating overall teacher performance? If so, what is the minimum number of rating categories that is required?</p> |
| <p>Requiring districts to use student achievement growth to evaluate some or all teachers and specifying the extent to which it must factor into teacher evaluations</p> | <p>TL38. Currently, do state regulations specify the extent to which evidence of student growth must factor into the overall teacher evaluation? For example, regulations may call for it to be a “significant” factor in evaluations or call for student growth to have a specific weight, such as 20 percent. If this is specified in state regulations, please describe those specific regulations.</p> <p>[Screener question: TL36. Currently, does the state require that districts use student growth as one measure when evaluating all teachers, some teachers, or no teachers?]</p> |
| <p>Requiring evaluations of all teachers (probationary and nonprobationary) take place at least annually</p> | <p>TL40. Currently, how often does the state require that teachers be evaluated during their probationary period? For example, every six months, annually, every other year, or at some other interval?</p> <p>TL41. Currently, how often does the state require that teachers be evaluated in their nonprobationary period? For example, every six months, annually, every other year, or at some other interval?</p> <p>[Screener question: TL39. Currently, how long is the probationary period (the period before tenure is awarded) for teachers in your state?]^a</p> |

| Policy | Interview Questions Addressing the Teacher Evaluation Policy |
|---|---|
| Requiring teacher evaluations to guide decisions about professional development and support for individual teachers | TL60. Currently, does the state prohibit, permit, or require teacher evaluation results to be used for any of the following purposes? Please select just one answer for each potential use. If your state both permits and requires that evaluation results be used for a particular purpose, please report that purpose as required. a. to guide decisions about professional development and support for individual teachers [Screener question: TL59. Currently, does the state mandate how the results of teacher evaluations must be used?] |
| Requiring teacher evaluations to guide decisions about dismissal of teachers | TL60. Currently, does the state prohibit, permit, or require teacher evaluation results to be used for any of the following purposes? Please select just one answer for each potential use. If your state both permits and requires that evaluation results be used for a particular purpose, please report that purpose as required. b. to guide decisions about dismissal of teachers [Screener question: TL59. Currently, does the state mandate how the results of teacher evaluations must be used?] |
| Requiring teacher evaluations to guide decisions about compensation | TL60. Currently, does the state prohibit, permit, or require teacher evaluation results to be used for any of the following purposes? Please select just one answer for each potential use. If your state both permits and requires that evaluation results be used for a particular purpose, please report that purpose as required. c. to guide decisions about annual teacher salary increases; d. to guide decisions about teacher bonuses or performance-based compensation other than salary [Screener question: TL59. Currently, does the state mandate how the results of teacher evaluations must be used?] |
| Requiring teacher evaluations to guide decisions about career advancement | TL60. Currently, does the state prohibit, permit, or require teacher evaluation results to be used for any of the following purposes? Please select just one answer for each potential use. If your state both permits and requires that evaluation results be used for a particular purpose, please report that purpose as required. e. to guide decisions about career advancement opportunities [Screener question: TL59. Currently, does the state mandate how the results of teacher evaluations must be used?] |

Source: Interviews with state administrators in spring 2012 (http://www.mathematica-mpr.com/publications/PDFs/Spring_2012_State_Interview_Protocol.pdf).

^a This question indicated whether states have a probationary period for teachers. If not, only TL41 was used in the analysis.

Table A.2. Sensitivity Analyses for Number of Teacher Evaluation Policies Aligned with the RTT Application Criteria in Spring 2012, Revised Construct Coding for Items Two and Three in Exhibit 1

| Number of Teacher Evaluation Policies Aligned with RTT Application Criteria: | Early RTT | Later RTT | Non-RTT |
|---|-----------|-----------|-----------|
| As presented in this brief | 3.7 | 3.6 | 2.2 |
| Item Two is constructed using three or more rating categories | 3.7 | 3.2 | 2.1 |
| Item Three is constructed to use a fractional approach for requiring student achievement growth and specifying the weight at which student achievement growth factors into teacher evaluations | 3.6 | 3.7 | 2.2 |
| Item Two is constructed using three or more rating categories <i>and</i> Item Three is constructed to use a fractional approach for requiring student achievement growth and specifying the weight at which student achievement growth factors into teacher evaluations | 3.6 | 3.3 | 2.1 |
| Number of States | 12 | 7 | 31 |

Source: Interviews with state administrators from 49 states and the District of Columbia in spring 2012.

Note: See endnote 20 for a description of the fractional approach used for Item Three.

ENDNOTES

¹ Weisberg, D., Sexton, S., Mulhern, J., and Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. Brooklyn, NY: The New Teacher Project. Retrieved February 2013 from <http://carnegie.org/fileadmin/Media/Publications/widget.pdf>.

² Jacob, B.A., and Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subject performance evaluation in education. *Journal of Labor Economics*, vol. 26, no. 1, pp. 101-136.

³ The RTT applications and materials for Phases 1, 2, and 3 can be found at <https://www2.ed.gov/programs/racetothetop/applicant.html>. The RTT application required applicants to address one absolute priority area focused on the extent to which states were taking a comprehensive approach to education reform. The absolute priority required states to address the four education reform areas laid out in the application (addressed in sections B through F of the application), as well as the “State Success Factors” criteria (addressed in section A of the application). The RTT application included two types of criteria: (1) state reform conditions criteria that were “used to assess a state’s progress and its success in creating conditions for reform in specific areas related to the four ARRA [American Recovery and Reinvestment Act of 2009] education reform areas” and (2) reform plan criteria that were used to “assess a state’s plan for future efforts in the four ARRA education reform areas.” For each of the four education reform areas, states were free to decide which policies to adopt and implement, as well as when, so long as they fulfilled the plans proposed in their applications within the four-year grant period. The RTT scoring rubric included 19 criteria worth up to 485 points. (Applicants that addressed a competitive priority area could earn an additional 15 points.) States’ plans for Section D(2), teacher and principal evaluation, were worth 58 points, or 12 percent, of the RTT application’s absolute priority areas. These 58 points were allocated as follows: 5 points for measuring student growth, 15 points for developing evaluation systems, 10 points for conducting annual evaluations, and 28 points for using evaluations to inform key decisions. ED designed and implemented a review process that identified areas in which RTT grantees need assistance and support to meet their goals. The review process includes on-site program reviews, grantee self-evaluations, and Annual Performance Reports (APRs) that are submitted by states. RTT grantees’ APRs can be found at <https://www.rtt-apr.us/>. States’ progress on the adoption and implementation of their plans is monitored against the plans laid out in their RTT applications. Interested readers can find each state’s plans for addressing RTT priority areas in their RTT applications, which can be found at <https://www2.ed.gov/programs/racetothetop>. The Office of the Inspector General also conducted an audit to (1) determine the extent to which RTT grantees have adhered to timelines established in their applications and related scopes of work, and achieved project performance measures and goals; (2) evaluate the effectiveness of program oversight to ensure that funds were used as intended and anticipated recipient performance was achieved in support of overall programmatic goals. The audit results are available at <http://www2.ed.gov/about/offices/list/oig/auditreports/fy2014/a19m0003.pdf>.

⁴ See section D(2) in the RTT application. This section also outlines criteria for evaluating states’ plans for improving principal effectiveness. For the purposes of this brief, we focused on teacher evaluation systems.

⁵ Mihaly, K., McCaffrey, D., Staiger, D., and Lockwood, J. (2013) A composite estimator of effective teaching. Retrieved February 2013 from http://www.metproject.org/downloads/MET_Composite_Estimator_of_Effective_Teaching_Research_Paper.pdf.

⁶ Kane, T., Taylor, E., Tyler, J., and Wooten, A. (2010). Identifying effective classroom practices using student achievement. Cambridge, MA: National Bureau of Economic Research. Retrieved February 2013 from <http://www.nber.org/papers/w15803.pdf>.

⁷ Gordon, R., Kane, T., and Staiger, D. (2006). Identifying effective teachers using performance on the job. Washington, DC: The Brookings Institution. Retrieved February 2013 from http://www.brookings.edu/views/Papers/200604hamilton_1.pdf.

⁸ Available at <https://www.rtt-apr.us/>.

⁹ *Race to the Top: States Implementing Teacher and Principal Evaluation Systems despite Challenges*. GAO-13-777. Washington, DC: September 2013.

¹⁰ Doherty, K., and Jacobs, S. (2013). Connect the Dots: Using evaluations of teacher effectiveness to inform policy and practice. Retrieved November 2013 from http://www.nctq.org/dmsStage/State_of_the_States_2013_Using_Teacher_EvaluationsNCTQ_Report.

¹¹ To thoroughly document the educational policies in each state, we conducted structured interviews with representatives from state education agencies in 49 states and the District of Columbia in spring 2012. The interviews covered six main topic areas: (1) state capacity to support education reforms, (2) state data systems, (3) standards and assessments, (4) teacher and principal certification and evaluation, (5) school turnaround, and (6) charter schools. Interviews were organized into modules

(one per topic area). Typically, we interviewed different respondents for each module. Respondents were chosen based on which state administrators were most knowledgeable about each topic area. The study team achieved a 98 percent response rate on the spring 2012 state interviews (Texas was the one state that did not complete a spring 2012 interview).

¹² We combined Round 1 and Round 2 RTT grantees for several reasons: the funding amounts and scope of the grants in those rounds were similar, only two states received Round 1 RTT grants, and the first two rounds of grants were awarded within roughly five months of each other. The two Round 1 grantees were Delaware and Tennessee. The ten Round 2 grantees were the District of Columbia, Florida, Georgia, Hawaii, Maryland, Massachusetts, New York, North Carolina, Ohio, and Rhode Island. The seven Round 3 grantees were Arizona, Colorado, Illinois, Kentucky, Louisiana, New Jersey, and Pennsylvania.

¹³ The size of the awards in each round was based on each state's share of the national student population and the overall amount of funds allocated to the program in that round.

¹⁴ The six areas in the RTT application were: (1) state success factors, (2) standards and assessments, (3) data systems to support instruction, (4) great teachers and leaders, (5) turning around the lowest-achieving schools, and (6) state reform conditions.

¹⁵ The RTT Phase 1 and 2 applications can be found at <http://www2.ed.gov/programs/racetothetop/applicant.html>. The RTT Program Guidance and Frequently Asked Questions document can be found at <http://www2.ed.gov/programs/racetothetop/faq.pdf>.

¹⁶ The interview protocols used for this analysis can be found at http://www.mathematica-mpr.com/publications/PDFs/Spring_2012_State_Interview_Protocol.pdf. The interview questions used from the teacher and leader module of the state administrator interview protocol for this analysis were TL36, TL38, TL39, TL40, TL41, TL42, TL43, TL44, TL59, and TL60.

¹⁷ The study team was unable to find explicit guidance in the RTT application or corresponding materials (such as the Frequently Asked Questions document provided to all applicants) regarding what was meant by "Establish clear approaches to measuring student achievement growth for individual students." For this reason, we did not align any interview questions to this criterion. In addition, for some criteria, the interview questions aligned with only part of the criteria. For example, for the first criteria listed in Exhibit 1, there were no interview questions explicitly aligned with implementing transparent evaluation systems.

¹⁸ For the first case (item one in Exhibit 1), the points assigned to each state were determined by responses to 19 questions about measures used to evaluate teachers in tested and nontested grades and subjects (see Table A.1 for a list of the 19 questions). A state that responded *yes* to all 19 questions received one point for that item, while states that responded *yes* to only some of the questions received a fraction of one point corresponding to the number of *yes* responses. For example, a state that responded *yes* to eight questions received eight-nineteenths of a point. For the second case (item seven in Exhibit 1), the points assigned to each state were determined by responses to two questions about requirements to use evaluation results to guide compensation decisions. (The first question was about annual salary increases, and the second was about performance-based compensation.) States that reported requiring evaluation results to inform both types of decisions received one point. States that reported requiring evaluation results to inform only one of those decisions received half a point. In both cases, if a response to one or more of the questions was missing, we used only the set of nonmissing responses to categorize the state. For example, if a state responded to only 15 of the 19 questions for item one, and 8 of the 15 responses were *yes*, then that state would receive eight-fifteenths of a point. Due to concerns that states that have not yet required an RTT-aligned policy addressed in a question may be more likely to not respond to that question, we conducted sensitivity analyses that instead treated missing responses as *no* responses. Our findings are not sensitive to this assumption, with no changes in the average number of policies required by early, later, and non-RTT states reported in Figure 1, and no changes to the tabulations in Table 2.

¹⁹ Table A.2 presents a sensitivity check showing the number of teacher evaluation policies required by states if three or more rating levels were used for this item instead of two or more rating levels.

²⁰ We interpreted the lack of specific thresholds to imply that ED considers it important to include student achievement growth in a substantial way in teacher evaluations but that they left it up to states to decide how to implement these reforms. For item three, states received one point if they responded *yes* to two questions about (1) requiring student achievement growth as an evaluation measure for some or all teachers, and (2) whether state regulations specified the extent to which student achievement growth must factor into evaluations. States that required student achievement growth as an evaluation measure for

some or all teachers, but did not specify the extent to which it must factor into evaluations, received zero points. We chose not to use a fractional method such as that used for item one, which would be interpreting the application wording as meaning that the higher the weight the better, because some states simply reported that they require student growth to be a “significant,” “substantial,” or “primary” factor, rather than providing the exact weight. Table A.2 presents a sensitivity check that displays the number of teacher evaluation policies required by states if the fractional method was used for this item. In this sensitivity check, the variable could take on values of 0 (if a state reported that it does not require student achievement growth), 1/2 (if a state reported that it requires student achievement growth but does not specify the weight at which student achievement growth factors into the teacher evaluation), or 1 (if a state reported that it requires student achievement growth *and* specifies the weight at which student achievement growth factors into the teacher evaluation).

²¹ For item four in Exhibit 1, the study team interpreted the RTT guidance to mean that the requirement to evaluate teachers annually applied to *all* teachers. During interviews we asked about the frequency of teacher evaluations for two types of teachers (probationary and nonprobationary) and we required that a state responded “every six months” or “annually” to both probationary and nonprobationary to receive one point. (The exception is states that don’t have a probationary period, for which we required a response of “every six months” or “annually” to a question about frequency of evaluation for nonprobationary teachers to receive one point). Therefore, only those states that required annual evaluations for all teachers received one point for requiring this policy. States that reported requiring annual evaluations for only some of their teachers (such as probationary teachers) received zero points. We combined these items into a single point, rather than each receiving its own point, to prevent overweighting these interview items relative to how they are represented in the application and because distinctions that were made in the interview questions were not explicit in the RTT application (the application referred to “all teachers,” whereas the interview asked about probationary and nonprobationary teachers separately).

²² If a state did not have any points for a particular policy due to complete nonresponse for that policy, we calculated the average number of points the state received for the non-missing policies and used that to rescale the result. For example, if out of the eight teacher evaluation policies, a state had five that were non-missing, two of which equaled 1 and the remaining three equaled 0, then the state’s average points for the non-missing policies would be 2/5, and the total number of points across the eight policies would be rescaled to $(2/5)*8$ or 3.2. Across all states, the average percentage of teacher evaluation policies that were missing was less than one percent. We conducted sensitivity analyses that instead assumed that missing responses were worth zero points and found no differences in our results (see endnote 18).

²³ Four of the non-RTT states did not apply for RTT grants (AK, VT, ND, and TX). Readers may be interested in the average number of required teacher evaluation policies aligned with the RTT application criteria in spring 2012 for only those non-RTT states that applied for RTT grants. After excluding AK, VT, and ND (TX was already excluded because state administrators did not participate in the spring 2012 interviews), non-RTT states reported requiring an average of 2.4 teacher evaluation policies aligned with the RTT application criteria.

²⁴ Gill, B., Bruch, J., and Booker, K. (2013). *Using alternative student growth measures for evaluating teacher performance: what the literature says*. (REL 2013-002). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Retrieved November 2013 from <http://ies.ed.gov/ncee/edlabs>.

²⁵ One state reported having a probationary period for teachers, but did not provide a valid response about the frequency of evaluation of probationary teachers.

²⁶ Two states that provided “not applicable” responses to a question about how often the state requires that teachers be evaluated in their nonprobationary period were excluded from the analysis of frequency of evaluation of nonprobationary teachers.

²⁷ Baker, A. (2013, January 18). More money at risk on teacher evaluations. *The New York Times*.

²⁸ McNeil, M. (2013, January 15). Hawaii completes most Race to Top work, without teachers’ contract. *Education Week*.

²⁹ *Race to the Top: States Implementing Teacher and Principal Evaluation Systems despite Challenges*. GAO-13-777. Washington, DC: September 2013.

³⁰ Collective bargaining is a state-level indicator from the National Right to Work Legal Defense Foundation (<http://www.nrtw.org/rtps.htm>).

For more information on the full study, please visit:

http://ies.ed.gov/ncee/projects/evaluation/other_racetotop.asp



This brief was prepared for NCEE by Kristin Hallgren, Susanne James-Burdumy, and Irma Perez-Johnson of Mathematica Policy Research under contract number ED-IES-10-C-0077, Project Officer, Thomas E. Wei. We acknowledge the contributions of Mason DeCamillis and Emily Evans for their expert programming support on this brief. The authors greatly appreciate the participation and cooperation of state administrators interviewed for the study.