

Using Student Achievement Measures to Evaluate Teachers in Non-Tested Subjects and Grades – A Brief Summarizing State Practices¹

Erika Hall, Center for Assessment
Douglas Gagnon, Center for Assessment
Andrew Brownstein
Scott Marion, Center for Assessment
M. Christina Schneider, Center for Assessment
Jeri Thompson, Center for Assessment

September 15, 2014

In recent years, the accumulation of research tying student progress to teacher quality has sparked a growing effort among states and districts to include measures of student achievement in teacher evaluations. The \$4.4 billion Race to the Top (RTTT) fund, announced in July 2009, added significant momentum to this effort, and the more recent waiver requirements for the Elementary and Secondary Education Act (ESEA) expanded the incentives for states and districts to invest in such evaluations. But the ongoing process presents significant challenges, among them: identifying or developing appropriate assessments to obtain measures of student achievement, and determining how to analyze, interpret and combine data to make valid inferences about a teacher’s impact on student learning. While these challenges apply to all teacher evaluations, they are especially complex when defining procedures for the roughly two-thirds of teachers associated with “non-tested subjects and grades” (NTSG). For such teachers, a lack of high-quality student performance data based upon large numbers of students precludes the use of statistically-sophisticated approaches to estimate a teacher’s impact on student learning. These approaches, which include value-added models (VAM) and Student Growth Percentiles (SGPs), describe the current score for students by taking into account students’ prior score history and, in the case of some VAM, additional factors such as student demographics. For teachers associated with tested subjects and grades these “conditional status” type approaches are the leading vehicles for examining a teacher’s

¹ We are grateful to the Bill & Melinda Gates Foundation for supporting the production of this paper. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Foundation



contribution to student performance. A major component of such models is the ability to compare student performance data from two distinct points in time for each student within a given subject. The dearth of such data for teachers in NTSG means the development of analogous evaluation procedures for this population has greatly lagged behind that of other teachers.

For this paper, we defined a NTSG teacher as one who does not receive a state-supplied score reflecting student growth or teacher value-added based on his/her students' performance on a common assessment in the primary subject area for which he/she teaches and test scores within that subject area from at least the previous year. In order to learn more about evaluation procedures for this pivotal group of teachers, we gathered information specific to each state and the District of Columbia. We reviewed state Department of Education (DOE) websites for state laws and statutes related to teacher evaluation, relevant policy documentation, teacher evaluation resources and implementation guidelines, ESEA waiver applications, and any other pertinent documentation. Our analysis focused on 17 questions constructed specifically to extend our understanding around the manner in which student achievement measures are being identified, defined and used to support educator evaluation in NTSG. To ensure confidence in the accuracy of the information collected, each state's DOE was subsequently contacted and asked to review, validate and/or correct the information. Responses were received from 30 states.

Key Findings

- Roughly 4 out of 5 states allow districts to partially or fully determine the manner in which student achievement measures are defined, aggregated, and weighted to establish a final “growth²” or teacher impact rating for NTSG.
- Most states impose few requirements related to the number and type of assessments that may be used to establish evidence of student achievement for NTGS; however, 1 in 5 requires state approval of such assessments prior to use.
- Roughly 2 out of 3 states propose the use of a Student Learning Objective (SLO) or goal-setting process to estimate student growth for NTSG.

² “Growth” is usually defined in these systems as some difference between two or more test scores over time, but in most cases, these “growth measures” rarely meet technical definitions of growth.



- Approximately 1 out of 2 states use some form of shared attribution to assign student growth ratings to NTSG teachers.
- Approximately 1 in 5 states require or recommend the use of school performance factors in addition to student growth measures to support educator evaluation for NTSG.
- RTTT states tend to give greater weight to student achievement measures and allow less local control than non-RTTT states in the design and implementation of NTSG teacher evaluation systems.
- The percent of a teacher's overall evaluation associated with student achievement measures is typically the same in tested and non-tested subjects and grades and ranges from 15-50%.

Design and Timing

Although the design of evaluation systems for teachers in NTSG has lagged behind that of teachers in tested areas, we find that deadlines related to operational implementation are generally the same for both populations. Approximately 14 states had fully operational systems in place as of 2013–2014 and 21 have plans to be fully operational in 2014–2015. Of the remaining 16 states, 9 indicated that their systems would be operational in 2015–2016 or later, and 7 did not provide dates. Not surprisingly given their head start, we found that most RTTT states are ahead of non-RTTT states in terms of when their teacher evaluation systems are planned to be fully operational. For example, nine of the twelve Phase 1 and 2 RTTT winners had fully operational systems in place as of the 2013–2014 school year, and the remaining three are scheduled to be fully operational in 2014–2015. Delaware and Tennessee, the first RTTT winners, have had systems in place for all teachers since 2012–2013 and 2011–2012, respectively.

Approaches Used to Evaluate Teachers in NTSG

While a variety of different approaches may be used to incorporate measures of student achievement into teacher evaluations, Student Learning Objectives (SLO) or shared attribution are most common for NTSG. SLO-type procedures, in which teachers evaluate student performance relative to established learning targets as a means of estimating student growth, were discussed in roughly two-thirds of states. SLO are content- and grade/course-specific



measurable learning goals that can be used to document student learning over a defined period of time. SLOs provide a means for educators to establish learning goals for groups of students, monitor students' progress toward these goals, and then evaluate the degree to which students achieve these goals. Shared attribution utilizes the performance of multiple teachers or a school to establish an aggregate indicator of teacher impact on student learning and was recommended (either alone or in conjunction with another approach) in nearly half of states. Several variations of shared attribution are currently used to support teacher evaluations in NTSG. These approaches differ in terms of the level of attribution applied (grade, instructional, team, school district) and/or the type of data used (e.g., Math and English Language Arts VAM scores or SGPs, graduation rates, etc.). The most common shared attribution approach used in practice, however, is the attribution of school-based VAM or SGP results from teachers in tested subjects and grades to teachers in NTSG.

Number and Type of Assessments

In general, states impose few requirements related to the number and type of assessments that can be used to establish measures of student achievement to support teacher evaluation in NTSG. In most cases, districts can select or propose any assessments they believe will appropriately reflect student achievement in the skill or domain of interest, including: state developed standardized assessments; locally developed end-of-course (EOC) assessments, performance tasks or portfolios; large scale norm-referenced assessments, and vendor-developed interim and benchmark assessments. Despite this flexibility, however, approximately 1 in 5 states, generally RTTT states, require some level of state approval of selected assessments or data prior to use. When states impose requirements related to the selection/development of student achievement measures they typically include: the use of data resulting from state or district administered EOC assessment (when available), the selection of assessments from a pre-approved list, or the inclusion of one or more measures of school performance (e.g., graduation rate, attendance, student surveys, achievement gap reduction). In fact, approximately 1 out of 5 states allow or require the use of school-based performance measures to support teacher evaluation, despite the fact that these measures do



not allow for inferences about the unique impact of an individual teacher, or even small groups of teachers (e.g., grade-level teams) on student growth.

Weight of Student Achievement Measures in Overall Evaluation

When specified, the relative weight and role of student achievement measures was typically the same for TSG and NTSG within a given state. The percentage of an educator's overall evaluation associated with student achievement measures varies from 15-50%, with approximately one-third of states indicating that student achievement measures comprise 50% of a teacher's overall evaluation, one-third suggesting a weight of 40% or less and, one third not specifying a relative weight or percentage for these measures. Predominantly, the emphasis given to student achievement measures was greater in RTTT states than among non-RTTT winners. The majority of RTTT states recommend that student achievement measures represent 50 percent of a teacher's overall evaluation, compared to approximately 1 in 5 non-RTTT winners.

Local Control

One compelling finding is the degree of local control afforded to districts in the development of NTSG teacher evaluation systems. The majority of states indicated that the approach used to estimate student growth in NTSG was either partially or fully determined by districts. In addition, 4 out of 5 states afford districts some flexibility to define the manner by which student achievement measures are defined, aggregated, and weighted to establish a final growth or teacher impact rating in NTSG. Therefore, even if districts within a state share a general approach related to the use of student test scores for teacher evaluation (e.g., SLOs or shared attribution), the manner in which that approach is implemented may differ greatly from district to district.

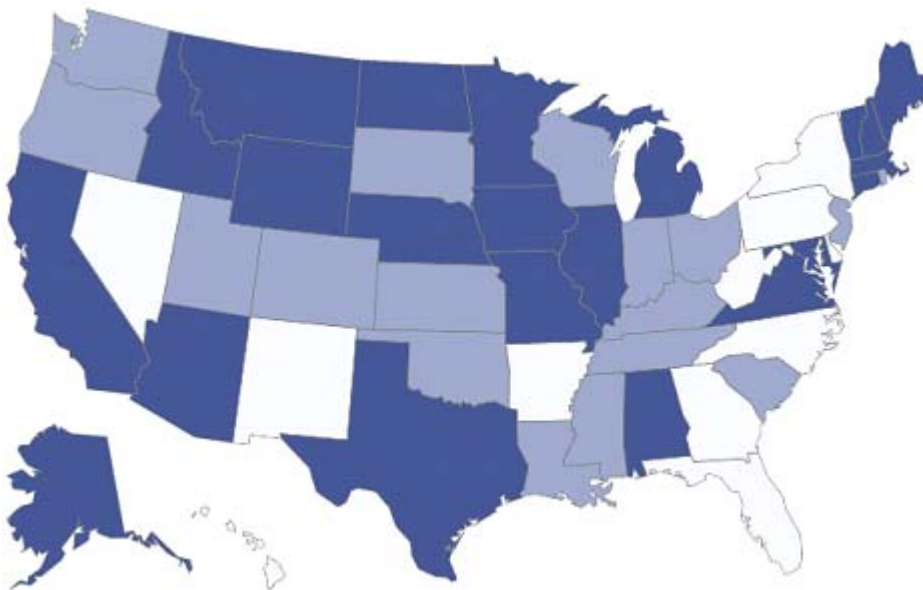
We find in roughly 1 in 5 states, NTSG teacher evaluation is almost completely in the purview of localities. For these states, decisions related to the selection, approval, use and aggregation of student achievement measures are left in large part to districts and/or schools. These states include those that have not received ESEA waivers, states that have yet to fully establish policy in this area, and states that strongly favor local control and have only issued guidelines around



teacher evaluation. In just as many states, however, states were likely to prescribe one or more key components of the teacher evaluation system. For example, in several states the approach(es) districts must use to calculate measures of student growth for inclusion in the teacher evaluation system are completely dictated by the state.

To capture variation in the discretion afforded to districts for NTSG teacher evaluation, an index of local control was established by quantifying and combining responses to survey questions which pertain to this issue. States were then placed into groups of low, moderate, and high local control based on their respective local control index value (see Figure 1). There is great deal of variability across states with respect to the degree of local control afforded to districts. Eleven states were categorized as providing low local control to districts, while 17 and 22 afforded moderate and high local control, respectively. Furthermore, we find that RTTT states are much more likely than non-RTTT states to curtail the amount of local control offered to districts. Approximately 40 percent of RTTT states (i.e., 7), compared to 12 percent of non-RTTT states (i.e., 4) were categorized as having strong state control.

Figure 1: Map Illustrating Low (Lightest), Moderate, and High (Darkest) Local Control States in NTSG Teacher Evaluation in US



The National Center for the Improvement of Educational Assessment, Inc.
31 Mount Vernon Street • Dover, New Hampshire 03820
phone: (603) 516-7900 • fax: (603) 516-7910 • www.nciea.org

Conclusion

In response to incentives associated with RTTT and ESEA waiver requirements, states have become more deliberate about specifying plans for evaluating teachers in NTSG. While implementation plans are at different stages of completion, we identified several trends across states related to how NTGS evaluation systems are being defined and operationalized. Three of the most prevalent include:

- the use of SLO or goal-setting procedures to establish estimates of student growth;
- the use of some form of shared attribution to assign student growth or school/student achievement estimates to individual teachers; and
- the granting of high local control to districts in the selection of assessments and/or data used as evidence of student achievement.

Given the diversity with which these procedures may be implemented, these findings highlight the need for targeted research that informs best practice related to the specification and implementation of these types of evaluation activities.

In addition to these trends, we identified several areas of variability across states including the:

- number and type of teachers that fall under the umbrella of NTSG;
- extent to which efforts are being put in place to move educators from NTSG to TSG status by implementing “test-based” approaches in formerly NTSG;
- emphasis given to school-based measures in teacher evaluation;
- degree of support provided to districts related to the selection, development and review of high quality assessment materials and teacher evaluation procedures; and
- the type and degree of local control afforded to districts in the design and implementation of their NTSG teacher evaluation systems.

In many instances, observed variability was related to RTTT status, as RTTT winners required earlier deadlines to be fully operational, exhibited higher state control, placed more weight on



student achievement measures in a teacher's overall evaluation, and provided for more detailed evaluation guidelines and procedures. Ongoing inspection of teacher evaluation results for RTTT and non-RTTT states may, therefore, provide useful information about those practices which do/do influence the attainment of desired student and teacher outcomes. Such analysis is crucial, because many of the practices currently in use to support teacher evaluation are based in large part on theories as to what motivates desired change (e.g., improved instruction, student learning) and how it can be measured. Research and evaluation can shed light on the conditions under which different teacher evaluation models will be most beneficial given the goals of evaluation.

In other cases, variability was tied to a state's plans for establishing measures of student achievement and incorporating them into the NTSG evaluation system. For example, North Carolina and Florida are working to develop end-of-course assessments for every subject area, so that all teachers have high quality measures of student achievement and ultimately are considered "tested." In most cases, we can only speculate as to the source of observed variability across states, as it likely relates to such things as the state's theory of action related to educator evaluation, the availability of certain assessments and data, fiscal and human resources, legislative constraints, and other factors.

Finally, while our research suggests that approximately 3 out of 4 states allow for districts to either develop their own evaluation system or propose modifications to components of the state-defined model, differences in the procedures used to support evaluation suggest greater local control is afforded to districts within the context of NTSG. For example, most states use VAM models or SGPs in lieu of or in conjunction with other approaches to establish estimates of teacher impact in tested subjects and grades. Such techniques necessitate the use of state developed standardized assessments and complex statistical calculations that preclude the provision of local control to districts around these aspects of the evaluation system. In contrast, the approaches utilized in NTSG, including SLOs and shared attribution, necessitate greater flexibility and local choice in order to provide for valid inferences related to student growth and teacher impact.



While the influence of differing levels of local control will likely vary across systems depending on a variety of factors (e.g., the availability of local resources and the selected analytic approach) significant differences in local control for tested and non-tested subjects and grades may be a source of concern for stakeholders in terms of apparent comparability of results. Whether such concerns are warranted can only be answered by additional research on the impact of different types and degrees of local control on desired outcomes, and the extent to which those findings interact with district, school and teacher-specific factors.



The National Center for the Improvement of Educational Assessment, Inc.
31 Mount Vernon Street • Dover, New Hampshire 03820
phone: (603) 516-7900 • fax: (603) 516-7910 • www.nciea.org