



# Teacher evaluation and local control in the US: an investigation into the degree of local control afforded to districts in defining evaluation procedures for teachers in non-tested subjects and grades

Douglas J. Gagnon, Erika L. Hall and Scott Marion

The National Center for the Improvement of Educational Assessment, Inc., Dover, NH, USA

## ABSTRACT

Many states only recently incorporated indicators of student achievement into teacher evaluation systems for Non-Tested Subjects and Grades (NTSG). This study examines how practices related to the inclusion of student achievement measures vary across states as to the discretion left to districts in defining and implementing evaluation systems for teachers in NTSG. For each state, information about current practices was obtained through document analysis and, when provided, feedback from state department representatives. We find substantial variation in state policies. Some states — notably those that received Race to the Top funding — afford districts with considerably less local control than others. Results presented here provide rich descriptive information and highlight the need for future studies related to local control and teacher evaluation.


## ARTICLE HISTORY

Received 8 December 2015  
Accepted 8 March 2016

## Introduction

In recent years, states and local school districts across the United States moved to incorporate indicators of student achievement into existing or newly envisioned teacher evaluation systems. This trend was largely the result of two federal policy initiatives. The first was Race to the Top (RTTT), a competitive grant programme which offered sizeable funds to the 12 states who were selected as Round 1 or 2 winners through a comprehensive application process. The second was Elementary and Secondary Education Act (ESEA) waiver requirements, which allowed states flexibility in meeting provisions of the No Child Left Behind (NCLB) Act in return for adhering to additional accountability measures. As of 2014, 43 states plus Puerto Rico and the District of Columbia received ESEA waivers, which necessitated the inclusion of measures of student growth in teacher evaluation. The rationale behind this policy is that effective teaching should result in improved student outcomes, and consequently teacher evaluation systems must, among other things, incorporate such measures if they are to identify effective teaching.

**CONTACT** Douglas J. Gagnon  [douglas.gagnon@unh.edu](mailto:douglas.gagnon@unh.edu)

 Supplemental data for this article can be accessed at <http://dx.doi.org/10.1080/0969594X.2016.1167669>.  
Douglas Gagnon is now a Research Associate at The Carsey School of Public Policy.

Two of the biggest challenges faced by states working to develop teacher evaluation systems are determining how to identify appropriate measures of student achievement and how to use those measures to make inferences about a teacher's impact on student learning. While these challenges apply to the evaluation of all teachers, they are especially problematic when defining procedures for teachers associated with subjects and grades for which high-quality student performance data (such as state-developed standardised assessments) are not readily available; ESEA only requires annual standardised testing for mathematics and English Language Arts in grades 3 through 8 and once in high school. For teachers in 'Tested Subjects and Grades' (TSG), econometric approaches have been developed, which can, to some degree, isolate a teacher's contribution to student learning. These techniques include value-added modelling (VAM) and student growth percentiles (SGP), which use data about a student's prior achievement to predict 'expected' performance, or compare achievement to a set of academic peers, respectively. For teachers not in these areas – those in the so-called 'non-tested subjects and grades' (NTSG) – the lack of vertical standardised achievement results largely precludes the use of VAM and SGP, and other techniques which purport to estimate teacher effectiveness using student outcomes had not been developed until much more recently.

Half a decade ago, the number of states with state-defined teacher evaluation systems in place was extremely limited – consisting mainly of those which had received RTTT or other external funds (e.g. Teacher Incentive Fund) – and most of the attention has focused on the development of procedures appropriate for TSG with relatively little attention paid to NTSG. In the span of just three years, however, most states had responded with teacher evaluation systems that include specifications for NTSG (Hull, 2013), albeit in very different ways. This was due to the novelty of using student achievement in NTSG teacher evaluation, the adoption of which was accelerated and shaped by RTTT and ESEA Waiver requirements. For these and other related reasons – a lack of consensus around best practices in this area and differing effects of RTTT for winners and non-winners – we hypothesise that states were likely to devolve varying degrees of authority to districts in the construction and implementation of teacher evaluation systems.

Until these recent policy developments, teacher evaluation was decentralised – a practice which essentially was completely under the purview of the nearly 14,000 local school districts in the United States (National Center for Education Statistics [NCES], 2012). A growing dissatisfaction with systems of educator evaluation, coupled with a renewed focus on the classroom teacher as a means to promote achievement, has led to teacher evaluation reform to be incorporated into the wider accountability movement. Broadly speaking, the insertion of accountability mechanisms through federal and state policy has the intended effect of creating high standards through the establishment of standardised assessments from which school quality is judged (Figlio & Loeb, 2011). Standardisation has ensured higher levels of comparability across schools, which, proponents would argue, allows the education community to better diagnose and replicate effective practices due to clearer (albeit narrow) definitions of school quality. This same line of reasoning argues for federal and state involvement in the creation of teacher evaluation systems: a level of standardisation and comparability allows for the identification and promotion of effective teaching practices, as well as consequences for continued ineffective teaching.

Conversely, a more centralised system of education brings with it potential disadvantages in the form of decreased flexibility necessary to tailor the educational process to

a variety of student populations. In regards to teacher evaluation, this could mean that successful teachers may look different – in terms of practice as well as the evidence used to support their effectiveness – depending upon the particular needs of the students they serve. This tension may only be resolved by understanding which aspects of education may require greater flexibility, and which aspects benefit from some standardisation. One may envision an ideal distribution of decision-making power in the creation of teacher evaluation systems, but the nascent body of literature does little to illuminate what such a distribution would look like. An important first step in this regard is to uncover key areas of variability that currently exist, as well as the policy mechanisms that created such a landscape; this paper makes this first step. We detail the extent to which local control varies across states in a crucial area of reform, and examine how variation in local control is related to a signature piece of federal legislation. We find very few studies that empirically examine local control, which adds to the importance of the findings presented here.

It is important to emphasise the relevance of such a line of inquiry within the context of US public education, which has long been seen as the responsibility of individual states as opposed to the federal government – a dynamic encoded in the Tenth Amendment of the Constitution. Despite growing financial involvement by the federal government in recent decades, state and local governments currently contribute roughly 90% of all public education funds (National Center for Education Statistics [NCES], 2015). Given that much power is devolved from the federal government, states may differ considerably in the extent to which the state governmental entities may create top-down education policies for the local school districts to follow, on one hand, and the amount that states may delegate decision-making power to local school districts, on the other. In this paper, we define ‘local control’ to be the complete or near complete devolving of education policy-making authority down to the level of school districts. During the 1990s and 2000s, there was an unprecedented expansion in the role of the federal government via accountability legislation (Figlio & Loeb, 2011). The signature federal policies of this era – NCLB, RTTT and ESEA Waivers – increased the federal role through mechanisms intended to broaden the influence of state governments and decrease local control (McGuinn, 2012). This expanded deferral role was questioned by many, as there exists wide support for general notions of local control in the United States (Jacobsen & Saultz, 2012). A scan of popular press articles after the re-authorisation of ESEA in late 2015 (Layton, 2015; Resmovits, 2015) suggests that issues of local control remain very visible in the United States.

### **Purpose**

In this paper, we briefly summarise current state practices related to the collection and evaluation of student achievement measures to support the evaluation of NTSG teachers. Then we examine how such practices vary across states as to the discretion left up to districts in determining their own systems. A focus on the role of student achievement, specifically, in teacher evaluation is appropriate in this respect as other features of teacher evaluation (i.e. observations of practice) can typically be applied to all teachers and therefore show much greater consistency across states. Furthermore, the use of student achievement in the evaluation for teachers in TSG has received considerably more attention in the literature, so there is less of a need to document policy surrounding these teachers. Thus, the use of student achievement measures in NTSG teacher evaluation provides a unique and compelling lens

from which to examine issues of local control. Specifically, we wish to answer the following research questions:

- (1) What are the dimensions on which NTSG evaluation systems vary?
- (2) Is there variation across states as to the discretion left up to school districts in establishing systems of teacher evaluation in NTSG?
- (3) Is the amount of local control in a state related to the receipt of RTTT funds?

## Literature review

### *The state of teacher evaluation in NTSG*

The paucity of literature around teacher evaluation in NTSG relative to TSG could be considered alarming, especially given the fact that NTSG teachers outnumber TSG teachers by more than two-to-one (Prince et al., 2009). Indeed, many conversations about the implications of VAM or SGP in teacher evaluation often take place without making mention of this fact. As recently as 2011, Sawchuk observed the lack of consensus about best practices surrounding the use of student achievement in NTSG teacher evaluation. This was due in part to the numerous challenges that arise when attempting to make inferences about teacher effectiveness. The use of student growth in teacher evaluation has several requirements: that well-defined standards exist, that well-designed assessments are created and reflect student growth and that student growth may be attributed to the teacher (Herman, Heritage, & Goldschmidt, 2011). Three years ago, states were still exploring potential approaches that could be reliable, valid, rigorous and enable cross-classroom comparisons. As Buckley and Marion (2011) as well as Goe (2010) observed, a few approaches were being pursued by a handful of RTTT states. These methods included models such as shared attribution and student learning objectives (SLO). Shared attribution is the practice of using common grade- or school-wide metrics for multiple teachers; SLOs are a broad family of approaches in which learning goals are set for students using a variety of possible achievement measures, with student progress towards those goals informing teacher evaluation. Alternatively, states could circumvent these practices through the creation of new end-of-course assessments and associated value-added scores and turn NTSG into TSG. Before examining the extent of use for these approaches, we first explore the roles that federal and state policy-makers have taken in creating new systems of teacher evaluation specifically as well as in education policy more generally.

### *Federal and state involvement*

Although the power over public education is formally under the purview of state governments, historically states have informally ceded control in most areas to localities, albeit to varying degrees. However, federal and state involvement in education policy has grown steadily in the past decade and a half, initially through the passage of NCLB in 2001 and the subsequent use of accountability mechanisms. Many have argued for a more centralised education system as a means to achieve equity (McDermott, 1999; Welner, 2001), and this presents one reason why accountability has received some level of support from both sides of the political divide. As Riley and Coleman (2011) note, the current policy agenda around equity is grounded in the broader principles of accountability set by federal and

state policy. Whether or not a greater federal and state presence in education in the past two decades has actually led to a shrinking of the achievement gaps faced by students in poverty and students of colour is another question altogether, but its use as a rallying cry for a more centralised education system in the United States is well established. More specifically, classroom teachers in particular were increasingly viewed as a primary means by which to address disparities in achievement along lines of race and income. The NCLB goal of ensuring a ‘highly qualified teacher in every classroom’ represented a pivotal shift in accountability, drawing focus away from schools and communities and placing the emphasis largely on teachers (Ingersoll, 2005). This movement has evolved, but not winnowed, now including a wider array of tactics intended to promote effective classroom teaching; professional development, accountability mechanisms for teacher preparation programmes, and teacher evaluation systems have all received renewed focus in recent years.

The most current federal imprint on education policy largely comes from two mechanisms: ESEA waiver applications and RTTT. Indeed, these two measures have influenced education policy in nearly every state. As of this analysis, all but seven states had been granted ESEA waivers, which included addressing Principal 3: a call for ‘supporting effective instruction.’ RTTT had a similar effect in spite of the fact that fewer than half of states received RTTT funding, because all but four states applied for grants and generally enacted or prepared to enact policies to be competitive in the application. Importantly, even applying for a RTTT grant entailed some stipulations, including the permissibility of linking of student achievement with teachers. The focus of RTTT shifted the timbre of federal involvement in education in two key ways. First, it represented a clear inducement rather than a penalty, provisioning extra funding rather than possibility withholding allocations that were long part of district budgets. Second, RTTT was concerned primarily with outcomes, rather than the process by which those outcomes are achieved (Nee, 2010). Because of this emphasis on ends – particularly in areas such as NTSG teacher evaluation where best practices were lacking – it is easy to see how many different approaches could be employed.

Indeed, these aspects of RTTT had profound implications. As McGuinn (2012) argues, ‘RTTT is fundamentally about two things: creating political cover for state education reformers to innovate and helping states construct the administrative capacity to implement these innovations effectively’ (p. 137). The former of these two goals is particularly germane in the realm of teacher evaluation, which has been a controversial area for states to address on their own. The result is that the increased role of the federal government in education policy has not diminished, but rather commensurately increased, the power of state governments in such areas (Marsh & Wohlstetter, 2013). In fact, this symbiotic expansion of the federal and state roles in education policy was observed in the years preceding NCLB as well (McDonnell, 2013). Many states moved from compliance monitors to active participants in the formation of teacher evaluation systems, often creating dedicated working units or partnering with outside agencies (Pennington, 2014). However, states also responded in very different ways, particularly as this relates to the state role in teacher evaluation. For example, Superfine, Gottlieb, and Smylie (2012) argued that New York and Rhode Island – both RTTT states – serve as contrasting case studies in terms of the level of local control afforded to districts in the creating of teacher evaluation systems. Those states intent on preserving local control may do so, to some degree, in the face of a greater federal presence; those states with historically strong roles may grow in reach in such an environment.

## **Local control**

A simple conclusion from this literature might be that individual school districts now have less sway in policy matters, on average. However, some have argued that a stronger national and state presence does not necessarily diminish the importance of district-level decision-making (Henig, 2009; Marsh & Wohlstetter, 2013). Rather than view the education policy labyrinth as a zero-sum game, one could infer that increased federal and state influence may actually empower localities. After all, successful national and state policy rests on the implementation efforts of local actors. As Henig (2009) observed, politics around education now more closely resemble that of other domestic issues such as housing and social services. However, if one takes a broad enough view, the locus of control must sum to zero; it flows as a tautology that all actors in a given arena cannot simultaneously grow in power. If all levels of school governance – as well as a slew of non-profits, foundations and think tanks – have greater influences on policy, then some group must be losing ground. The views advocated by teachers unions has in some ways shifted to be more in line with the federal (and many states’) policy agenda, particularly with regards to the evaluation of teachers (Toch, 2011; Tucker, 2012). Whether or not this is seen as a decrease in the influence of unions may be semantic. A key takeaway in regards to this study is that a greater overall federal and state involvement in teacher evaluation might not ensure less of a role for all districts in all ways, and even more likely not a diminished influence for all districts. A well-articulated state plan that devolves substantial decision-making to the local level could actually empower districts compared to the status quo. Conversely, one must also acknowledge that a highly prescriptive system could limit the reach of local decision-making.

At the heart of the debate over the locus of control is the trade-off between maintaining high standards, the efficiencies gained with state and federal control (avoiding the ‘reinvention of the wheel’), and the flexibility that is necessary when crafting effective solutions for schools that serve very different student populations. Indeed, the lack of flexibility and one-size fits-all critique of NCLB is some of the impetus behind ESEA flexibility waivers. The causal argument made for local control is that when individual schools and teachers are given greater autonomy and responsibility in decision-making, the process becomes more nimble and solutions may be targeted to the needs of their students. The body of research which connects the locus of control to student outcomes, however, is far from equivocal. There is some evidence to suggest that greater accountability – one form of decreased local control – raises overall student achievement. For instance, a number of studies have shown that accountability in the form of high-stakes testing may improve student achievement (Braun, 2004; Dee & Jacob, 2011; Hanushek & Raymond, 2004). However, such gains may not be distributed uniformly: Wei (2012) took advantage of the fact that several facets in NCLB led to variability in the stringency across states to examine the effect of accountability on achievement, finding substantially different effects of across grades, subjects and ethnic subgroups. The empirical support for the effectiveness of accountability should be further tempered by the notion that such gains in achievement may be caused by the narrowing of curricula, teaching to the test and other related concerns about high-stakes testing in general (Darling-Hammond, 2004), and there is some evidence to support this (Amrein & Berliner, 2002). Along these lines, a study by Honig and Rainey (2012) suggests that increased school-level decision-making may lead to a deeper form of school improvement, representing gains which may be difficult to measure.



This discussion could leave one wondering: What does local really mean? In essence, local control can have very different practical implications. For example, in examining cases where a state-defined value-added score versus a state-defined goal-setting process are suddenly required in high-stakes teacher evaluation – both instances of diminished local control – one may conclude that teachers are likely to feel much more empowered in the latter case than in the former one. Similarly, the implications of high local control are also muddled, as there are multiple levels within localities. Henig (2009) notes that school district bureaucracies have probably played the single most important role in diminishing school-level autonomy. Thus, even in cases where school districts are free to make their own decision, the true implementers of policy – principals and teachers – might have little control over some important facets of their work if district administration creates rigid procedures, or vice versa. To this point Honig and Rainey (2012) conclude that in order for school-level autonomy to be realised, district offices may need to undergo considerable systemic change by removing administrative barriers while helping schools build capacity to engage in such endeavours as budgeting and academic planning. These examples serve to highlight nuances when investigating issues of local control, but by no means should be seen as reasons to not delve into such a topic. The body of literature which empirically examines the effects of local control is still developing, and this study is in part an effort to push research deeper into this area by seeking to document its variability and connect it to policy levers.

Ultimately, this discussion of local control must venture beyond the overall matrix of decision-making to include those specific areas of education that should be loosely controlled and those which might require a greater state presence. Clearly the values of flexibility and accountability may collide at times, and in some ways flexibility is antithetical to educational practices whose aim is to ensure comparability and equity across schools. However, we believe that these values may coexist in education. As local control relates to alternative assessments in the classroom, for instance, Gong and Marion (2006) observed that policy-makers generally tolerate and even value flexibility in the classroom when it comes to helping students learn – but not when it comes to holding schools accountable. A pragmatic approach will seek to understand the elements of education are better served with a stronger centralised role, and those that should remain local. Such an approach rests on first knowing where decision-making rests, and how federal policy shaped the landscape.

A small number of previous studies have sought to operationalize local control, often times conceptualised as accountability. Carnoy and Loeb (2002) use an accountability index created by the Centre for Policy Research in Education (CPRE) to estimate the impact of state testing policies on student achievement. This index is on a zero to five scale, with zero corresponding to no state-level accountability, five corresponding to a high school competency test graduation requirement and intermediate scores determined by degrees of sanctions, rewards and frequency associated with state testing. Hull (2013) operationalized local control in teacher evaluations, specifically, using a rather simple three-point scale. States were deemed to have high involvement if components of evaluation were mandated, medium involvement if districts could adopt a model or develop a model with certain requirements, and low involvement if states played a small role in the implementation of a teacher evaluation system. Our current study extends this work considerably by applying an extensive protocol to aspect of teacher evaluation systems that is both most variable and contentious: the use of student outcomes.

## Data and methods

### Data

The data used in this study come from document analysis, which uncovered design aspects in systems of teacher evaluation for all 50 states plus the District of Columbia. A list of questions, which we refer to here as the ‘protocol’ (see Supplementary Material), was uniformly applied to each state in order to determine their respective policies related to teacher evaluation. The set of questions in the protocol serves to inform not only this but two additional studies, and therefore it contains numerous items of various relevance to this study. The protocol is exhaustive in nature, as it captures general aspects of teacher evaluation as well as those particular to student achievement measures, NTSG and local control.

The document analysis included reviews of state Department of Education (DoE) websites and linked policy and guideline documents, ESEA waiver applications, related state laws and statutes and other pertinent documentation. In most cases, it was rather straightforward to find supporting documentation in these areas which directly answered questions on the protocol document. For instance, ESEA Waiver applications, which among other things typically outline proposed changes to teacher evaluation systems, are all readily available on the US DoE website. In addition, most states have published supporting documentation on teacher evaluation systems, which usually provide immense detail on these matters. Many states have also made changes in state code regarding teacher evaluation, and thus legal documents also provided relevant information. Note that all these documents, state websites, etcetera, need not be fully reviewed; once a protocol was completely answered, document analysis for that state was concluded. In order to evaluate and ensure accuracy, these completed questionnaires were sent to a DoE personnel identified as being in charge teacher evaluation in that state. For instance, most states clearly identify a point of contact on their website for questions regarding teacher evaluation. Three follow-up emails were sent in roughly one week intervals, urging the state official to review and confirm the protocol document. We received feedback from 30 states, which greatly increased our confidence in the quality of the data used in this study.

### Methods

To provide context for this paper, we begin with a qualitative description of the evaluation practices commonly applied across the US for teachers in NTSG. This is meant to address our first research question, and to provide the reader with a broad overview of teacher evaluation policies. Subsequently, to answer our second and third research questions, we move to quantify the amount of local control in all 50 states in the area of teacher evaluation. To do so, questions 6, 7, 12, 13, 15 and 17 from the protocol – which are the only questions that address local control – were quantified and analysed for inclusion in an index variable.<sup>1</sup> Responses to questions 6, 7, 12, 13 and 15 were operationalized as 0, 1 or 2, with 0 indicating that a policy was completely dictated by the state, 1 suggesting that some amount of local control was preferred within some state prescriptions, and 2 indicating that full local control is granted in such a policy decision. Questions 17 was operationalized dichotomously (0 or 1) because the practical implications of this item in terms of local control – whether or not a district has the ability to choose between a compensatory or conjunctive evaluation



model – were deemed less important than the implications of the other five items. For each state, an NTSG Local Control Index (NTSGLCI) was created by summing the values assigned to each of these questions, resulting in score scale of 0 to 11. The states with a high NTSGLCI leave most/all decisions regarding the use of student achievement up to individual districts, including the structure of evaluation systems, how assessments are identified and how scores are operationalized and aggregated. Conversely, states with a low NTSGLCI generally dictate how assessments may be selected, what analytic approaches are used, and how scores are incorporated into evaluation systems.

Once the NTSGLCI is calculated, reliability testing is conducted through the calculation of Cronbach's alpha. A high Cronbach's alpha suggests that these items of local control are related, and thus our index measure has a high internal consistency. Ideally, the construct validity of this measure would be established through comparisons of multiple other scales that measure local control of states during the same time period. The only such scale is one developed by Hull (2013), which divides states into three categories of involvement in teacher evaluation system based on broad policy measures. A Spearman rank correlation coefficient is calculated for the relationship between the NTSGLCI and Hull's, 2013 index. Next, descriptive statistics of NTSGLCI are evaluated across states, including both measures of central tendency (mean, mode) and variation (standard deviation, qualitative descriptions of its distribution). We then graphically examine trends in NTSGLCI through the use of a choropleth map of the US. Since RTTT winners were overwhelmingly found on the East Coast, this would provide a visual confirmation of a relationship between RTTT receipt and local control. Finally, we test for this relationship more explicitly by comparing average NTSGLCI scores for RTTT and non-RTTT states.

## Findings

### *General trends*

By the 2014–2015 school year, 36 states had fully operational teacher evaluation systems in NTSG. We find considerable variation in the type of student achievement measures used in NTSG evaluation, as well as in how such measures are incorporated into an evaluation score. The types of assessments vary tremendously, from more traditional vendor- or state-developed end-of-course type assessments, to teacher-created assessments, to more broadly conceived measures of achievement such as graduation rates and achievement gap-reduction statistics. There is also considerable variability in how student achievement measures are then analysed and aggregated to inform evaluation. The use of goal setting procedures (usually through an SLO process) and shared attribution (which can include shared SLOs in addition to VAM results, gap reduction, graduation rates, etc.) are the two broad categories most commonly in use: an SLO process is being used in roughly two-thirds of states, while nearly half of states use some form of shared attribution. These two practices are by no means mutually exclusive, as the majority of states that promote shared attribution also incorporate some form of SLO. Portfolio approaches, which rely on the collection and evaluation of a body of student evidence to make inferences about the value a teacher adds to his/her students may be used in cases where districts have the freedom to define achievement measures, and are specifically referred to in states such as Tennessee and Massachusetts.

**Table 1.** Spearman correlation coefficients, measures of local control in NTSG teacher evaluation.

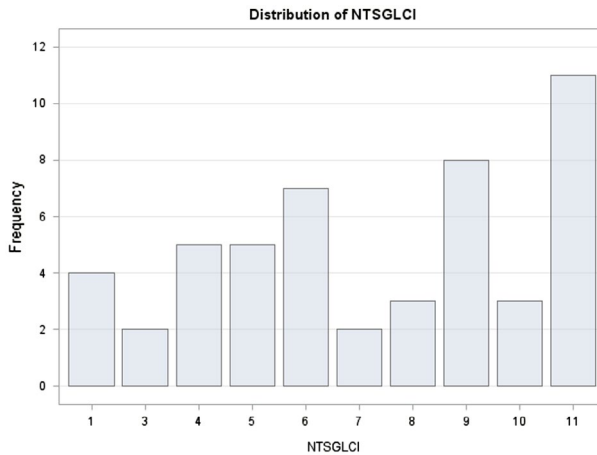
	Selects achieve	Approves achieve	Selects analytic	Aggregates scores	Alternate systems	Model structure	NTSGLCI
Who selects measures of student achievement?	1.00						
Who approves measures of student achievement?	0.58*	1.00					
Who selects the analytic approach?	0.28*	0.32*	1.00				
Who aggregates scores?	0.24	.31*	0.59*	1.00			
May alternate systems be developed?	0.24	0.33*	0.57*	0.52*	1.00		
Who determines model structure?	0.30*	0.27	0.50*	0.36*	0.60*	1.00	
<b>NTSGLCI</b>	<b>0.56*</b>	<b>0.62*</b>	<b>0.80*</b>	<b>0.73*</b>	<b>0.78*</b>	<b>0.70*</b>	<b>1.00</b>

\*Significant at the  $\alpha = 0.05$  level.

Even in states that share a general approach, the practical impact on students and teachers can differ substantially. For instance, the use of SLOs differs considerably in terms of who selects assessments, how many SLOs must be used, how weighting occurs and the amount of guidance provided to teachers and evaluators. For instance, the State of Georgia reviews and approves SLOs, whereas states such as New Hampshire, Rhode Island and Colorado largely leave this process up to the districts. The implementation of shared attribution is even more nuanced, as the level of attribution (grade, instructional team, school, district, etc.) as well as the metric used (mathematics and English Language Arts VAM scores, graduation rates, etc.) can completely change the meaning and consequences of shared attribution in teacher evaluation. There is, of course, variability across grades and subjects within states and districts, and this serves to limit the generalizability of practices to all NTSG teachers within a state. For instance, in Tennessee there are different procedures used in different content areas: a portfolio-based approach is used in pre-kindergarten, kindergarten, physical education, world languages and the fine arts, while shared attribution is used for teachers in other NTSG.

### **NTSGLCI: an index of local control**

The NTSGLCI was created by operationalizing and summing together the six questions related to local control from the protocol. These individual measures exhibited meaningful variability across states, as judged by their standard deviations. Furthermore, each individual measure was positively correlated to the other measures, and consequently the NTSGLCI (see Table 1), suggesting their suitability for inclusion in the index. The internal



**Figure 1.** Histogram of the index of local control in the use of student achievement measures in NTSG teacher evaluation, US states.

consistency of this composite was evaluated using Cronbach's alpha, which yielded a moderately high coefficient,  $\alpha = 0.79$ . This exceeds Nunnally's rule of thumb of 0.7 for a sufficiently high alpha (Nunnally, 1978), confirming the reliability of the composite. Next, a Spearman rank correlation coefficient is calculated between the NTSGLCI and Hull's (2013) index of state involvement. We find this correlation to be 0.32, which may be considered moderate correlation, and provides some construct validity for the NTSGLCI as a measure of local control.

We found that the degree of local control as it pertains to NTSG teacher evaluation varies substantially across US states, as illustrated by the histogram of NTSGLCI values in Figure 1.<sup>2</sup> The average NTSGLCI across all states was 7.6, with values ranging from 1 to 11. The modal score for the NTSGLCI was 11, which corresponds to essentially complete local control related to the specification and use of student achievement measures in teacher evaluation for NTSG. A total of 11 states received this maximum score, states which include those that have not received ESEA waivers, states that have yet to fully establish policy in this area and states that strongly favour local control and have only issued suggestive guidelines around teacher evaluation. Outside of this fair proportion of states with near-complete local control, the overall distribution of NTSGLCI was rather flat, with nearly the full range of possible outcomes being presented. For instance, 23 states exhibited an NTSGLCI of 6 or less, suggesting that nearly half of the states have a fairly strong presence in the creation of teacher evaluation systems in NTSG.

In order to graphically represent the distribution of NTSGLCI, values were aggregated into three categories: high, moderate and low local control: states with a NTSGLCI greater than 8 were deemed to have high local control; those with an NTSGLIC less than 5 were identified as low local control states; states in between were deemed moderate.<sup>3</sup> Figure 2 is a choropleth map of the US that shows these three categories, with darker shading corresponding to higher levels of local control. The general trends seen on this map – higher levels of local control in the north and west; lower levels in the south – are fairly consistent with prior studies that have operationalized the amount of local control permitted within states (Carnoy & Loeb, 2002; Hull, 2013), albeit with some differences in individual states.



**Table 2.** Descriptive statistics, measures of local control in NTSG teacher evaluation (higher numbers indicate greater local control) for RTTT and non-RTTT States.

Abbreviated protocol question	Question number	Range	Mean (All states)	SD (All states)	Mean (non-RTTT states, only)	Mean (RTTT states, only)	Effect size of difference**
Who selects measures of student achievement?	6	0 to 2	1.6	1.6	1.7	1.3	0.7
Who approves measures of student achievement?	7	0 to 2	1.6	1.6	1.7	1.3	0.5
Who selects the analytic approach?	12	0 to 2	1.1	1.1	1.3	0.7	0.7
Who aggregates scores?	13	0 to 2	1.3	1.3	1.4	1.2	0.2
May alternate systems be developed?	15	0 to 2	1.2	1.2	1.4	0.7	1.1
Who determines model structure?	17	0 to 1	0.4	0.4	0.5	0.2	0.6
NTSGLCI		1 to 11*	7.6	3.1	8.1	5.5	0.9

\*Although a value of 0 is possible, no states were observed with this value.

\*\*Calculated as follows: effect size of difference = (non-RTTT mean – RTTT mean)/(pooled standard deviation).

be of interest to researchers and policy-makers alike. We also ask: What is the policy-making authority on such matters? Here there are numerous, important philosophical considerations at play, including the trade-off between flexibility and efficiency when allowing for local control. We do not argue that local or state control is good or bad, in and of itself. Rather, we feel that in order for informed decisions to take place around such matters, such as when the provision of flexibility is beneficial and when best practice should be defined, it is important to document and evaluate current practices.

There is essentially no research that illuminates the connection between levels of state control and student outcomes (Wei, 2012). Furthermore, no other study has examined the variability in state approaches to teacher evaluation in NTSG – a category which, despite including the majority of teachers in the United States, has received less attention than teachers who receive a state-supplied growth or value-added score. Because of the difficulties that arise when trying to include measures of student growth in the evaluation of NTSG teachers, and because such challenges are only beginning to be addressed, there is an astounding amount of variation in state approaches. In particular, the extent to which states allow districts to create their own systems as opposed to follow rather rigid state-developed procedures varies considerably across states. In which areas of teacher evaluation might local control be beneficial, and in which areas might a greater or lesser role for the state be warranted? When does flexibility in the use of student growth measures lead to better systems, and when might such flexibility create inefficiencies or lead to poor practice?

These are important questions, and the results from this paper will provide a platform from which future studies can address them.

Although a complex concept such as local control is difficult to operationalize, this study finds very real differences in this regard across states as it pertains to NTSG teacher evaluation. We find considerable range in the NTSGLCI for states, which translates to meaningful, practical differences for districts across the United States. Furthermore, the correlation between NTSGLCI and receiving RTTT funds represents empirical evidence for the connection between federal and state roles in education policy in general, and in teacher evaluation specifically. Although it is not the case that the level of local control is consistent within a state, district or even school, as different standard practices of evaluation exist across subjects and grades even within NTSG, nonetheless compelling trends emerge. This study provides the rare empirical evidence that a federal policy initiative decreased local control; although this is often assumed, it is seldom tested. Specifically, RTTT appears to have led to more prescriptive teacher evaluation systems, including how student achievement measures are used in such systems, for those states that won RTTT awards. Future studies should investigate the implications and consequences of these policies by leveraging the variability in local discretion which they created. For instance, a future study could utilise the NTSGLCI to examine the impact of local control in NTSG teacher evaluation on state-level performance on the National Assessment of Educational Progress (NAEP). However, conversations about these matters must also go beyond student achievement to look at an array of indicators which scrutinise the very nature of public education. As McDonnell (2013) asserts, we should seek to understand the impact of increased federal and state roles 'on public attitudes toward the schools; on how parents, educators, and the public define their responsibility to public education; and on the relationship between information about school performance and trust in government more broadly' (p. 184). In other words, public education is at its heart a public endeavour, with its body of stakeholders reaching far and wide. To this point, a poll found that the public exhibits a general support for local control, but unequivocal support does not always extend to more specific policy areas (Jacobsen & Saultz, 2012). Thus, the public may appreciate a stronger role for federal government, but in nuanced and often narrow ways. It is important that when states restrict local control in the name of efficiency or holding high standards, it is done with explicit intentions and rationale, and in areas where research and practice suggest a lack of capacity at the local level would benefit from a greater state role.

### Notes on contributors

*Douglas J. Gagnon* is currently a research associate at the University of New Hampshire's Carsey School of Public Policy. At the time of this work, he was a researcher at The National Center for the Improvement of Educational Assessment, Inc. He may be reached at [douglas.gagnon@unh.edu](mailto:douglas.gagnon@unh.edu).

*Erika L. Hall* is a senior associate at The National Center for the Improvement of Educational Assessment, Inc.

*Scott Marion* is the executive director of The National Center for the Improvement of Educational Assessment, Inc.



## Notes

1. Question 9 was also examined for inclusion in this index. However, state responses to this question proved difficult to classify. Furthermore, reliability testing found that the operationalized response to question 9 decreased the internal consistency of the index variable. For these reasons, it was not included.
2. Washington DC is excluded from these analyses, given that it is a district and not a state.
3. The number of groups and cut points between these groups could be considered somewhat arbitrary. One may have included only two groups, for instance, or set the cut points differently between high and moderate local control. Ultimately the decisions made with two criteria in mind: (1) preserve that natural groupings that seemed to emerge from the distribution of the NTSGLCI and (2) generate a somewhat-even distribution of states into each group. For example, there was a steep drop off between values of 8 and 9, so that was deemed an appropriate cut for high and moderate categories. The cut between moderate and low categories was made using the latter of the two criteria.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported in part by the Bill & Melinda Gates Foundation [grant number OPP1099484\_2013\_NCIEA\_Results].

## References

- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18). Retrieved from <http://epaa.asu.edu/epaa/v10n18/>
- Baker, B. D., Oluwole, J. O., & Green, P. C. (2013). The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the race-to-the-top era. *Education Policy Analysis Archives*, 21(4/5). Retrieved from <http://epaa.asu.edu/ojs/article/view/1298/1043>
- Braun, H. (2004). Reconsidering the impact of high-stakes testing. *Education Policy Analysis Archives*, 12(1). Retrieved from <http://epaa.asu.edu/epaa/v12n1/>
- Buckley, K., & Marion, S. (2011). *A Survey of approaches used to evaluate educators in non-tested grades subjects*. Dover, NH: Harvard University and National Center for the Improvement of Educational Assessment.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24, 305–331.
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106, 1047–1085.
- Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30, 418. doi:10.2307/23018959.
- Figlio, D., & Loeb, S. (2011). School accountability. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbooks of the economics of education*, Vol. 3 (pp. 383–421). North-Holland: Elsevier B.V.
- Finn, C.E., & Petrilli, M.J. (2013). The failures of U.S. education governance today. In P. Manna, & P. McGuinn (Eds.), *Education Governance for the twenty-first century: Overcoming the structural barriers to school reform* (pp. 21–35). Washington, DC: Brookings Institution Press.
- Goe, L. (2010). *Evaluating teachers' contributions to student learning growth in the non-tested subjects and grades*. Wichita, KS: Educational Testing Service.
- Gong, B., & Marion, S. (2006). *Dealing with flexibility in assessments for students with significant cognitive disabilities (synthesis report 60)*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

- Hanushek, E. A., & Raymond, M. E. (2004). The effect of school accountability systems on the level and distribution of student achievement. *Journal of the European Economic Association*, 2, 406–415.
- Henig, J. R. (2009). The politics of localism in an era of centralization, privatization, and choice. *Yearbook of the National Society for the Study of Education*, 108(1), 112–129.
- Herman, J. L., Heritage, M., & Goldschmidt, P. (2011). *Developing and selecting assessments of student growth for use in teacher evaluation systems*. Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Honig, M. I., & Rainey, L. R. (2012). Autonomy and school improvement: What do we know and where do we go from here? *Educational Policy*, 26(3), 465–495.
- Hull, J. (2013). *Trends in teacher evaluation: How states are measuring teacher performance*. Alexandria, VA: The Center for Public Education.
- Ingersoll, R. M. (2005). The problem of underqualified teachers: A sociological perspective. *Sociology of Education*, 78(2), 175–178.
- Jacobsen, R., & Saultz, A. (2012). The polls – trends. Who should control education? *Public Opinion Quarterly*, 76(2), 379–390. doi:10.1093/poq/nfs017.
- Layton, L. (2015, December 9). Senate overwhelmingly passes new national education legislation. *The Washington Post*. Retrieved from [https://www.washingtonpost.com/local/education/senate-overwhelmingly-passes-new-national-education-legislation/2015/12/09/be1b1f94-9d2a-11e5-a3c5-c77f2cc5a43c\\_story.html](https://www.washingtonpost.com/local/education/senate-overwhelmingly-passes-new-national-education-legislation/2015/12/09/be1b1f94-9d2a-11e5-a3c5-c77f2cc5a43c_story.html)
- Marsh, J. A., & Wohlstetter, P. (2013). Recent trends in intergovernmental relations: The resurgence of local actors in education policy. *Educational Researcher*, 42(5), 276–283. doi:10.3102/0013189X13492193.
- McDermott, K. A. (1999). *Controlling public education: Localism versus equity*. Studies in government and public policy. Lawrence, KS: University Press of Kansas.
- McDonnell, L. M. (2013). Educational accountability and policy feedback. *Educational Policy*, 27(2), 170–189. doi:10.1177/0895904812465119.
- McGuinn, P. (2012). Stimulating reform: Race to the Top, competitive grants and the Obama education agenda. *Educational Policy*, 26(1), 136–159.
- National Center for Education Statistics (NCES). (2012). Digest of education statistics 2012 Number of public school districts and public and private elementary and secondary schools: Selected years, 1869–70 through 2010–11. Retrieved from [https://nces.ed.gov/programs/digest/d12/tables/dt12\\_098.asp](https://nces.ed.gov/programs/digest/d12/tables/dt12_098.asp)
- National Center for Education Statistics (NCES). (2015). *The condition of education 2015: Public school revenue sources*. Washington: DC. Retrieved from [http://nces.ed.gov/programs/coe/pdf/coe\\_cma.pdf](http://nces.ed.gov/programs/coe/pdf/coe_cma.pdf)
- Nee, E. (2010, November 19). Q&A: Joanne Weiss. *Stanford Social Innovation Review*, 15. Retrieved from [http://www.ssireview.org/articles/entry/qa\\_joanne\\_weiss/](http://www.ssireview.org/articles/entry/qa_joanne_weiss/)
- Nunnally, J. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.
- Pennington, K. (2014). *ESEA Waivers and teacher-evaluation plans: State oversight of district-designed teacher-evaluation systems*. Washington, DC: Center for American Progress.
- Prince, C. D., Schuermann, P. J., Guthrie, J. W., Witham, P. J., Milanowski, A. M., & Thorn, C. A. (2009). *The other 69 percent: Fairly rewarding the performance of teachers of non-tested subjects and grades*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education, Discussion paper.
- Resmovits, J. (2015, December 10). Obama signs every student succeeds act, marking the end of an era. *Los Angeles Times*. Retrieved from <http://www.latimes.com/local/education/standardized-testing/la-me-edu-essa-obama-signs-end-no-child-left-behind-20151210-story.html>
- Riley, R. W., & Coleman, A. L. (2011). Turning the page on the equity debate in education: How to give all children a real opportunity. *American Educator*, 35(1), 26–30.
- Sawchuk, S. (2011). Wanted: Ways to assess the majority of teachers. *Education Week*, 30(19). Retrieved from [http://www.edweek.org/ew/articles/2011/02/02/19teachers\\_ep.h30.html](http://www.edweek.org/ew/articles/2011/02/02/19teachers_ep.h30.html)
- Superfine, B. M., Gottlieb, J. J., & Smylie, M. A. (2012). The expanding federal role in teacher workforce policy. *Educational Policy*, 26(1), 58–78.

- Taylor, E. S., & Tyler, J. H. (2012). Can teacher evaluation improve teaching? *Education Next*, 12(4), 78–84.
- Toch, T. (2011). The teacher unions' Odysseus. *Phi Delta Kappan*, 93, 66–67.
- Tucker, M. (2012). A different role for teachers unions. *Education Next*, 12(1), 16–20.
- Wei, X. (2012). Are more stringent NCLB state accountability systems associated with better student outcomes? An analysis of NAEP results across states. *Educational Policy*, 26(2), 268–308.
- Welner, K. G. (2001). *Legal rights, local wrongs: When community control collides with educational equity*. Albany, NY: State University of New York Press.