

Presenting Findings from Measures of
Teacher Effectiveness:
Key Reporting Considerations for States and Districts

Authored by:
Carole Gallagher, Ph.D.
September 20, 2012

**Presenting Findings from Measures of Teacher Effectiveness:
Key Reporting Considerations for States and Districts**

Authored By:

Carole Gallagher, Ph.D.

Copyright © 2012 by WestEd

All rights reserved.

ACKNOWLEDGEMENTS

This white paper grew out of discussions among a group of state accountability directors representing close to twenty states and has benefitted specifically from input and reviews from a number of critical friends. I especially appreciated the targeted guidance from Charlene Tucker (formerly with the Connecticut State Department of Education) during the design, development, and refinement of this work. I also would like to thank Doug Rindone (formerly with the Connecticut State Department of Education), April Zenisky (University of Massachusetts), George Michna (formerly with the Connecticut State Department of Education), Shanthia Washington (Alabama State Department of Education), Pat Roschewski (formerly with the Nebraska Department of Education), Joanne Reihm (Delaware Department of Education), Rich Maraschiello (Pennsylvania Department of Education), Wes Bruce (Indiana Department of Education), Carla Olivares (Michigan Department of Education), Duncan MacQuarrie (formerly with the Washington State Office of Superintendent of Public Instruction), Janice Poda (formerly with the South Carolina State Department of Education), and Pauline Kinsella (New York State United Teachers). The comments, clarifications, and recommendations for improvement they provided helped ensure that the information reported was accurate, timely, and relevant for the target audience. Finally, I would like to acknowledge the research support provided by my WestEd colleague, Patricia Morris.

Table of Contents

Chapter 1: Guidance in Developing Responsible Reporting Practices	1
Purpose.....	3
Comments about Seminal Resources Cited	3
Organization	4
Chapter 2: Key Features of Comprehensive Reports that Present Findings from Measures of Teacher Effectiveness	5
A. Purpose and Target Audience(s).....	5
B. Measures from Which Results Are Reported	6
C. Scoring, Rating, and/or Performance Levels	7
D. How Scores Are Calculated or Performance Rated, with Criteria Used During Decision Making	8
E. Interpretation of Results	9
F. Use of Report or Database	11
Chapter 3: Teacher Effectiveness Reporting Practices in Seven Jurisdictions	12
A. Colorado	12
B. Indiana	16
C. New York City.....	18
D. North Carolina	32
E. Pennsylvania	35
F. Tennessee	36
G. District of Columbia	38
Chapter 4: Legal Considerations.....	40
Types of Laws and Policies That Affect Reporting.....	40
Case Studies	41
Chapter 5: Conclusions and Implications for Future Research	47
Key Issues for Further Exploration	48
References.....	49
Appendix A: Sample Teacher Effectiveness Reports from Selected Jurisdictions	52
Appendix B. Summary of Guidance on Reporting From Seminal Resources	54
Appendix C: Key Features of Educator Effectiveness Reports	63
Appendix D. Legislative and Policy Considerations that May Affect Reporting Practices.....	64

Chapter 1: Guidance in Developing Responsible Reporting Practices

Under the No Child Left Behind Act of 2002 (NCLB), states receiving Title I funding are required to report specific information about teacher qualifications via public report cards and accountability workbooks for peer review. This information includes the number of teachers in each district who (a) are fully certified and/or licensed by the state, (b) hold at least a bachelor's degree from a four-year institution, and (c) demonstrate competence in each core academic subject in which they teach. These requirements are aimed at holding state education agencies (SEAs) accountable for improving student learning outcomes through documentation of teachers' qualifications and monitoring of the distribution of highly qualified teachers across each state.

Under President Obama's administration, the focus of federal requirements has shifted *away* from documenting teachers' qualifications and *toward* examining teachers' instructional effectiveness. States that have received Race to the Top funding and/or have been granted flexibility in relation to particular NCLB requirements are required to annually measure and report on the effectiveness of their teachers. In doing so, states can select models or methods that most effectively meet their needs and context, but a measure of student academic growth must be one component.¹ At a minimum, they must report results in aggregate in their state accountability workbooks. While Elementary and Secondary Education Act (ESEA) reauthorization discussions are still underway, it appears likely that states receiving certain types of federal funding will be required to develop and implement plans that show they are holding their teachers accountable for contributing to students' annual academic growth.²

Scant attention has been paid to the issues that these agencies face in determining how best to *report* these results. As they consider how, when, and to whom to report findings from measures of teacher effectiveness, SEAs and local education agencies (LEAs) will want to consider the rights, needs, and diverse interests of many different stakeholder groups. Opportunities and challenges have emerged for those responsible for collecting and reporting this information, not only in determining what to report to whom, but also in identifying the optimal reporting format (Colmers, 2007; Jaeger, 2003). Also to be considered are the myriad of professional, legal, and political implications associated with each strategy.

The situation for states and other jurisdictions seeking guidance is made murkier by virtue of two conditions. First, states are in very different places or stages in terms of reporting on measures of teacher effectiveness. Research for this paper revealed significant diversity across states in terms of how they currently are or plan to report information about teacher effectiveness. Some are following a determined course of action and already are reporting findings at some level, while others are in an early planning or pilot stage. Even those states with reporting plans in place frequently have only internal, nontransparent, or as-yet undefined reporting policies and structures that are not easily characterized (Bellwether Education Partners, 2011; National Comprehensive Center on Teacher Quality

¹ Findings emerging from the Bill & Melinda Gates Foundation (BMGF)-funded Measures of Effective Teaching project suggest that two measures are particularly useful tools when used in conjunction with student growth: (1) classroom observations conducted by a school administrator using a research-supported rubric, and (2) well-designed student surveys (BMGF, 2010; 2012).

² Current and emerging federal directives include expectations for the evaluation of the effectiveness of principals, as well as teachers. While the focus of this paper is K-12 teachers, we expect that much of the guidance presented will be applicable to reporting findings from measures of principal effectiveness.

[TQ Center], 2012). Second, those states that are reporting on teacher effectiveness have adopted different approaches. A few have policies that require standardized reporting across all schools, while most of the others leave decision making about what and how to report entirely to district discretion. Reports developed at the district level vary widely, and may include a range of practices:

- sharing observation comments or survey results during a private conversation between a teacher and school administrator;
- sending student-specific letters to parents to inform them about their child’s teacher’s effectiveness rating;
- maintaining interactive, password-protected websites that allow different levels of access to different stakeholder groups and that may enable teachers to dig more deeply into their own performance ratings; and
- publicly releasing individual teacher names and effectiveness ratings and/or rankings, as compared to their peers.

For illustrative purposes, it may be helpful to consider scenarios in which different types of reports of teacher effectiveness may be more useful than others.

Scenario 1

A district decides to report school-specific percentages of effective and ineffective teachers via a publicly available annual report or school report card. These aggregate data will be of particular interest to school and district administrators, parents contemplating a move to this community, and policymakers. But parents with children in this district who are viewing these data may be more interested in a different type of report, one that tells them specifically whether their child’s teacher is rated as effective.

Scenario 2

A state policy requires districts to report directly to parents a composite score derived from multiple measures of effectiveness at the teacher level. Parents appreciate such specificity because they now have information directly linked to their child’s classroom teacher(s). But parents who learn that their child has been assigned to a classroom with a teacher who is found to “need improvement” may need more information to help them understand what the rating is based on and the relationship between those criteria and the needs of their child.

Scenario 3

Findings collected annually from formal observations may be used only by a teacher and building principal as feedback that can be used formatively for annual goal setting and during decision making about professional development opportunities. But this teacher’s colleagues have noted that her students tend to perform poorly on state assessments in subsequent grades. This teacher’s peers wish that she attended more closely to her students’ performance on benchmark and end-of-year tests.

Scenario 4

Using value-added analyses of students’ end-of-grade test scores, researchers estimate an effectiveness score for each teacher in a large urban district and provide the local newspaper with a rank ordered list (most to least effective) of teachers based on this one criterion. Without a guide to interpretation, this information is easily misinterpreted—or *over*interpreted—by readers.

States clearly have many issues to consider as they weigh tradeoffs, wrestle with public and private rights and interests, and commit to a plan for reporting results from measures of teacher effectiveness. While there are no easy answers to this dilemma, research- and best practice-based guidelines suggest that attending to certain considerations early on in this process increases the likelihood that outcomes

will be intentional and defensible. Such considerations may help states and districts develop strategies for fairly and clearly reporting information about a teacher’s instructional effectiveness that can be used to improve teaching and learning.

Purpose

This paper is intended to support states and other jurisdictions at all stages of planning and development as they move forward with implementing systems for reporting on measures of teacher effectiveness. In this forum, we seek to raise important questions, to provide research- and best-practice-supported guidelines for several features of a comprehensive teacher effectiveness report, and to share illustrative examples from jurisdictions that have developed reporting tools and strategies.³

Specifically, we address important questions about the following report features:

- Purpose and target audience(s);
- Measures from which results are reported;
- Scoring, rating, and performance levels;
- How the score was calculated and/or performance rated (e.g., criteria used);
- Interpretation of results; and
- Use of report or database.

Comments about Seminal Resources Cited

While guidelines specific to reporting findings from measures of teacher effectiveness—not traditional teacher evaluation data—are just emerging, much can be learned from lessons learned in other domains. Critical resources exist that provide general recommendations for effective reporting practices. These include the following:⁴

1. *Code of Ethics* (American Educational Research Association [AERA], 2011);
2. *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], the American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999; also draft version posted for review and comment, 2011);
3. *Graduate Record Examination [GRE] Guide to the Use of Scores* (Educational Testing Service [ETS], 2011);
4. *NAEP Validity Studies: Reporting the Results of the National Assessment of Educational Progress* (Jaeger, 2003);
5. *Information Quality Guidelines* (U.S. Department of Education [ED], n.d.);
6. *Standards and Assessments Peer Review Guidance* (ED Office of Elementary and Secondary Education, 2009);
7. *Public Reporting and Transparency* (Colmers, 2007);
8. *Student Test Score Reports and Interpretive Guides: Review of Current Practices and Suggestions for Future Research* (Goodman & Hambleton, 2004); and
9. *Guidelines for Effective Score Reporting* (Aschbacher & Herman, 1991).

³ See Appendix A for a list of sample reports and links to each.

⁴ See Appendix B for a summary of key recommendations from each.

In combination, these resources provide useful, relevant, and timely considerations for states that are in the process of formulating strategies for reporting teacher effectiveness data. They suggest what types of data to include and the appropriate level of detail. They help developers consider the audience of the report and how best to convey information to that audience via text and graphic displays. They remind reporters of their responsibilities and offer cautions. Together, the resources send a common message: educational agencies should be proactive in determining a strategy and framework for reporting information of interest to their constituents. Though states may not yet know what types of data and at what levels to report, or in what form to report them, they can engage in thoughtful planning around their stakeholders' needs and priorities for state-specific, customized reporting of data from measures of teacher effectiveness.

Organization

This paper is organized as follows:

- Chapter 1: Guidance in Developing Responsible Reporting Practices
- Chapter 2: Key Features of Comprehensive Reports that Present Findings from Measures of Teacher Effectiveness
- Chapter 3: Teacher Effectiveness Reporting Practices in Seven Jurisdictions (Colorado, Indiana, New York City, North Carolina, Pennsylvania, Tennessee, and District of Columbia)
- Chapter 4: Legal Considerations
- Chapter 5: Conclusions and Implications for Future Research
- References
- Appendix A: Sample Educator Effectiveness Reports from Selected Jurisdictions
- Appendix B: Guidance on Reporting from Seminal Resources
- Appendix C: Key Features of Educator Effectiveness Reports
- Appendix D: Legislative and Policy Considerations that May Affect Reporting Practices

Chapter 2: Key Features of Comprehensive Reports that Present Findings from Measures of Teacher Effectiveness

Measures of teacher effectiveness are tools intended to help stakeholders better understand the strengths and limitations of teachers in preparing their students for the academic challenges of schooling at the next grade or level. Findings from these measures can be used for different purposes and may be intended for different target audiences. For example, findings may be used to provide feedback to teachers that can be used formatively to improve instruction, but they also may be used for high-stakes purposes, such as classroom- or school-level accountability or decision making about hiring, firing, and promotion. As a result, reporting these findings requires adherence to guidelines that will ensure that the information communicated to teachers, administrators, policymakers, parents, and the general public is complete, accurate, trustworthy, and delivered in a timely manner.

As discussed in Chapter 1, a number of seminal documents provide general guidance to support states and other jurisdictions with implementation of responsible reporting practices. In this chapter, the key features of a high-quality report are identified and discussed. These include the following:⁵

- A. purpose and target audience(s);
- B. measures from which results are reported;
- C. scoring, rating, and/or performance levels;
- D. how the score was calculated or performance rated and the criteria used;
- E. interpretation of results; and
- F. use of report or database.

Following a description of each, elements of reports currently used by states or districts for reporting on measures of teacher effectiveness will be highlighted.⁶

A. Purpose and Target Audience(s)

Guiding Questions:

1. *Why is a report needed? What purpose does it serve?*
2. *What does the report explain?*
3. *Who is the target audience for this report? With whom will information be shared? Who will have access to reports?*
4. *What reporting format (e.g., paper copy or computer-supported database) and display features (e.g., data tables and graphics) best address these needs?*

Reports are developed for a number of reasons. States and districts may be required by law or regulation to report findings about teachers' effectiveness. They also may seek to provide feedback annually to each teacher based on classroom observations. In either case, the purpose of the report should be transparent, with a clear statement that alerts readers to the developers' intent (AERA et al., 1999; Goodman & Hambleton, 2004). In general, these purposes can be grouped by the stakes associated with how the reports will be used (Gallagher, Rabinowitz, and Yeagley, 2011).

⁵ See Appendix C for a summary of these features of reports used for low-, moderate-, and high-stakes purposes.

⁶ See Appendix A for links to sample reports.

- A *low-stakes* report may be used as a professional development tool to facilitate teacher growth or to initiate school-level conversations about school improvement. These reports are intended to be used formatively to improve instructional practices.
- A *medium-stakes* report may be used to guide decision making about the value of a particular initiative or intervention; it also may inform hiring or promotion decisions. The purpose may be to explore a research question, such as investigating the regions or districts with high concentrations of teachers who are deemed effective, or examining the characteristics of teacher preparation programs at which effective teachers are trained.
- A *high-stakes* report might be used for accountability purposes at the classroom, school, district, state, and/or federal level or to inform decisions about incentive pay, tenure, or termination. This type of report usually includes a summative score or judgment about a teacher’s performance that can be used in comparison with other teachers (norm-referenced) or in relation to a policy-based standard for performance (criterion-referenced).

The purpose of the report is closely linked to the target audience. Developers will want to focus on both the purpose and the target audience when determining the content that will be presented and the format in which it will be presented. An effective and useful report takes into consideration what the target audience will find to be of interest and tailors the information presented—including the level of technical language—to meet its needs (Aschbacher & Herman, 1991; Jaeger, 2003).

In many cases, states and districts may find that they need to develop different reporting strategies (e.g., bulletins, manuals, checklists, databases, and interactive websites) for different audiences (Aschbacher & Herman, 1991; Jaeger, 2003). Seeking input and/or feedback from key audiences can help developers verify their assumptions about the clarity, credibility, and usefulness of different reporting strategies. Finally, the reasons for reporting findings from measures of teacher effectiveness may change over time, thereby triggering the need to re-think the target audience and re-evaluate the content and/or format to ensure they remain appropriate for the purpose intended.

B. Measures from Which Results Are Reported

Guiding Questions:

1. *Which measures of teacher effectiveness system are used in reporting?*
2. *How was this measure identified, administered, and validated? What evidence supports claims that this measure is appropriate for the purpose of measuring teacher effectiveness?*

The research community has emphasized that states and districts should plan and implement systems that include multiple measures of teacher effectiveness (BMGF, 2010; 2012; Gallagher et al., 2011; TQ Center, 2011; New Teacher Project, 2010a, 2010b). For teachers in tested grades and subjects, the systems may include student growth estimates that use performance on annual standardized tests or estimates from teacher value-added models. For teachers in non-tested grades and subjects, the systems may include vendor-developed measures, customized end-of-year or end-of-course examinations, or pre-post measures. For all teachers, the set of measures may include surveys, observation protocols, performance assessments, or results from content area assessments.

Reports may include information about the steps taken during development to ensure content validity and alignment to the standards for performance. Mode of administration may be reported. Comprehensive reports also may include a theory of action or theoretical framework that links each measure to the scores

or judgments assigned to teachers. In addition, presenting evidence of predictive validity in these reports can help readers judge the usefulness of the measures for the purpose of measuring teacher effectiveness.

Report readers will have questions about the quality and appropriateness of the measures, methods, and tools used to determine teacher effectiveness. They will expect the report to identify each measure, provide sufficient information from which to draw conclusions about the degree to which the data or findings presented are reliable and accurate, and the validity of each measure for the purpose intended (Glazerman, Goldhaber, Loeb, Raudenbush, Staiger, & Whitehurst, 2011; Goodman & Hambleton, 2004; Jaeger, 2003; National Council on Teacher Quality [NCTQ], 2011; Zenisky & Hambleton, 2012). Report developers should strive to include evidence of the technical adequacy of each measure, such as details about field testing, internal consistency coefficients, degree of precision (or measurement error) associated with scores or judgments, classification consistency data, and findings about fairness from bias and sensitivity reviews. Yet developers also must be mindful of the purpose and target audience of the report so the amount and type of technical information provided is appropriate. For this reason, many jurisdictions may find that they need to develop multiple reports, each tailored to the unique needs and interests of the intended user. As when reporting results from annual end-of-grade assessments, a reasonable practice would be to ensure that a separate technical report or interpretive guide is accessible to report users who seek more detailed information about methodology, analyses, or ratings.

C. Scoring, Rating, and/or Performance Levels

Guiding Questions:

1. *What performance outcome is reported (e.g., a change or gain score, performance level, checklist summary, or descriptive information)? Is this outcome familiar to the target audience?*
2. *At what level (individual vs. aggregate [subgroup, school, district, etc.]) are results reported? What evidence supports reporting at this level?*
3. *Does the report format support the presentation of findings in a clear, attractive, and interpretable manner?*

A key element of any report is a listing of the scores or other judgments (e.g., performance levels or qualitative descriptions) assigned to each teacher or group of teachers (e.g., all those in one school or in one subject area or grade). These data may include test scores (in aggregate form or disaggregated at the individual or teacher group levels), percentile ranks, score ranges, performance level designations, and/or descriptive information. Recommendations include providing group or subgroup sizes (Ns), score ranges, measures of central tendency, confidence intervals or error bands in which the educator's "true" score or rating is expected to be located, and effect sizes to guide interpretation about the practical importance of differences in scores (AERA et al., 1999; ETS, 2011; Jaeger, 2003; Zenisky & Hambleton, 2012).

While providing a depth and breadth of data results and contexts for those data results is helpful, an argument can be made for limiting information to only what will be useful and easily interpretable by the intended audience. Substantial guidance has been shared by experts in data display. Such guidance includes attending to the amount of information presented and apportioning space according to the importance of the ideas; using visual techniques, such as graphics and colors; and incorporating headings and sidebars (Aschbacher & Herman, 1991; Goodman & Hambleton, 2004; Jaeger, 2003; Wainer, Hambleton, and Meara, 1999). Tables, graphs, and charts are frequently included as user-friendly formats

for quantitative data. Developers are advised to make the report readable, clear, and uncluttered, and to use strategies for presenting the information in a visually attractive and interpretable manner, but they also should include sufficient technical information to support valid inferences and replicability of findings; attaining this “just right” balance is always a challenge (Goodman & Hambleton, 2004; Zenisky & Hambleton, 2012).

NCES experts provide guidance about reporting quantitative and qualitative data. Their reports can help developers determine the number or categories of data that should be included and how the chosen numbers/categories should be visually depicted (Jaeger, 2003). Other researchers recommend presenting only a small number of categories, using summaries, and providing rich descriptions that help paint a picture (Aschbacher & Herman, 1991) and following universal design principles so that the report will be maximally accessible to wide audiences (Forte Fast & ASR SCASS, 2002).

D. How Scores Are Calculated or Performance Rated, with Criteria Used During Decision Making

Guiding Questions:

1. *What methods are used to calculate scores or rate performance? Are scores accompanied by clear statements about the degree of measurement error associated with each?*
2. *If a composite score is reported, what is the basis and rationale for arriving at the composite?*
3. *If a weighting strategy is used, what evidence supports use of that process?*

Report readers will have questions about how the measures were scored and how teachers were judged and categorized based on those data. They will expect to see scoring protocols, rubrics, scales, descriptions of the performance levels, and information about the standards or criteria against which each teacher was judged. States and other jurisdictions responsible for developing reports will want to ensure that the methodology (or formulae) for deriving scores, judgments, or performance levels is transparent, that steps in the process were documented, and that analyses are replicable.

Many states and districts intend to or are implementing a system that includes multiple measures. If they seek to combine subscores from different measures into a composite, the report will want to provide information about the methodology used, i.e., how each individual measure contributed to the composite (e.g., subscore weighting strategy), and if the model used was compensatory or conjunctive (Burling, 2012; ED, 2004; Zenisky & Hambleton, 2012).⁷ Reporting inter-rater reliability statistics and information about standard setting are encouraged (NCTQ, 2010).

⁷ In compensatory scoring, all of the subscores that contribute to a comprehensive effectiveness rating are combined in such a way that a teacher who scores low on one aspect of his or her evaluation composite (e.g., value-added estimate) will not necessarily receive a low rating overall, if he or she receives high scores on other components of his or her evaluation composite (e.g. observation rating). This, of course, depends largely on how each component is weighted—if, for example, student value-added test scores comprise 65 percent of a teacher’s final rating, while observation ratings make up 40 percent, it is possible for the teacher’s comprehensive score to remain low. Conjunctive scoring, on the other hand, treats each subscore of a comprehensive rating as its own distinct element, and each subscore must indicate proficiency or better in order for a teacher’s comprehensive rating to be at the proficient level. If one subscore indicates below-proficient performance, even where other scores indicate proficiency or better, the teacher’s comprehensive rating will be below proficient.

Peer Review guidance (ED, 2004) encourages states to report subscores for different domains (i.e., to report more than a single composite rating). This recommendation has relevance for those states and districts using multiple measures, as each measure may produce a subscore that then may (or may not) become part of a composite (see Section D of this paper). Specific guidance about reporting subscores is provided in the Joint Standards:

When interpretation of subscores, score differences, or profiles is suggested, the rationale and relevant evidence in support of such interpretation should be provided.” (AERA et al., p. 20)

Strategies for reporting subscores responsibly have been studied extensively (e.g., see Goodman & Hambleton, 2004; Haladyna & Kramer, 2004; Ryan, 2006; Sinharay, Puhan, and Haberman, 2011). Report developers are expected to follow specific rules when reporting at the subscore level and to present evidence to support the validity of inferences drawn from results reported at that level (e.g., reliability coefficients, intercorrelations, findings from factor analyses, analyses of variance, and/or alignment studies).

States and districts will want to consider strategies for tailoring the amount and type of information presented to fit the level of understanding of the target audience; for example, if the target audience is parents, statistical jargon should be minimized and key terms defined (Goodman & Hambleton, 2004).

E. Interpretation of Results

Guiding Questions:

1. *Does the report enable accurate interpretation of results by the target audience and other stakeholders?*
2. *What rationale (e.g., a comprehensive summary of evidence and theory bearing on the intended interpretation) supports the recommended interpretation? Were alternative explanations for the teacher’s performance considered?*
3. *If comparing results at the group level, what relevant contextual information is presented to explain group differences (where possible) and enable meaningful interpretation of these differences?*
4. *Is any supplemental information provided that will minimize possible misinterpretations of the data?*

Findings from measures of teacher effectiveness are challenging to interpret (Corcoran, 2010). Without guidance—as well as cautions—readers from any stakeholder group can misinterpret or over-generalize findings (AERA, 2011). Even when developers strive to minimize technical jargon, results need to be accompanied by explanatory text that provides context for the findings and suggests appropriate interpretations (ETS, 2011; Zenisky & Hambleton, 2012). In many cases, general information can be shared in a shorter report if an interpretive guide is readily available that provides more detailed information about the measures used and the evidence to support inferences drawn from results (AERA et al., 1999; ETS, 2011; Goodman & Hambleton, 2004).

According to the Joint Committee on Standards (Standard 5.10):

When test score information is released to students, parents, legal representatives, teachers, clients, or the media, those responsible for testing programs should provide appropriate

interpretations. The interpretations should describe in simple language what the test covers, what the scores mean, and how the scores will be used (AERA et al., 1999, p. 65).

As detailed previously, well-designed reports support appropriate interpretation of results by describing the knowledge, skills, and abilities that were measured; how information was collected, scored, and analyzed; and how and by whom results are intended to be used (Aschbacher & Herman, 1991; Koretz & Diebert, 1993). Well-designed reports provide guidance about appropriate comparisons and help readers reach decisions about whether individual and group or subgroup differences (or changes over time) are meaningful. They may include tests of significance and effect sizes so that readers understand the statistical and practical importance of noted differences (ED, 2004). Known limitations of the measures used to generate results should be acknowledged so that the findings are kept in perspective. Similarly, a precise description of the teachers who were included (Which grades? Which subject areas?) and excluded should be part of any statement about the findings. That is, to ensure that limits to generalizability are explicit, a high-quality report might repeat a qualifying statement, such as the following, in all sections of the report:

- “Findings are reported for only a small subset of teachers in this state, those who taught reading or mathematics to students in grades 5-8 during a specific three-year period.”
- “Effectiveness scores, or estimates, were calculated using students’ scores over time on the state’s annual standardized assessment in reading and mathematics; no other measures of effectiveness were included in these analyses.”
- “Performance levels were assigned based on information collected during formal classroom observations conducted on three occasions by a school administrator.”
- “The score reported describes an estimate of instructional effectiveness for one time period. If this teacher were evaluated again, it is likely that his/her rating would fall into the range provided.”

Responsible report developers seek to reinforce to readers that (a) the findings come from specific measures or sources; (b) these measures contain known limitations; (c) a teacher’s “true” effectiveness is likely to be in a range around the effectiveness rating reported; and (d) the evidence presented in support of the claims made should be examined carefully to ensure that inferences drawn from these results are valid.

Educators in **Delaware**, the **District of Columbia**, **New York City**, **Tennessee**, and **Pennsylvania** receive comprehensive guidelines for report use that include a Frequently Asked Questions (FAQ) section and a glossary of terms.

F. Use of Report or Database

Guiding Questions:

- *Does the report include guidance about the ways in which reported data are intended to be used?*
- *Is rationale for each recommended use of this report provided?*
- *Are findings released to the public and/or policymakers? If so, is a plan in place for providing them with supplemental information that will promote responsible use of data?*
- *Are teacher evaluation results reported in a way that decreases the likelihood of inappropriate decisions or actions based on findings? Have potential unintended consequences of reporting these data been considered and addressed?*
- *Does a state-level mandate or ruling dictate the types of information that must be disclosed and/or specify the types of information that cannot be shared or reported publicly?*

Until recent years, what, how, and to whom information about teacher quality was reported was largely the domain of individual districts. But federal funding and other initiatives have encouraged states to build comprehensive systems for measuring and monitoring teacher effectiveness, with increasing pressure to report findings from these systems to different stakeholder groups. States now bear responsibility for implementing policies and/or disseminating best-practice guidelines to ensure responsible use of data generated from these systems (NCTQ, 2011). Steps taken to ensure data security and protection of individual's rights to privacy must be balanced with state policies regarding the public's right to access information—even if it is potentially sensitive—about its teachers (AERA, 2011; ED, n.d.). This concern is discussed at greater length in Chapter 3 of this paper, with research-supported guidelines to help states and districts develop short- and long-term plans for monitoring the consequences (positive and negative, intended and unintended) of their reporting practices.

Helpful advice born from lessons learned and years of research is available from those who report findings from the ETS-developed GRE (ETS, 2011). Strategies for effective reporting of scores and indicators of level of performance have been studied extensively. ETS's experiences can support states and districts in understanding how report developers can remain alert to cases of alleged misuse. Please see the References section of this paper for more information.

Finally, Colmers (2007) provides suggestions based on lessons learned from the field of healthcare that have relevance to this context. Specifically, as do healthcare researchers, states and other jurisdictions must grapple with the intersection of confidentiality and the public's right to information and carefully weigh tradeoffs when deciding when, how, and to whom they report findings. Colmers highlights that a key benefit of public reporting is helping stakeholders make informed choices about key issues that impact their futures. See the References section at the end of this paper for a link to more information on Colmers's report.

Chapter 3: Teacher Effectiveness Reporting Practices in Seven Jurisdictions

In this chapter, actual reports or screen shots from publicly accessible websites are presented. These are intended to provide readers with examples of the strategies that jurisdictions from across the country are using to report findings or results from measures of teacher effectiveness. They also are intended to illustrate the broad range of the *types* of information communicated to stakeholders as well as the format in which it is reported. While none of these examples incorporates *all* of the principles of effective reporting practices discussed in Chapter 2, each provides readers with a concrete example of a report that describes findings or results from measures of teacher effectiveness. Sample reports from the following jurisdictions are included:

- Colorado
- Indiana
- New York City
- North Carolina
- Pennsylvania
- Tennessee
- District of Columbia

A. Colorado

Several sections from the Colorado Department of Education's (CDE's) *Draft Rubric for Evaluating Colorado's Teachers* (April 10, 2012) are provided below. The target audiences for this report are teachers (to ensure they are informed about the criteria on which they will be judged) and the school or district administrators who will conduct evaluations and share results with teachers. CDE's teacher effectiveness system addresses five standards:

- I. Mastery of and Pedagogical Expertise in the Content They Teach
- II. Safe, Inclusive, and Respectful Learning Environment for Diverse Population of Students
- III. Effective Instruction and an Environment that Facilitates Learning
- IV. Reflection on Practice
- V. Leadership

The first section of the full report for one standard, Standard I, is provided below. Note that this introduction includes a statement of purpose (see second paragraph that describes how the rubric is intended to be used). This rubric can be used as a checklist for rating the teacher on specific behaviors and certain student-level outcomes that are associated with one facet of Standard I, *Element a*.

Effective Teachers in the state of Colorado have the knowledge, skills, and commitments needed to provide excellent and equitable learning opportunities and growth for all students. They strive to support growth and development, close achievement gaps and to prepare diverse student populations for postsecondary and workforce success (See Appendix A). Effective Teachers facilitate mastery of content and skill development, and employ and adjust evidence-based strategies and approaches for students who are not achieving mastery and students who need acceleration. They also develop in students the skills, interests and abilities necessary to be lifelong learners, as well as for democratic and civic participation. Effective Teachers communicate high expectations to students and their families and utilize diverse strategies to engage them in a mutually supportive teaching and learning environment. Because effective Teachers understand that the work of ensuring meaningful learning opportunities for all students cannot happen in isolation, they engage in collaboration, continuous reflection, on-going learning and leadership within the profession.

The Teacher Quality Standards outline the knowledge and skills required of an effective Teacher and will be used to evaluate Teachers in the state of Colorado. All School Districts and BOCES shall base their evaluations of licensed classroom Teachers on the full set of Teacher Quality Standards and associated detailed Elements included below, or shall adopt their own locally developed standards that meet or exceed the Teacher Quality Standards and Elements.

Observation	Quality Standard I: Teachers demonstrate mastery of and pedagogical expertise in the content they teach. The elementary Teacher is an expert in literacy and mathematics and is knowledgeable in all other content that he or she teaches (e.g., science, social studies, arts, physical education, or world languages). The secondary Teacher has knowledge of literacy and mathematics and is an expert in his or her content endorsement area(s).				
	Not Evident	Partially Proficient	Proficient (Meets State Standard)	Accomplished	Exemplary
	Element a: Teachers provide instruction that is aligned with the Colorado Academic Standards; their District's organized plan of instruction; and the individual needs of their students.				
✓	<p>The Teacher:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Rarely plans instruction aligned to Colorado Standards, their District's organized plan for instruction, or student needs. <input type="checkbox"/> Uses instructional objectives that are inappropriate for some students. 	<p>The Teacher:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Uses the Colorado Standards, the district's plan of instruction, and individual student needs as the foundation for lesson plans. 	<p>... and</p> <p>The Teacher:</p> <p>Aligns instruction with:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Student learning objectives. <input type="checkbox"/> District plan for instruction. <input type="checkbox"/> Colorado Standards. <input type="checkbox"/> Needs of students. <input type="checkbox"/> Collaborates with other school staff to vertically and horizontally articulate the curriculum. 	<p>... and</p> <p>Students:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Move to the next curriculum level or next higher course in sequence without remediation. <input type="checkbox"/> Interact with the rigorous and challenging content in meaningful ways and without becoming frustrated. 	<p>... and</p> <p>Students:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Discuss with their Teacher, families, and significant adults any potential gaps in their learning in order to address them.

The next section of the comprehensive report summarizes ratings for all facets of Standard I, *Elements a-f*. Report users can see the criteria on which they were judged, their ratings for each, and an explanation of how their Overall Rating for Standard I was calculated. This form also provides space for written comments from the evaluator. Finally, this section of the report lists the types of evidence that were considered during decision making.

	Ratings (# Points per rating at this level)	NE	PP	P	A	E	Total Points
		(0)	(1)	(2)	(3)	(4)	
a.	Provides instruction that is aligned with the Colorado Academic Standards; their District's organized plan of instruction, and the individual needs of students.						
b.	Demonstrates knowledge of student literacy development in reading, writing, speaking and listening.						
c.	Demonstrates knowledge of mathematics and understands how to promote student development in numbers and operations, algebra, geometry and measurement, and data analysis and probability.						
d.	Demonstrates of the content, central concepts, tools of inquiry, appropriate evidence-based instructional practices and specialized character of the disciplines being taught.						
e.	Develops lessons that reflect the interconnectedness of content areas/disciplines.						
f.	Makes instruction and content relevant to students and takes actions to connect students' background and contextual knowledge with new information being taught.						
0 to 5 Total Points = Not Evident 6 to 9 Total Points = Partially Proficient 10 to 14 Total Points = Proficient 15 to 19 Total Points = Accomplished 20 to 24 Total Points = Exemplary		Overall Rating for Standard I: <div style="border: 1px solid black; height: 20px; width: 100%; background-color: #cccccc;"></div>					
Evaluator Comments (Required for Ratings of "Not Evident" or "Partially Proficient" and recommended for all rating levels). Please indicate the element for which the comment applies if not for the standard as a whole.							
Comments of person being evaluated. (Optional)							

Examples of Artifacts that may be used to provide evidence of performance:	Evidence of performance provided by artifact:
<input type="checkbox"/> Student Achievement Data	
<input type="checkbox"/> Student feedback	
<input type="checkbox"/> Parent feedback	
<input type="checkbox"/> Lesson plans/units of study	
<input type="checkbox"/> Feedback from walkthrough observations	
<input type="checkbox"/> Instructional activities schedules	
<input type="checkbox"/> Student journals/learning logs	
<input type="checkbox"/> Student work	

Colorado evaluators create similar reports for each of the five standards. They then complete a summary document that combines ratings for each standard and the total number of points for all standards. Note that this section of the report provides a “translation” to support teachers in interpreting findings from their evaluation.

Determining the Overall Rating for Professional Practices		
Standard	Rating for the Standard	Number of Points for the Rating
I. Mastery of and Pedagogical Expertise in the Content They Teach		
II. Safe, Inclusive and Respectful Learning Environment for Diverse Population of Students		
III. Effective Instruction and an Environment that Facilitates Learning		
IV. Reflection on Practice		
V. Leadership		
Total Points for All Standards		

Translating the Total Points for All Standards to Overall Professional Practices Rating		
Total Number of Points Received	Rating for Number of Points Received	Total Number of Points Received for this Evaluation = <input type="text"/>
0 to 3 Points	Not Evident	Overall Professional Practices Rating = <input type="text"/>
4 to 7 Points	Partially Proficient	
8 to 12 Points	Proficient	
12 to 16 Points	Accomplished	
17 to 20 Points	Exemplary	

B. Indiana

The Indiana Department of Education has taken a different approach. In a presentation to teachers that explains the state’s new teacher evaluation system, the following information was shared.

The summative rating reflects a teacher’s professional practice and contribution to student growth.

1) Professional Practice – Assessment of instructional knowledge and skills
Measure: Indiana Teacher Effectiveness Rubric (TER)

2) Student Learning – Contribution to student academic progress
Measure: Individual Growth Model (IGM)*
Measure: School-wide Learning Measure (SWL)
Measure: Student Learning Objectives (SLO)

* Only teachers in grades 4-8 ELA/Math have individual growth model data

- Each of these measures is scored separately and combined for the summative rating.



If you have received this document from any source other than the RISE website, it may have been altered from its original version. For the official, and most up-to-date version, please visit www.RISEIndiana.org.

Evaluators use the Teacher Effectiveness Rubric to rate a teacher **at the end of the year** using a four step process.

1) Professional Practice – Assessment of instructional knowledge and skills
Measure: Indiana Teacher Effectiveness Rubric (TER)

- 1 Compile ratings and notes from multiple observations, drop-ins, and other sources of evidence
- 2 Use professional judgment to establish three, final ratings in Planning, Instruction, and Leadership

Example

Competency	1.1	1.2	1.3	1.4	1.5
Teacher’s Rating	3	2	2	3	3

Use Professional Judgment

Final Domain 1 Rating: **3**

Competency ratings based on notes from observations, conferences and other sources of evidence.



If you have received this document from any source other than the RISE website, it may have been altered from its original version. For the official, and most up-to-date version, please visit www.RISEIndiana.org.

The last two steps convert domain ratings to a final, overall Professional Practice rating

3

Use established weights to roll-up three domain ratings into one rating for Domains 1-3



4

Multiply each domain rating by its designated weight. Add up the weighted ratings. Subtract one point if the teacher did not meet all of the core professionalism expectations

Example

	Rating (1 - 4)	Weight	Weighted rating
Domain 1: Planning	3	X 0.10	0.30
Domain 2: Instruction	2	X 0.75	1.50
Domain 3: Leadership	3	X 0.15	0.45
			Total = 2.25
All professionalism expectations met:			- 0
			Final TER Score: 2.25



If you have received this document from any source other than the RISE website, it may have been altered from its original version. For the official, and most up-to-date version, please visit www.RISEIndiana.org.

Weighting Example

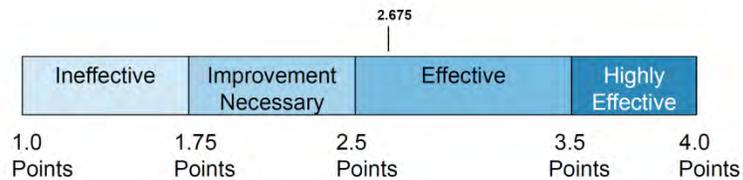
- Mrs. Smith teaches three sections of 8th grade ELA and three sections of 8th grade Social Studies.
- Because half or more of her classes taught have individual growth model data, she is a **Group 1** Teacher.
- We use the **Group 1** weights from the previous slide to calculate her summative score.

Component	Raw Score	Weight*	Weighted Score
Teacher Effectiveness Rubric	2.25	x 50%	= 1.125
Individual Growth Model Data	3	x 35%	= 1.05
Student Learning Objectives	4	x 10%	= .4
School-wide Learning Measure	2	x 5%	= .1
Sum of the Weighted Scores			2.675



If you have received this document from any source other than the RISE website, it may have been altered from its original version. For the official, and most up-to-date version, please visit www.RISEIndiana.org.

The weighted score determines the final rating.



Note: Borderline points always round up.

- In the Mrs. Smith example, the weighted score of 2.675 is mapped to this scale. The final rating is “Effective”.



If you have received this document from any source other than the RISE website, it may have been altered from its original version. For the official, and most up-to-date version, please visit www.RISEindiana.org.

This proposed reporting strategy is intended to inform teachers about the four steps that comprise the evaluation process. The reports that emerge will describe the process used to calculate domain-level ratings, including details about the relative weight of each domain. A final Professional Practice rating also will be provided.

In addition, the department plans to make explicit the ways in which the Teacher Effectiveness Rubric (TER) score is combined with other types of information to create a weighted score that will be used in assigning a final overall rating. This type of report provides transparency for teachers and administrators about how the final rating was calculated.

C. New York City⁸

The New York City (NYC) Department of Education has historically been forward-thinking in developing and implementing data reports for teachers that are comprehensive, explanatory, and easily understood. An example of one of their value-added data reports is presented below. Importantly, supporting documents not shared in this paper include language clarifying that student growth (the data being calculated and presented in this report) comprises only one factor in teacher evaluation efforts. Providing such an explanation ensures that teachers do not assume that this score is reflective of the full range of their abilities.

⁸ In early 2012, much attention was given to New York City’s public release of student growth data calculated with a value-added model. See Chapter 4 for additional information about this release.

In NYC, each teacher in the district receives a performance report annually. The following teacher-level report, which was showcased in a study of measures of teacher effectiveness (Rockoff, Staiger, Kane, and Taylor, 2010), is from the 2006-2007 school year and contains information about the teacher's value-added growth data. Note that developers include information about the comparison group, sample size, and upper and lower bounds for error estimate (see sections labeled Citywide Horizon and Peer Teacher Horizon). This report is missing information about how scores were calculated, however we assume that that type of information would have been available in supporting documents.

NYC Department of Education
Value-added Data for Teachers Initiative

Teacher: Swain, Winthrop **Years in Current Grade/Subject:** 3
Grade: 5.0th Grade **Experience Category:** 10+Yrs
School: PS 006 Lillie D. Blake **Classroom Quintile:** Fourth
Year: 2006-2007

Teacher Performance
The Difference-from-Predicted gain in the average student proficiency level for this teacher is the difference between the average actual gain of all the teacher's students and the average predicted gain for students with similar characteristics.

Teacher Compared to Citywide Teacher Performance Horizon - All schools, all teachers; same grade

	Sample Size	Actual Gain	Predicted Gain	Difference from Predicted (Teacher's Value Added)	Citywide Horizon Teacher value-added relative to range of results for all in same grade in the City
ELA - This year (lower / upper bound)	40	.22	.07	.15* (.02, .27)	76.4% (56%, 96.9%) -0.31 .15* 0.49
ELA - History: up to 3 years (lower / upper bound)	144	.11	.04	.07* (.00, .14)	67.4% (54.9%, 79.9%) -0.30 .07* 0.25
Math - This year (lower / upper bound)	43	.40	.17	.23* (.09, .37)	78.8% (62.1%, 95.5%) -0.45 .23* 0.41
Math - History: up to 3 years (lower / upper bound)	152	-.03	-.09	.06 (-.01, .13)	62.4% (52.6%, 72.2%) -0.40 .06 0.34

Teacher Compared to Peer Teacher Performance Horizon – Similar experience, similar classrooms; same grade

	Sample Size	Actual Gain	Predicted Gain	Difference from Predicted (Teacher's Value Added)	Peer Teacher Horizon Teacher value-added relative to range of results for all teachers in the same grade with similar experience and similar classrooms
ELA - This year (lower / upper bound)	40	.22	.14	.07 (-.04, .19)	69.8% (43.3%, 96.2%) -0.23 .07 0.21
ELA - History: up to 3 years (lower / upper bound)	144	.11	.09	.03 (-.04, .09)	60.1% (45.3%, 75%) -0.25 .03 0.21
Math - This year (lower / upper bound)	43	.40	.22	.18* (.05, .31)	75.4% (59.7%, 91.1%) -0.47 .18* 0.39
Math - History: up to 3 years (lower / upper bound)	152	-.03	-.04	.01 (-.06, .08)	54.4% (41.3%, 67.6%) -0.28 .01 0.25

Note: The lower and upper bound means that there is a very high probability (95%) that the teacher's actual contribution to student gains in proficiency falls within this interval. The (*) means that there is a very high probability that the contribution is positive (or negative). All comparisons are among teachers in the same grade.

In the associated report shown below, the teacher's performance is reported for different student subgroups (similar prior performance, English language learners, and students with disabilities). The teacher's percentile rank in comparison to other teachers at the same grade level is reported, as is the percentage of students in that teacher's classroom whose actual scores met or exceeded their predicted scores.

NYC Department of Education				
Value-added Data for Teachers Initiative				
Teacher: Swain, Winthrop				
Teacher Performance by Student Characteristics				
Teacher's value-added for sub-groups of students compared to teacher's value-added overall for history: up to 3 years				
Types of Student	Sample Size / (% of Sample)	Actual Gain	Predicted Gain	Difference from Predicted (Teacher's Value Added)
English Language Arts				
All Students	144 (100%)	0.11	0.04	0.07*
Citywide:				
Bottom Third	94 (62.8%)	0.27	0.16	0.10*
Middle Third	39 (29.3%)	-0.13	-0.14	0.01
Top Third	11 (7.9%)	-0.32	-0.37	0.04
School				
Bottom Third	51 (32.5%)	0.39	0.24	0.16*
ELL	-	-	-	-
Special Education	15 (10.1%)	0.19	0.02	0.17
Mathematics				
All Students	152 (100%)	-0.03	-0.09	0.06
Citywide:				
Bottom Third	106 (64.2%)	0.11	0.01	0.10*
Middle Third	37 (28.4%)	-0.33	-0.30	-0.03
Top Third	9 (7.4%)	-0.46	-0.45	-0.02
School				
Bottom Third	48 (25.2%)	0.24	0.14	0.11
ELL	10 (6.8%)	-0.14	0.01	-0.15
Special Education	15 (9.1%)	-0.01	-0.11	0.11
The (*) means that there is a very high probability that the contribution is positive (or negative).				
Teacher Percentile				
The percent of teachers in the comparison group whose value added falls below this teacher				
Comparison Teachers	English Language Arts		Mathematics	
	This Year	History: up to 3 years	This Year	History: up to 3 years
All teachers, all schools	88	77	90	69
Teachers with similar experience, similar classrooms	86	71	86	56
Note: All comparisons are among teachers in the same grade				
Basic Student Progress				
The percent of students in the teacher's classroom making at least the predicted gain				
	English Language Arts		Mathematics	
	This Year	History: up to 3 years	This Year	History: up to 3 years
This Teacher	70.0%	60.4%	72.1%	53.3%
All teachers, all schools	48.4%	47.3%	49.9%	47.3%
Teachers with similar experience, similar classrooms	56.7%	51.8%	48.5%	54.3%
Note: All comparisons are among teachers in the same grade				

A second illustrative example from the NYC Department of Education is provided below. The target audience for this score report is teachers. The report describes how value-added scores are estimated, lists the student- and classroom-level factors taken into consideration (“controlled for”) during the estimation process, and explains how percentile ranks were assigned.

I.S. 000
8th Grade
Years Teaching in NYC: More Than 3 Years

NYC TEACHER DATA REPORT
MATH Mark Jones

How does the value-added model work?

- Teacher Data Reports calculate teachers’ value-added by controlling for factors that can influence student achievement but that are outside of a teacher’s control, including, but not limited to, students’ prior test scores, class size, and the percentage of students in the school with disabilities and living in poverty.
- These factors are used in a statistical model to predict each student’s test score in math and ELA from one year to the next.
- The “value-added” is the difference between the predicted and actual scores.
- A teacher’s value-added is calculated by averaging this difference for all students in his/her class.

What factors does this report take into account?

Student Characteristics

- ✓ Prior year reading
- ✓ Prior year math
- ✓ Free or reduced price lunch
- ✓ Special education status
 - ✓ Related Services
 - ✓ Self-contained classroom
 - ✓ Team Teaching
 - ✓ Special Education Teacher Support
- ✓ English Language Learner status (current and former ELL)
- ✓ Number of suspensions and absences (prior-year)
- ✓ Student retained in grade
- ✓ Attended summer school
- ✓ New to school
- ✓ Race
- ✓ Gender

Classroom Characteristics

- ✓ Average prior year reading and math
- ✓ Percent free or reduced price lunch
- ✓ Percent special education status
 - ✓ Related Services
 - ✓ Self-contained classroom
 - ✓ Team Teaching
 - ✓ Special Education Teacher Support
- ✓ Percent English Language Learner status
- ✓ Average number of suspensions and absences (prior)
- ✓ Percent of students retained in grade
- ✓ Percent attended summer school
- ✓ Class size
- ✓ Percent by race
- ✓ Percent by gender

How is my percentile calculated?

Teacher Data Reports order teachers from lowest to highest value-added to determine a percentile rank.

With whom am I compared?

Throughout the report, teachers are only compared to other teachers who teach in the same grade, in the same subject, and are in the same experience group. Experience groups are determined by total teaching experience and include: 1 year, 2 years, 3 years, and 4+ years.

Presenting Findings from Measures of Teacher Effectiveness

21

A third example, highlighted below, is intended for educators and school administrators. Its stated purpose is to “improve instruction and student learning” (see caption in section titled *What is the Teacher Data Report?*). Under the section titled *What Data Goes Into the Calculations on This Report?*, report developers describe each of the measures included (state test scores, years of experience, and measureable factors considered outside the teacher’s control).

Note also that the graphic displays report results in three different ways: (1) in comparison to all NYC teachers; (2) in comparison to peers at same grade in same subject area; and (3) disaggregated by subgroup. In each case, a caption explains how displayed results were calculated or estimated. Additional tables are provided on subsequent pages of the report that drill more deeply into results for each. The importance of providing this level of transparency is discussed further in Sections D and E. It is also important to note that the report calls attention to those numbers that require extra caution in interpreting due to the large error band surrounding the numbers (as explained in Chapter 2, Section C). Sharing information about potential error rates alerts the audience to the reliability of these findings. In addition, results are reported in reference to all NYC teachers as well as grade-level peers. This report supports stakeholders in drawing valid inferences from results.

TEACHER DATA REPORT: ENGLISH LANGUAGE ARTS
SUMMARY SHEET

Teacher: Travis, Mary
 School: PS 31 - Lincoln Elementary
 Years with data: 2005-06, 2006-07, 2007-08

Grade Level: 5th
 Years Teaching in NYC: 4

What Is the Teacher Data Report?

- The Teacher Data Report is a new tool for teachers and school leaders to use to improve instruction and student learning.
- The information in this report is calculated by using a statistical model to isolate the effect of a teacher's instruction on student achievement from factors about students, classrooms and schools that are outside of a teacher's control. The model uses these factors to predict gains for each student.
- A teacher's result, also called by the statistical term "Value-Added," is the difference between the average "actual gain" and the average "predicted gain" for all students in the classroom.

What Data Goes into the Calculations on This Report?

Standardized NYS Test Scaled Scores in: Math and English Language Arts (ELA) from 2004-05 to 2007-08 (Baseline achievement data for 2004-05 includes some city tests)

Teacher Experience: The number of years the teacher taught in NYC and in this grade/subject

Student, Classroom and School Data: Measurable factors about students and classrooms outside of the teacher's control, including: prior year's standardized NYS test scaled scores, Special Education and ELL status, student demographics and class size.

This Page Summarizes Three Ways to Look at Teacher Data
 More Details on the Following Pages

1 My Results, Compared to All NYC Teachers Citywide:
 How do my results compare to other teachers in my grade and subject area throughout NYC?

	My Percentile	Range	My Percentile (0%-100%)					What Results Are Shown?
			0%	25%	50%	75%	100%	
2007-08	58%	39% 77%			▼		⇨ Shows my results from 2007-08, and the last three years (when available) ⇨ Comparison group: All teachers in my grade and subject area ⇨ NOT adjusted for teacher experience level	
Last 3 years	48%	37% 62%			▼			

2 My Results, Compared to Peer Teachers:
 How do my results compare to other teachers in my grade and subject area throughout NYC?

	My Percentile	Range**	My Percentile (0%-100%)					What Results Are Shown?
			0%	25%	50%	75%	100%	
2007-08	65%	46% 84%			▼		⇨ Shows my results from 2007-08, and the last three years (when available) ⇨ Comparison group: Peer teachers in my grade and subject area* ⇨ Adjusted for teacher experience level*	
Last 3 years	53%	40% 66%			▼			

3 My Results with Student Sub-groups:
 How do my results for student sub-groups compare with other teachers'?

0%-20% My Result Is Between these Percentiles	20%-80% My Result Is Between these Percentiles	80%-100% My Result Is Between these Percentiles	What Results Are Shown?
Citywide Top 3rd*	Citywide Middle 3rd School Top 3rd School Middle 3rd Male Students Female Students	Citywide Lowest 3rd School Lowest 3rd Special Education	

* If an asterisk appears, the range is large. Interpret with caution.

TEACHER DATA REPORT: ENGLISH LANGUAGE ARTS

3 STUDENT SUB-GROUPS

Teacher: Travis, Mary

School: PS 31 - Lincoln Elementary

Years with Data: 2005-06, 2006-07, 2007-08

Grade Level: 5th

Years Teaching in NYC: 4

How do my results for student sub-groups compare with other teachers'?

- ⇒ Uses three years of data (when available)
- ⇒ Comparison group: Peer teachers in my grade and subject area
- ⇒ Adjusted for teacher experience levels, overall and in grade

Average										My Percentile				Performance with sub-groups	
Number of Students	Prior Proficiency Rating	Actual Gain	Predicted Gain	Value-Added	Percentile (0-100%)	0%	25%	50%	75%	100%					
Prior Student Achievement Level															
Students in the Citywide Top 3rd	13	3.4	-0.41	-0.17	-0.24	14%	▼				Low*				
Range						0%-55%									
Students in the Citywide Middle 3rd	30	2.3	0.02	0.05	-0.03	46%	▼				Medium				
Range						25%-71%									
Students in the Citywide Lowest 3rd	30	1.6	0.48	0.29	0.19	81%	▼				High				
Range						63%-100%									
Prior Student Achievement Level															
Students in the School's Top 3rd	25	3.1	-0.21	-0.09	-0.12	35%	▼				Medium				
Range						18%-52%									
Students in the School's Middle 3rd	28	2.3	0.02	0.05	-0.03	47%	▼				Medium				
Range						25%-69%									
Students in the School's Lowest 3rd	20	1.6	0.48	0.29	0.19	81%	▼				High				
Range						58%-100%									
Gender															
Male Students	35	2.4	0.09	0.07	0.02	54%	▼				Medium				
Range						38%-70%									
Female Students	38	2.5	0.07	0.07	0.00	51%	▼				Medium				
Range						35-67%									
Other Sub-groups															
ELL Students	-	-	-	-	-	-									
Range						-									
Special Education	10	1.4	0.02	-0.18	0.20	83%	▼				High				
Range						51%-100%									

Range: Statistically, your result most likely lies near, but may not be exactly equal to, the highlighted percentile result. Therefore a range is provided.

High: My result is between the 80th and 100th percentiles
* If an asterisk appears, the range is large. Interpret with caution.

Medium: My result is between the 20th and 80th percentiles

Low: My result is between the 0 and 20th percentiles
* If an asterisk appears, the range is large. Interpret with caution.

TEACHER DATA REPORT: ENGLISH LANGUAGE ARTS

- 1 COMPARISONS TO ALL TEACHERS CITYWIDE
- 2 COMPARISONS TO PEER TEACHERS

Teacher: Travis, Mary
 School: PS 31 - Lincoln Elementary
 Years with Data: 2005-06, 2006-07, 2007-08

Grade Level: 5th
 Years Teaching in NYC: 4

1 My Results, Compared to All NYC Teachers Citywide:

How do my results compare to all teachers in my grade and subject area throughout NYC?

- ⇒ Shows my results from 2007-08, and the last three years (when available)
- ⇒ Comparison group: All teachers in my grade and subject area
- ⇒ NOT adjusted for teacher experience level

	Number of Students	Prior Proficiency Rating	Average			Percentile (0-100%)	My Percentile				
			Actual Gain	Predicted Gain	Value-Added		0%	25%	50%	75%	100%
This year: 2007-08 <i>Range</i>	24	2.1	0.19	0.12	0.07	58% 39-77%					
2006-07 <i>Range</i>	24	2.4	0.08	0.07	-0.03	46% 26-66%					
2005-06 <i>Range</i>	25	2.5	0.03	0.09	-0.06	40% 19-59%					
Last 3 years average <i>Range</i>	73	2.4	0.10	0.11	-0.01	49% 37-62%					

2 My Results, Compared to Peer Teachers:

How do my results compare to other teachers in my grade and subject area throughout NYC, whose classrooms have similar predicted gains, adjusted for teacher experience levels?

- ⇒ Shows my results from 2007-08, and the last three years (when available)
- ⇒ Comparison group: Peer teachers in my grade and subject area
- ⇒ Adjusted for teacher experience level

	Number of Students	Prior Proficiency Rating	Average			Percentile (0-100%)	My Percentile				
			Actual Gain	Predicted Gain	Value-Added		0%	25%	50%	75%	100%
This year: 2007-08 <i>Range</i>	24	2.1	0.19	0.08	0.11	65% 46-84%					
2006-07 <i>Range</i>	24	2.4	0.08	0.08	0.00	50% 30-70%					
2005-06 <i>Range</i>	25	2.5	0.03	0.06	-0.03	43% 22-64%					
Last 3 years average <i>Range</i>	73	2.4	0.10	0.07	0.03	53% 40-66%					

***Range: Statistically, your result most likely lies near, but may not be exactly equal to, the highlighted percentile result. Therefore a range is provided.*

In addition to providing comprehensive score reports for student growth data, the NYC Department of Education has also historically included a “Frequently Asked Questions” document to explain score reports and address those questions likely to be asked by key stakeholders. Below is an example from 2011—these FAQ’s have since been updated on the department’s website. Note the visually appealing, easy-to-digest sample graphs depicting the distribution of teacher effectiveness.

5/9/2011
Teacher Data Reports, NYCDOE

Teacher Data Reports Frequently Asked Questions Regarding Relationship Between Value-Added and Student Achievement

We have received questions from principals and teachers wanting to better understand how teachers’ value-added scores are related to their students’ performance on the state tests. We have provided this guide to help principals and teachers understand this relationship. If you have additional questions, please do not hesitate to contact the TDI Support Helpdesk at tdisupport@schools.nyc.gov.

- 1. How do I understand the relationship between student performance in z-scores, scale scores, and raw scores (questions correct) on the state tests?**
- 2. How do I understand the relationship between a teacher’s percentile score and his/her value-added z-score?**
- 3. Is there a relationship between value-added percentile scores and the prior achievement of teachers’ students?**
- 4. Are value-added scores more sensitive to changes in student performance for teachers of particularly low- or high-performing students than other teachers?**
- 5. How do different factors impact a teacher’s predicted score? Why is a teacher’s predicted score different from other teachers with students of similar starting achievement levels?**
- 6. What should I do if my views of teachers’ performance differ from their reported value-added results?**

1. How do I understand the relationship between student performance in z-scores, scale scores, and raw scores (questions correct) on the state tests?

[Lookup Table #1](#) shows the relationship between raw scores (questions correct) on the state tests, scale scores, and z-scores.

Teachers' predicted score, actual score, and value-added score are reported in standard deviations (also known as standardized scores or z-scores). Every year, the State Education Department converts students' raw scores on the state tests into scale scores. Students' scale scores on the state tests are converted into z-scores to perform the value-added calculations. More information about z-scores is available [here](#).

Teachers can download the Report Data file from the report website to see their predicted score, actual score, and value-added scores for each section of their report.

In Lookup Table #1, there is one table for each grade/ subject (e.g., 4th grade math). As this table shows, the conversion of raw scores to scale scores and z-scores varies by year – this is because 1) z-scores are based upon the distribution of test results for all students in the city for that year and 2) the state tests have different conversion of raw scores to scale scores for each test year (see the [NY State Education Department's website](#) for more information about the state tests and raw score to scale score conversions).

2. How do I understand the relationship between a teacher's percentile score and his/her value-added z-score?

[Lookup Table #2](#) shows the relationship between percentile scores and average value-added z-scores for teachers in different experience categories. There is one table for each grade/ subject (e.g., 4th grade math teachers).

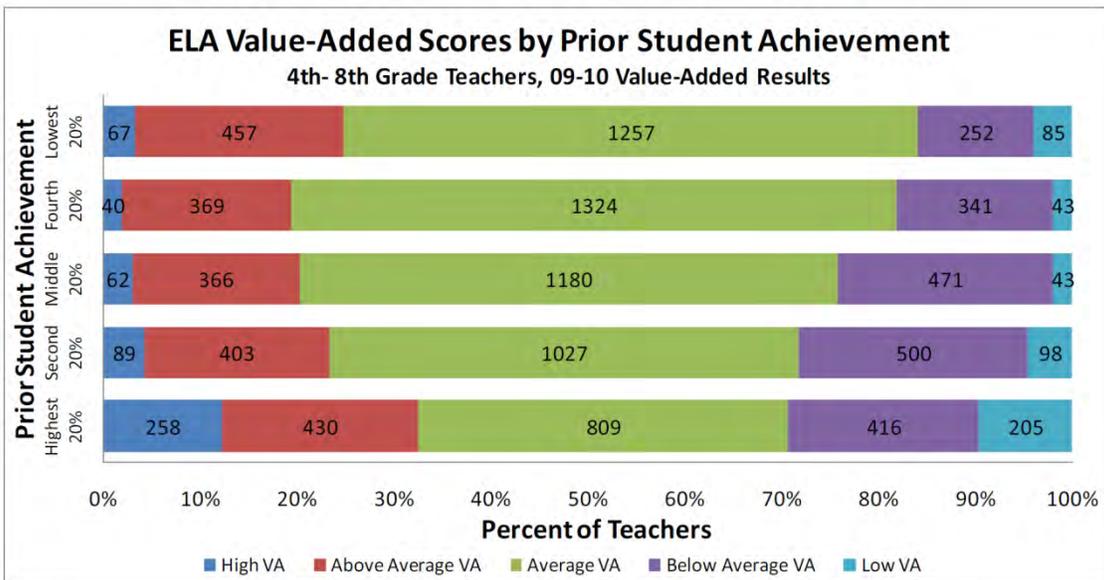
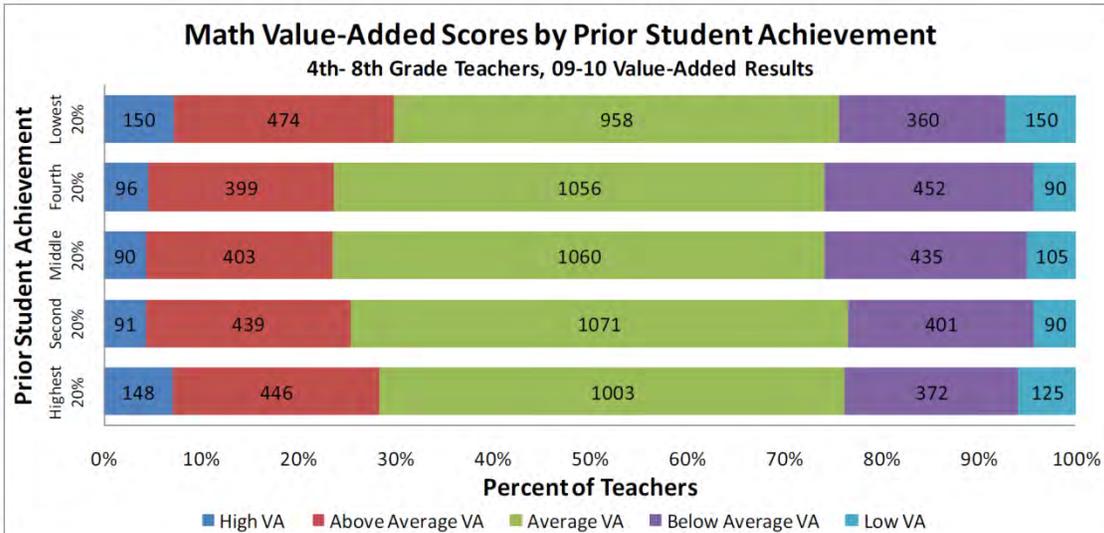
For each section of a teacher's report, the percentile shows the percentage of teachers in the same subject, grade, and experience category whose value-added score was lower than the teacher's. The experience categories are: 1st year teachers, 2nd year teachers, 3rd year teachers, teachers with more than 3 years experience. For example, if Teacher A, a 1st year 7th grade math teacher, has a percentile score of 87 that means that Teacher A's value-added score is higher than 87% of 1st-year 7th grade math teachers.

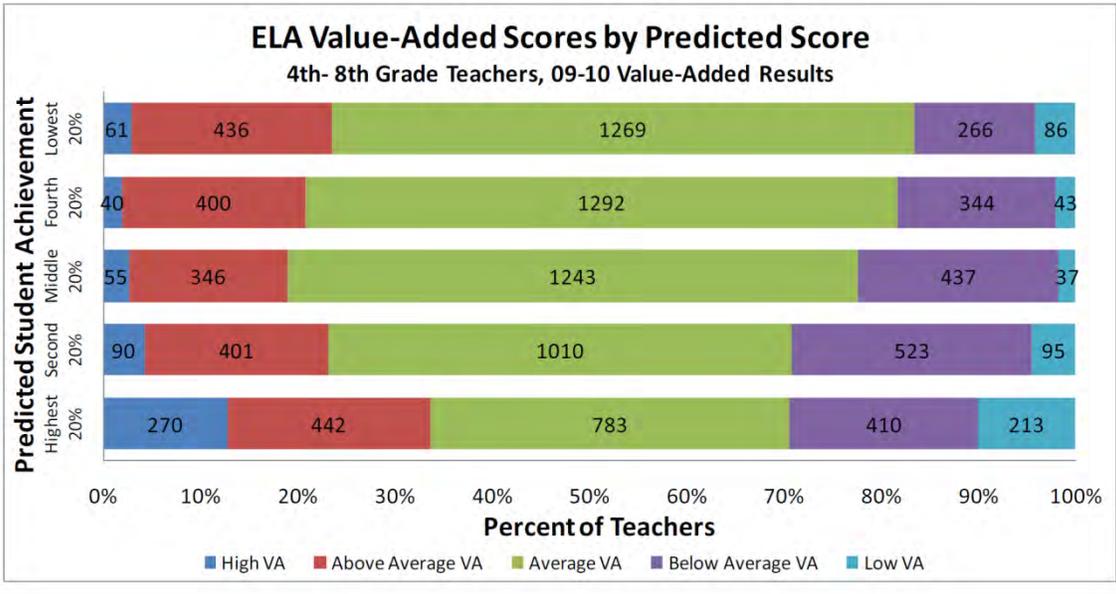
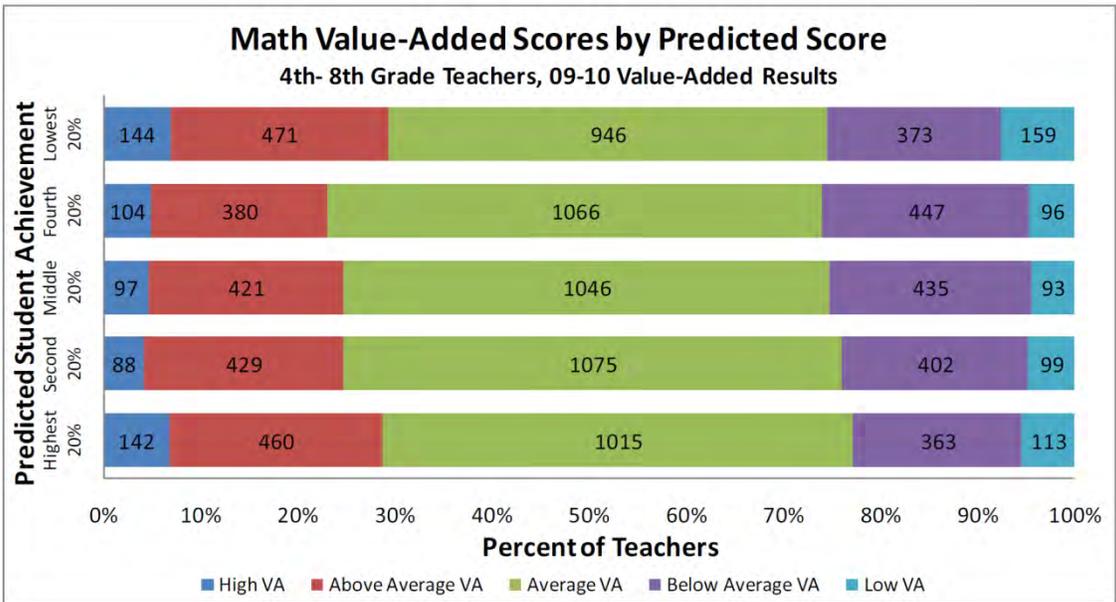
As Lookup Table #2 shows, the same value-added z-score tends to result in higher percentile scores for first- and second-year teachers than third-year teachers and teachers with more than 3 years of experience. Teachers and principals can use this table, alongside Lookup Table #1, to understand how changes in student performance could have affected a teacher's value-added z-score and percentile score.

3. Is there a relationship between value-added percentile scores and the prior achievement of teachers' students?

There is no systematic relationship between value-added percentile scores and the prior achievement of teachers' students. Teachers of high-performing students are as likely to have high value-added scores as low value-added scores. Similarly, teachers of low-performing students are as likely to have high value-added scores as low value-added scores.

The charts below show this pattern of results in two ways. The first two charts show the percent of teachers with different value-added scores by the prior achievement level of their students. The second two charts show the percent of teachers with different value-added scores by students' predicted scores.





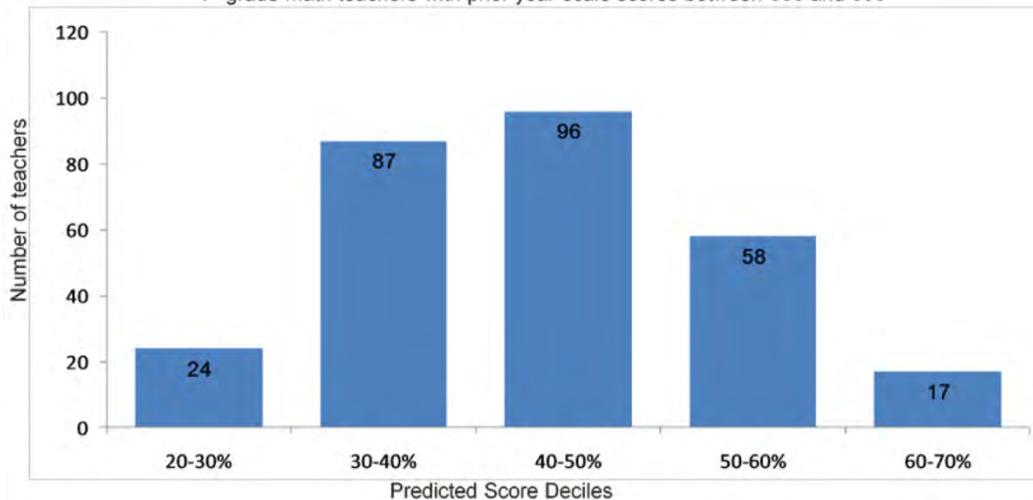
5. How do different factors impact a teacher’s predicted score? Why is a teacher’s predicted score different from other teachers with students of similar starting achievement levels?

The two charts below illustrate the impact of different variables on predicted scores. The first chart shows the range of predicted scores for 4th grade math teachers with the same entering student achievement level – this chart shows that teachers with the same entering student achievement levels can have a range of predicted scores. The second chart shows the difference in student characteristics of 4th grade math teachers with similar entering achievement levels – this chart shows that differences in student characteristics are the drivers of differences in predicted score for teachers with similar entering student achievement levels. (4th grade math teachers’ results are shown for illustrative purposes –the same pattern of results is found for other grades and subjects.)

Teacher Data Reports calculate teachers’ value-added by controlling for factors that can influence student achievement but that are outside of a teacher’s control, including students’ prior test scores, class size, and the percentage of students in a class with disabilities and living in poverty. These factors are used in a statistical model to predict each student’s test scores in math and ELA. The entering level of achievement is one of the biggest predictors of student achievement. However, even when classes have the same entering achievement level, other student and classroom characteristics can impact student achievement and therefore change a teacher’s predicted score.

Range in Predicted Scores for Teachers with Similar Prior Student Achievement

4th grade math teachers with prior year scale scores between 686 and 690



Note: This chart only shows predicted score deciles with more than 5 teachers

Differences in Student Characteristics for Teachers With Similar Prior Student Achievement

4th grade math teachers with prior year scale scores between 686 and 690

Characteristic	Classroom Average of Students Characteristics	
	Lowest 5% predicted score	Top 5% predicted score
Average prior year scale score in ELA	654	678
# of absences in days (prior year)	15	8
% students receiving free lunch	91%	28%
% students who are Special Education receiving related services or support services	17%	10%

D. North Carolina

The North Carolina Department of Public Instruction provides open access to all district-level reports of teacher effectiveness through their public website. These reports include aggregate data only; no teacher-level results are accessible to the public. Each report includes an introduction (see below) that clearly explains the purpose of the report and target audience. It also offers information about the standards on which teachers were evaluated, as well as the expected level of growth for the average teacher. In addition, report users are cautioned that only newer teachers are evaluated and that these teachers are more likely than their experienced peers to receive low ratings. Finally, performance levels are listed and explained. Providing this information supports report users in drawing valid inferences from evaluation results.

The following screen shot is from an online report and represents one district, the Alamance-Burlington Schools. Users who are accessing the state database are presented with a drop-down list that allows them to explore any district in North Carolina.

READ THIS FIRST: The following is critical to a thorough understanding of the information presented.

Information on Educator Effectiveness Data:

North Carolina is reporting on the effectiveness of teachers and administrators across the State. The reports show ratings of the performance of all teachers who were evaluated in the 2010-11 school year. This group includes mostly new teachers and some veteran teachers. Data for all principals and assistant principals also are reported.

The following pages provide data on educator effectiveness in this school district. In North Carolina, educator effectiveness is gauged through the use of the North Carolina Educator Evaluation System and other informal means. This system engages teachers, principals, and their evaluators in dynamic discussion that recognizes educators' individual strengths and focuses on how they can improve their craft.

Important Information:

Please keep the following in mind when reviewing the attached educator effectiveness data.

1. North Carolina's Educator Evaluation System is a **growth instrument**. It identifies the knowledge, skills, and dispositions expected of teachers, and measures the level at which they meet the standard as they move from ratings of "developing" to "distinguished."
2. As lifelong learners, teachers and school leaders are constantly learning and growing. Since they always have room to improve, it is uncommon to see a school in which the majority of teachers are distinguished. It is **expected that teachers in a school would be distributed across the categories**.
3. During the 2010-11 school year, the State only required school districts to complete **evaluations for teachers in their first three years in the classroom and tenured teachers renewing their licenses**. As a result, new teachers are heavily over-represented in the data, and the data do not fully represent the school or school district. New teachers are more likely to be rated lower on the evaluation standards as they are still learning and developing new skills and knowledge.

Table 6b: District Report of Aggregate Principal/AP Evaluation Ratings for All Standards													
District: Alamance County Schools													
Creation Date: 19 September 2011													
Filter Period: 1 August 2010 to 31 July 2011													
Standard	Not Demonstrated (1)		Developing (2)		Proficient (3)		Accomplished (4)		Distinguished (5)				
	#	%	#	%	#	%	#	%	#	%			
Standard1	1.0	1.3	9.0	11.4	33.0	41.8	31.0	39.2	5.0	6.3			
Standard2	2.0	2.5	8.0	10.1	18.0	22.8	38.0	48.1	13.0	16.5			
Standard3	0.0	0.0	8.0	10.1	33.0	41.8	33.0	41.8	5.0	6.3			
Standard4	0.0	0.0	6.0	7.6	27.0	34.2	39.0	49.4	7.0	8.9			
Standard5	1.0	1.3	5.0	6.3	25.0	31.6	39.0	49.4	9.0	11.4			
Standard6	0.0	0.0	9.0	11.4	26.0	32.9	40.0	50.6	4.0	5.1			
Standard7	0.0	0.0	2.0	2.5	29.0	36.7	40.0	50.6	8.0	10.1			

As North Carolina endeavors to complete the work proposed in the state’s winning Race to the Top grant application and implement a revised evaluation system, a different type of summary report is emerging. As shown below, this report provides individual teachers with information about the standards on which they were judged and their level of performance (Not Demonstrated, Developing, Proficient, Accomplished, or Distinguished) on each. This at-a-glance format is visually appealing, convenient, and useful to teachers and policymakers, but the measures used to calculate ratings are not immediately evident.

Draft January 2012: Assume SBE Adoption of Option One in TCP-C-006

North Carolina Educator Evaluation System Evaluation Summary Sheet

Name: Martha Washington **School:** Independence Elementary School
LEA: Freedom County Schools **Licensure:** Career-Status
Overall Status: Effective

Standard One: Teachers demonstrate leadership.				
Not Demonstrated	Developing	Proficient	Accomplished	Distinguished
Standard Two: Teachers establish a respectful environment.				
Not Demonstrated	Developing	Proficient	Accomplished	Distinguished
Standard Three: Teachers know the content they teach.				
Not Demonstrated	Developing	Proficient	Accomplished	Distinguished
Standard Four: Teachers facilitate learning for their students.				
Not Demonstrated	Developing	Proficient	Accomplished	Distinguished
Standard Five: Teachers reflect on their practice.				
Not Demonstrated	Developing	Proficient	Accomplished	Distinguished
Standard Six: Teachers contribute to the academic success of students.				
<i>*Only three-year rolling average is used to determine overall status*</i>				
Year One (2009 - 2010)		Year Two (2010 - 2011)		Year Three (2011 - 2012)
Individual Student Growth: .8 School-wide Student Growth: .1 Year One Growth: .73		Individual Student Growth: 1.2 School-wide Student Growth: .5 Year Two Growth: 1.13		Individual Student Growth: .7 School-wide Student Growth: .5 Year Three Growth: .68
Does not meet expected growth	Meets expected growth	Exceeds expected growth	Does not meet expected growth	Meets expected growth
				Exceeds expected growth
				Does not meet expected growth
				Meets expected growth
				Exceeds expected growth
				Three-Year Rolling Average*
				.85
				Does not meet expected growth
				Meets expected growth
				Exceeds expected growth
Overall Status:	Needs improvement		Effective	
			Highly Effective	

E. Pennsylvania

The Pennsylvania Department of Education utilizes a value-added system (Pennsylvania Value-Added Assessment Model or PVAAS) to measure a teacher’s unique contribution to student growth. The PVAAS team has developed a “Guide to Public Reporting” to support interpretation of its district-level reports of teacher effectiveness. This report uses a color-coded system to guide readers in understanding the implications of a district’s average growth rating. To further support understanding, the report uses a medical analogy, indicating the severity of need for intervention at each rating level. A sample report is provided below.

District Value-Added Summary Report for Grades 4 through 8 for Reading and Math

The District Value Added Summary Report provides educators with an overall look at the progress, or growth, of groups of students by subject within each district and charter school across the entire Commonwealth. This report indicates if the district/charter school met or exceeded the standard for PA Academic Growth in grades 4-8 Reading and/or Math.

District Name		Growth Measure over Grades (Math) Relative to	
		Growth Standard	State
Sample District	2011	0.8	-0.5
	3-Yr.Avg	2.6	1.3
Sample District	2011	1.5	0.1
	3-Yr.Avg	2.4	1.1
Sample District	2011	-5.6	-6.9
	3-Yr.Avg	-2.8	-4.1
Sample District	2011	3.1	1.8
	3-Yr.Avg	1.9	0.6
Sample District	2011	1.5	0.1
	3-Yr.Avg	2.8	1.5
Dark Blue		Significant evidence that the district exceeded the standard for PA Academic Growth	
Light Blue		Moderate evidence that the district exceeded the standard for PA Academic Growth	
Green		Evidence that the district met the standard for PA Academic Growth	
Yellow		Moderate evidence that the district did not meet the standard for PA Academic Growth	
Red		Significant evidence that the district did not meet the standard for PA Academic Growth	
--		The district does not have data for this test and subject in the most recent year.	

The two levels below the Growth Standard (Yellow and Red) can be understood using a medical analogy:

- A Yellow, or Y, is comparable to taking your temperature and recording a fever of approximately 100.5°F. It is unlikely that you would go to the emergency room with that temperature but this may warrant a call to the doctor’s office. We say that a Yellow indicates moderate evidence that the district did not meet the standard for Pa Academic Growth.
- A Red, or R, indicator requires immediate attention, comparable to a temperature of perhaps 104.0°F. This temperature provides significant evidence that the patient needs immediate attention, just as a Red indicator provides significant evidence that the district did not meet the standard for PA Academic Growth.

In this same way, the Light Blue, LB, and Dark Blue, DB, indicators signify moderate and significant evidence of exceeding the standard for PA Academic Growth.

What do we need to know about this report?

- The numbers are reported in NCE – Normal Curve Equivalent (NCE) units. It is necessary to convert PSSA scaled scores to a common scale (NCEs) so a growth measure can be yielded. The use of NCEs allows PSSA scores in any school year and grade level to be compared across years.
- The value displayed next to each district for the most recently tested year, is the *Growth Measure* for the specified grades and subject. The *Growth Measure*, displayed across grades, indicates how much movement a group of students has made on the NCE scale (that ranges from approximately 1 to 100) as compared to the previous year. A *Growth Measure* of 0 indicates that this group of students has maintained its position from the previous year in the statewide distribution of PSSA scores (baseline of 2006); this group has met the Standard for PA Academic Growth.
- If a Green (G), Light Blue (LB), or Dark blue (DB) rating is listed for some schools but not others, it is beneficial for all schools to dig deeper and begin conversations involving questions such as, “How has this result been achieved?” and “How can we learn from this school to impact growth in other schools in our district?”

F. Tennessee

Another example of comprehensive score reporting for value-added data is provided below from the Tennessee Department of Education’s *Value-Added Assessment System (TVAAS)*. This is a screen shot taken from their TVAAS database, which is password protected. These reports are particularly unique and useful because they are interactive; teachers can dig deeper into embedded links to get detailed information about a particular score or judgment. Teachers see information that includes an aggregate score, as well as subscores, each accompanied by a graphic display of results. The report includes a summary of findings, average estimated progress with associated standard error, and an average effectiveness level rating. The initial login screen assures teachers that, per Tennessee law, the value-added estimates contained within the website are only available to the “specific teacher, the teacher’s appropriate administrators...and school board members.”

The screenshot shows a web interface titled "Teacher Report". It contains a login form with two input fields: "New Teacher Licensure Number" and "TVAAS Password", both masked with dots. Below the form is a legal disclaimer: "As required by TCA 49-1-606(b), 'The estimates of specific teacher effects on the educational progress of students will not be a public record, and will be available only to the specific teacher, the teacher's appropriate administrators as designated by the local board of education, and school board members.'" There is a checkbox labeled "I accept the above statement" which is checked. At the bottom of the form are "Cancel" and "Submit" buttons. Below the form are three links: "Teacher Report Module", "Help", and "Contact Us".



[Click here for more information](#)

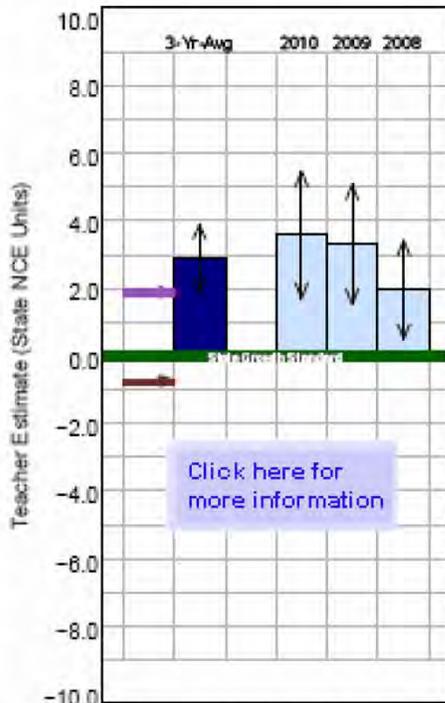
Sample District
Sample School

Sample Teacher
TCAP Reading/Language, Grade 4

3-Year-Average Estimated Progress = 2.9
(Standard Error = 1.0)

3-Year-Average Effectiveness Level = FIVE

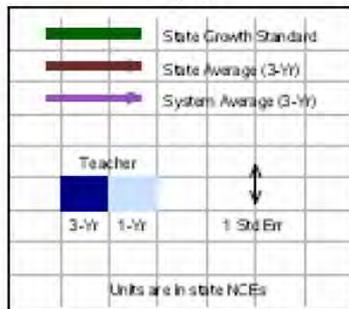
TVAAS Teacher Graph



[Click here for more information](#)

Teacher progress estimates and standard errors are presented in the chart above. This allows each teacher to compare their students' progress with the state growth standard, state average, and system average.

Legend



Teacher estimates are from SAS® EVAAS® multivariate, longitudinal analyses using all available data for each student (up to 5 years). Copyright © 2010 SAS Institute Inc., Cary, NC, USA. All Rights Reserved.

TVAAS Teacher Table

State Growth Standard	0.0
State Average (3-Yr)	-0.8
System Average (3-Yr)	1.9

[Click here for more information](#)

Teacher Progress Estimates and Standard Errors

	Estimate	StdErr	Index	
- Teacher 3-Yr-Avg	2.9	1.0	2.86	Level 5
- Teacher 2010	3.6	1.9	1.92	Level 4
- Teacher 2009	3.3	1.8	1.87	Level 4
- Teacher 2008	2.0	1.5	1.35	Level 4

(All metrics except Index expressed in state NCEs reflecting a base year of 2009)

What is a teacher value-added estimate?

It is an indicator of how much the teacher influences his or her students' academic progress. The associated standard error is a measure of the uncertainty around the teacher's estimate.

How is this used to determine an effectiveness level?

The ratio of the teacher's estimate to its standard error (shown in the column labeled 'Index') is used to determine the teacher's effectiveness level based on the rules shown below. The effectiveness level at the top of this report is based on a multi-year average, when available.

What is my teacher effectiveness level?

The teacher's effectiveness level is 'FIVE,' most effective, because the teacher's index is 2 or greater. The educational outcome for the average student taught by this Level 5 teacher is that the student made decidedly more progress than the state growth standard of 2009.

Rules for Effectiveness Level Determination

- Level Five, Most Effective: Teachers whose students are making substantially more progress than the state growth standard (the teacher's index is 2 or greater).
- Level Four, Above Average Effectiveness: Teachers whose students are making more progress than the state growth standard (the teacher's index is equal to or greater than 1 but less than 2).
- Level Three, Average Effectiveness: Teachers whose students are making the same amount of progress as the state growth standard (the teacher's index is equal to or greater than -1 but less than 1).
- Level Two, Approaching Average Effectiveness: Teachers whose students are making less progress than the state growth standard (the teacher's index is equal to or greater than -2 but less than -1).
- Level One, Least Effective: Teachers whose students are making substantially less progress than the state growth standard (the teacher's index is less than -2).

[Click here for more information](#)

3-Year-Average State Distribution of Teachers
(TCAP Reading/Language, Grade 4)

Level Five, Most Effective	137
Level Four, Above Average Effectiveness	183
Level Three, Average Effectiveness	677
Level Two, Approaching Average Effectiveness	325
Level One, Least Effective	266

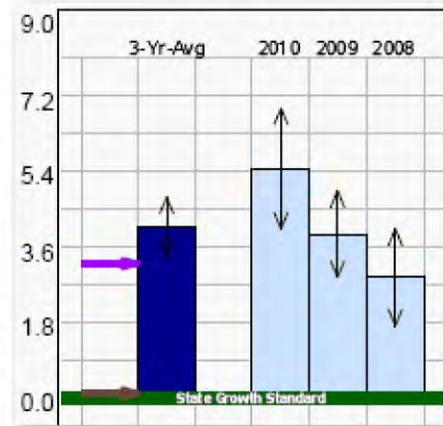
Below is a screen shot of one of the domains included in the above report—the teacher’s three-year average as graphed against the state’s growth standard. Also provided are the state and district’s average effect for that grade and subject, as well as the confidence interval surrounding each data point.

The TVAAS Teacher Graph for TCAP [\(back to image\)](#)

The graph is a visual representation of the progress your students made, compared to the State Growth Standard. The State Growth Standard is represented by the green zero (0.0) line on the chart. Results are presented in State NCEs, basis 2009.

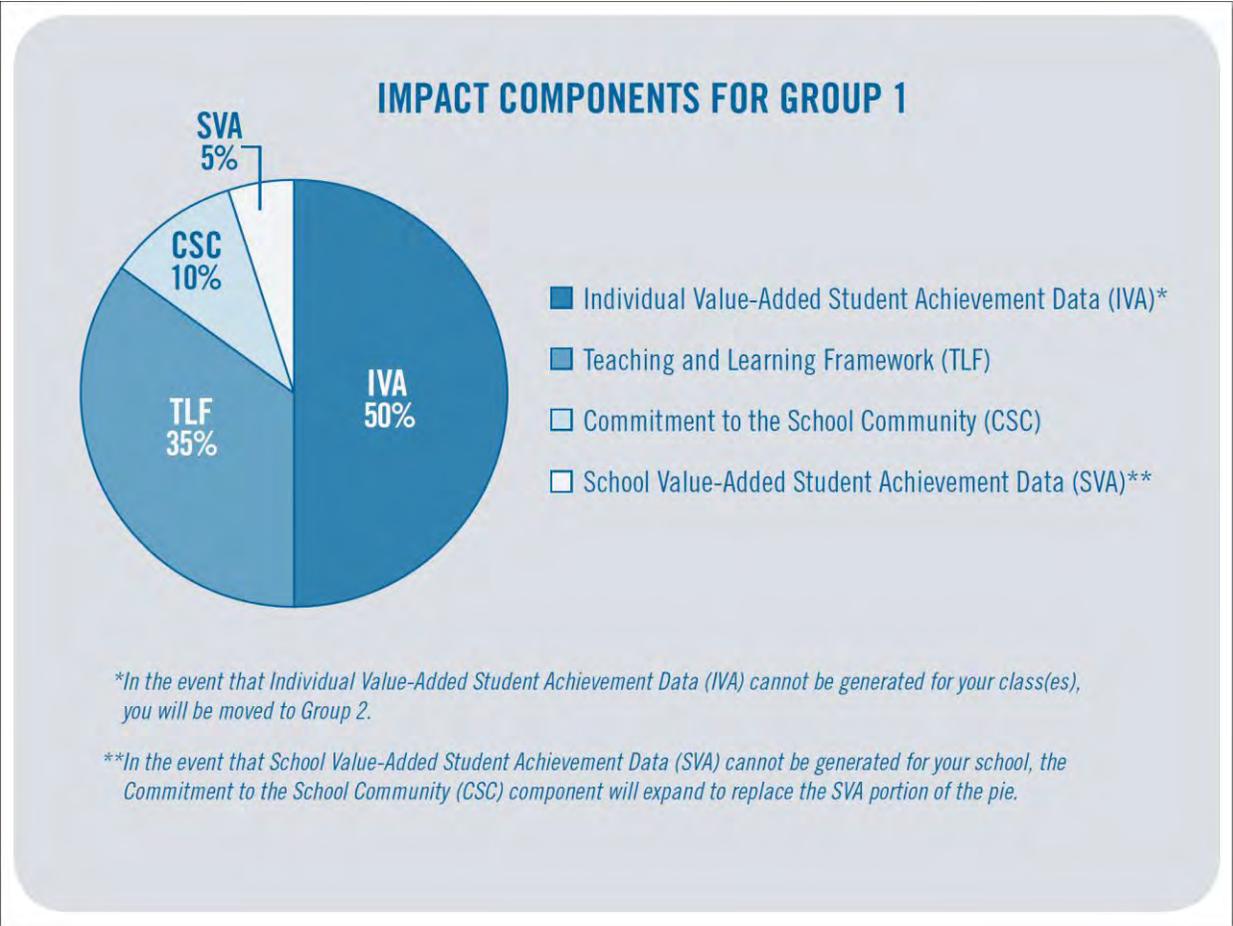
The dark blue bar is your multi-year average estimate, when available. It will be the average of your estimates for the most recent three years, when sufficient data exist. The light blue bars are your single year estimates for the most recent three years, when available. The double-pointed, vertical arrows indicate one standard error above and below the estimate, providing a confidence band around it.

The red arrow identifies the 3-Yr. State Average Effect for this grade and subject. The purple arrow indicates the 3-Yr. Average System Effect for this grade and subject for the system in which you taught in the most recent testing year.



G. District of Columbia

In their instructional materials for the DC IMPACT teacher evaluation system, the District of Columbia (DC) Public Schools includes this graphic that describes each component of a teacher’s evaluation rating. This particular report is for teachers in DC’s “Group 1,” teachers in tested grades and subjects.



Similar graphics have been developed to describe the performance of teachers in non-tested grades and subjects and of support staff; reports for each group are customized to explain the components on which educators will be rated and the relative weight of each component in the composite score. Score reports for DC's *Guidebook for Teacher Evaluation* also provide the cut-scores for each performance level (see example shown below).



Chapter 4: Legal Considerations

In this chapter, additional considerations are discussed to guide states and other jurisdictions that are developing plans for reporting on measures of teacher effectiveness or are already producing such reports. These considerations include (1) laws or regulations that affect what information can and cannot be shared with different audiences, including a state-of-the-states profile, and (2) case studies from the **Los Angeles Unified School District (LAUSD)** and **New York City Public Schools**, where effectiveness ratings for individual teachers were made publicly available. This chapter specifically addresses questions such as the following:

- How might states’ reporting strategies be affected by laws, court-based rulings, and/or policies regarding disclosure of information?
- Will a formal state-level ruling about the public’s right to information⁹ require disclosure of certain types of information?
- Do states have laws or policies governing teachers’ rights to privacy and/or confidentiality of evaluation findings?

Types of Laws and Policies That Affect Reporting

Some states have legal requirements associated with teacher evaluation and/or teacher effectiveness systems that dictate the types and amount of information that must or must not be shared with audiences other than the teacher and his or her school and district administrators. In general, a state’s approach can be classified as *prescriptive*, *restrictive*, or *flexible* in relation to the extent to which release of evaluation information or results from measures of effectiveness is left up to the districts (Epstein & Miller, 2011, p. 7).

State Approach to Managing District Reporting Practices	Description	Practical Example
Prescriptive	Requires all districts to follow specific common, standardized procedures in relation to reporting of information from teacher evaluations or measures of teacher effectiveness.	Districts must publicly report findings from value-added analyses using state-developed templates.
Restrictive	Prohibits all districts from certain actions.	Districts may not publicly report information associated with value-added analyses at the teacher level.
Flexible	Provides for district-level decision making within certain boundaries.	Each district may choose whether to publicly release a value-added score or a performance level (highly effective, effective, or ineffective) for each teacher or a school-level aggregate score.

States with a flexible approach will have less control over the districts’ reporting practices. These states may want to conduct a statewide survey so that they are clear on what may or may not be emerging from each district in terms of reporting the findings from measures of teacher effectiveness.

⁹ This category includes Freedom of Information Acts and “Sunshine Laws.”

Similarly, the types and amount of information that may be reported by individual districts is prescribed by that district's policy regarding the following question: *Are evaluation data part of a teacher's confidential file?* States and other jurisdictions can face legal challenges if they have not carefully examined existing policies and legislation that dictate what information *cannot* be released publicly and is not subject to Freedom of Information Act requests, what information may be part of the public record and may be requested or voluntarily released, and which data are confidential. For example, some districts share specific information (value-added score) about a teacher's level of effectiveness with certain parents via a written letter, while keeping other aspects of that teacher's evaluation confidential.

Appendix D of this paper provides a description, by state, of the differences in the laws or other regulations that have an impact on reporting findings from measures of teacher effectiveness. These findings suggest that access to teacher evaluation results are allowed under open records laws in at least 18 states and the District of Columbia. Others states allow disclosure under certain conditions, (e.g., only in cases of compelling public interest where the evaluation led to suspension or termination [Arkansas] or to inform parents that their students are being taught by two or more ineffective teachers [Florida, Indiana]). Nineteen states do not allow access or require teacher or third-party approval to release information.

Case Studies

Media outlets have different priorities from states and districts; media outlets tend to see increasing public knowledge as a mission and ensuring transparency as critical. Journalists frequently state that parents and other stakeholders have a right to know about the instructional effectiveness of teachers in their local schools, and are committed to dissemination of information to a broad audience (Epstein & Miller, 2011). This section describes two high-profile cases that highlight the dilemma facing all jurisdictions: Who should have access to information about an individual teacher's effectiveness? Balancing the needs and rights of diverse stakeholders and setting appropriate boundaries may prove challenging for those responsible for deciding what and how much information about teachers' effectiveness can or should be reported to different audiences.

Los Angeles Unified School District (LAUSD). In August of 2010, the *Los Angeles Times* released a searchable, online database of LAUSD teachers' value-added ratings. The newspaper was able to secure the data through the California Public Records Act, which legislates that all records of public agencies must be available to the public unless there is confidential personal material, and/or unless the consequences of releasing the information outweigh the benefits. These ratings were based on California Standards Tests in mathematics and English for students in grades 3-5 from the 2004-2005 through the 2009-10 academic years. The *Los Angeles Times* hired their own expert, Richard Buddin (2011), to conduct the analyses. Approximately 11,500 teachers and 470 elementary schools were included in the database. Teachers were included even if they did not teach for the full span of time covered in the analysis.

Los Angeles Teacher Ratings

About **11,500** Los Angeles Unified elementary school teachers and **470** elementary schools are included in The Times' updated database of "value-added" ratings.

Most third-, fourth- and fifth-grade instructors who taught at any point during the 2004-05 through 2009-10 academic years were given ratings in the Times analysis. Most district elementary schools are included. Test scores for most charter schools were not available.

A teacher's value-added rating is based on his or her students' progress on the California Standards Tests for English and math. The difference between a student's expected growth and actual performance is the "value" a teacher added or subtracted during the year. A school's value-added rating is based on the performance of all students tested there. Small differences in ratings are not statistically significant, particularly for those rated near the average.

Although value-added measures do not capture everything about a teacher or school's performance, The Times decided to make the ratings available because they bear on the work of public employees who provide an important service, and in the belief that parents and the public have a right to the information.

Find a teacher...

Or, find a school

Top value-added schools

Ninety-Second Street Elementary
Broadway Elementary
Maywood Elementary
Nora Sterry Elementary
West Hollywood Elementary
Emelita Street Elementary
Huntington Park Elementary
Cantara Street Elementary
Mayall Street Elementary
Amestoy Elementary

[View the full list »](#)

Lowest value-added schools

Valley Alternative Magnet
Third Street Elementary
Topeka Drive Elementary
Cowan Avenue Elementary
Wilbur Avenue Elementary
Tweedy Elementary
Broadacres Avenue Elementary
Montara Avenue Elementary
Cabrillo Avenue Elementary
Wooderest Elementary

[View the full list »](#)

As shown in the screen shot above, members of the public can search the database by individual or by school, with links to the top-performing and bottom-performing schools and to the top-performing teachers provided on the front page of the ratings site. School ratings are provided via a chart with bands spanning from least effective to most effective, and a diamond indicating where the school falls on the spectrum. Data are given for overall effectiveness, as well as for mathematics and English language arts (ELA) separately; no average or composite rating is provided. A list of each teacher within that school is then provided, with a link to his or her individual ratings. In May 2011, the data were updated to include "error bands" that describe the margin of error associated with estimates for both schools and teachers.

In order to provide context for the ratings, the newspaper offers a sidebar with links to information, such as research articles on value-added analysis, value-added model comparisons, other related articles from the *Los Angeles Times*, and the analyst's methodology. One link provides answers to 28 questions that include the following:

- Do value-added scores tell you everything you need to know about a teacher or school?
- How do I read the new teacher ratings?
- What factors did *The Times* consider in its value-added analysis?
- How accurate are the teacher/school ratings?

The ratings themselves provide sparse information; however, the *Los Angeles Times* provides a number of useful resources for understanding and interpreting the ratings. The addition of the aggregate school rating provides further context and prevents any individual teacher from being singled out as most or least effective. Nonetheless, many readers will be drawn to just the ratings and may never access this detailed, contextual information.

When the *Los Angeles Times* launched this database, even with accompanying cautions and explanatory content, there remained substantial controversy over the highly public and personalized nature of the ratings (Fensterwald, 2012). Though the newspaper provided cautions that value-added data are only one component of a teacher's overall effectiveness rating, many readers misinterpreted the findings as representative of a given teacher's capacity to teach. Though the LAUSD may have preferred to keep individual teacher's value-added rankings private, as mentioned previously, the California Public Records Act prevented them from being able to do so. Districts that are concerned about similar repercussions should consult Appendix C for information about the legal policies governing this issue in their state.

Links to more information can be found in Appendix A of this paper.

Los Angeles Teacher Ratings

Recommend 0

Tweet 0

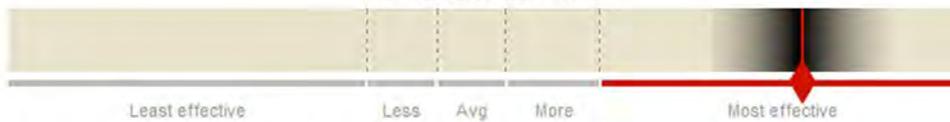
Amestoy Elementary

1048 West 149th St., Gardena, 90247

A **most effective school**, according to The Times “value-added” analysis.

Each school's rating was based on the performance of all students who took the California Standards Tests in math and English between 2003-04 to 2009-10. Value-added measures the difference between the expected progress of the student body and its actual performance. It is designed to estimate what the school contributes to students' learning. By contrast, the state's Academic Performance Index measures students' achievement level, which is tied closely to students' socio-economic background.

Overall effectiveness



Math effectiveness



English effectiveness



API: 803/1000



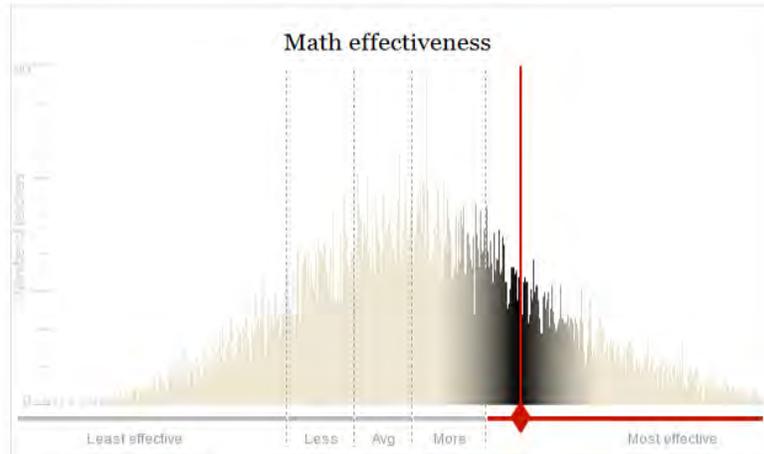
Learn more about test scores and demographics at Amestoy Elementary using the The Times' [California Schools Guide](#) ».

Doris J. Johnson

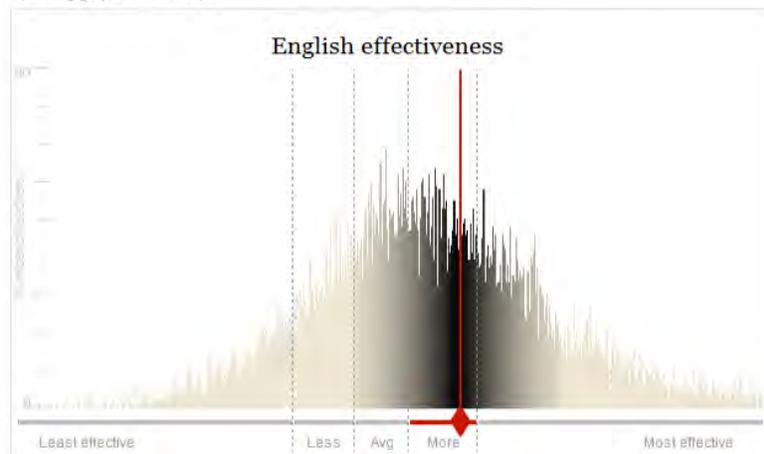
A 4th grade teacher at [Amestoy Elementary](#) in 2010

These graphs show a teacher's "value-added" rating based on his or her students' progress on the California Standards Tests in math and English. The Times' analysis used all valid student scores available for this teacher from the 2003-04 through 2009-10 academic years. The value-added scores reflect a teacher's effectiveness at raising standardized test scores and, as such, capture only one aspect of a teacher's work.

← Drag graph for more →



← Drag graph for more →



See how this teacher would change under different statistical models >

About this rating

The red lines show The Times' value-added estimates for this teacher. Johnson falls within the "most effective" category of district teachers in math and within the "more effective than average" category in English. These ratings were calculated based on test scores from 127 students.

Because this is a statistical measure, each score has a degree of uncertainty. The shading represents the range of values within which Johnson's actual effectiveness score is most likely to fall. The score is most likely to be in the center of the shaded area, near the red line, and less likely in the lightly shaded area. Teachers with ratings based on a small number of student test scores will have wider shaded range.

The beige area shows how the district's 11,500 elementary school teachers are distributed across the categories.

Johnson's LAUSD teaching history
 Years used for value-added rating. See [FAQ](#) for details.

- Amestoy Elementary, 2010 - 2004

New York City Public Schools. In late February 2012, the New York City (NYC) Department of Education publicly released effectiveness ratings for over 18,000 teachers in grades 4-8 in response to The *New York Post* and other media outlets' challenges under the Freedom of Information Act. Ratings were distributed as percentiles from 0 to 99, based on student growth data from state ELA and mathematics tests. Teachers' rankings were determined by comparing a given teacher's value-added results against other teachers of "similar experience." Percentile figures were then assigned based on where a teacher's data falls on the spectrum of effectiveness, as compared to his or her peers in terms of years of experience teaching at that grade level.

Unlike LAUSD where the *Los Angeles Times* provided its own value-added analysis, the New York media sources were given the already-calculated value-added ratings, along with their associated margins of error. The *New York Post* article offers little information about how the ratings actually were calculated, other than stating that the value-added model used in the calculations "attempts to control for a host of factors such as class size, poverty levels, and students' test scores in prior years." The article further states that NYC's model is "among the most advanced in the country" according to educational researchers (Gonen, 2012, p. 2).

On its websites, the *New York Post* includes a link to a database of teacher rankings. As a preamble to the posting of the scores, introductory text explains to site visitors that the rankings include an average error range of 35 points for mathematics and 53 points for ELA, though the estimate varies by teacher. The newspaper also states that the data were calculated for the 2007, 2008, and 2009 academic years, though not all teachers have data for all three years. Site visitors can search the database by school, though school aggregate scores are not reported.

In contrast, *School Book*, a platform for education information and debate hosted by the *New York Times*, while also searchable by school, offers school-average ratings of effectiveness for ELA and mathematics as well as individual teacher ratings. The school averages are then compared to the city-wide average for each subject. Individual-level data are displayed similar to the *Los Angeles Times* (i.e., as points on a continuum with associated error bands), though the actual percentile rank is also listed. Interested viewers can click on a teacher's name and see the number of tested students for that subject, the expected average growth, actual average growth, and the estimated added value. Scores are listed for the 2010 school year only—no multi-year averages are included. A sidebar to the ratings offers answers to frequently-asked questions, such as the following:

- What is being measured?
- Does a low rating mean my child has a bad teacher?
- Why are some ratings more precise than others?

The *New York Times* provides greater context for the ratings than does the *New York Post*; however, neither newspaper provides much information about how scores were calculated, and how to interpret the error bands associated with each score.

Importantly, prior to this public release, such findings were shared only with a teacher and his or her principal and superintendent. The value-added estimates were *not* protected as part of a teacher's evaluation, since the data were not collected for evaluation purposes. See Appendix A of this paper for links to more information.

Chapter 5: Conclusions and Implications for Future Research

It is clear from the research conducted during development of this paper that states and other jurisdictions are pursuing a number of strategies to ensure that information about teacher effectiveness is available to the right audiences at the right times and that these data may be used summatively for accountability purposes and/or formatively for school improvement and teachers' professional development. In addition to the reporting practices highlighted in earlier chapters, jurisdictions are developing and/or refining information systems that allow for accurate matching of teachers and their students; using findings from measures of effectiveness to identify those teachers who may serve as mentors to others; and disseminating different types of information about teachers' qualifications to school boards and to teacher training programs at institutions of higher education.

The SEAs in at least two states (**Rhode Island** and **Indiana**) propose to use findings from measures of teacher effectiveness to prevent students from being assigned for two consecutive years to teachers who were rated as ineffective. The state of **Florida** requires each school district to annually report to the parents of any student assigned to a teacher who has received two consecutive "unsatisfactory" ratings; **Michigan** plans to follow similar guidelines starting in the 2015-16 school year. These requirements are intended to pressure district and school administrators into making tough decisions about the right of all students to have effective teachers (Epstein & Miller, 2011).

SEAs in other states (**Tennessee** and **North Carolina**) are using reported data to examine the distribution of effective teachers across their respective states. They seek to answer questions such as, *Are the least effective teachers over-represented in rural or high-poverty schools?* They also want to better understand how graduates from various teacher preparation programs fare as in-service teachers. These research efforts track new teachers for three years and explore which institutions of higher education are associated with the highest and lowest percentages of teachers rated as instructionally effective.

Indiana and **North Carolina** are conducting presentations that explain to educators and policymakers what they can expect from reports of teacher effectiveness, with whom results will be shared, and how results can be accessed. Other states, such as **Colorado** and **Georgia** have developed print documents that describe rubrics and scoring processes and alert teachers to the performance standards to which they will be compared. These documents are available to the general public, thereby promoting full disclosure of information about their educator effectiveness evaluation systems and ready access to specific information.

Importantly, many states and districts are thinking proactively about how best to monitor short- and long-term consequences (intended and unintended) of their reporting practices. These efforts are helping to address important questions, such as the following:

- How are reports on measures of teacher effectiveness currently being used?
- What benefits are associated with these practices? Are the costs of reporting findings from measures of teacher effectiveness greater than the benefits realized?
- Whose needs are and are not well-addressed with current reporting practices?
- In what ways might changes to the type of information reported or the reporting format improve the usefulness of the reports to different stakeholder groups?

- Are changes to state or district policies compromising the intent of reporting such information?

Those responsible for report development at the state or district levels will want to follow the recommendations for best practice spelled out in this paper and document the steps they have taken to promote responsible reporting activities in their respective communities. They may consider administering needs assessments to uncover current gaps in what information can and should be more accessible and to identify the types of information that would be most useful to specific audiences. All jurisdictions are encouraged to continue to attend to lessons learned from other states and districts. It is hoped that use of this framework will ensure that outcomes from reporting are evaluated over time as an ongoing process, rather than as a static, one-time event.

Key Issues for Further Exploration

The intent of this paper is to explore strategies for promoting responsible practices when reporting findings from measures of teacher effectiveness. Yet it also seeks to raise important questions about the purpose of such reports and how reported data are used. Four such questions warrant further inquiry:

1. While value-added estimates are only one component of teacher evaluation systems in many jurisdictions, they have been reported publicly in at least two districts as the sole indicator of instructional effectiveness. What factors have contributed to the public reporting of these data and not to composite scores that reflect information from multiple sources?
2. How might the focus of reporting findings from measures of teacher effectiveness be shifted away from the teachers' impact on standardized test scores and more directly linked to efforts to improve teaching and learning?
3. How can the public's right to information be balanced with policies that protect the rights of educators and limit the types of information that can be reported?
4. What will "next generation" reports look like? How will technology support ready access to websites and databases that allow different types of information to be accessible to different audiences? What types of innovative reporting practices are emerging?

Jurisdictions seeking to responsibly report information about the effectiveness of its teachers will benefit from research that explores these issues.

References

- American Educational Research Association (AERA). (2011). *Code of ethics*. Washington, DC: Author. Retrieved September 14, 2012, from [http://www.aera.net/Portals/38/docs/About_AERA/CodeOfEthics\(1\).pdf](http://www.aera.net/Portals/38/docs/About_AERA/CodeOfEthics(1).pdf).
- American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) Joint Committee on Standards for Educational and Psychological Testing. (1999). *Standards for educational and psychological testing*. Washington DC: AERA.
- Aschbacher, P.R. & Herman, J.L. (1991). *Guidelines for effective score reporting*. CSE Technical Report. Los Angeles: UCLA Center for Research on Evaluation, Standards, and Student Testing. Retrieved September 12, 2012, from <http://www.cse.ucla.edu/products/reports/TR326.pdf>.
- Bellwether Education Partners. (2011). *Recent teacher effectiveness legislation: How do the states stack up?* Washington, DC: Bellwether Education Partners. Retrieved September 10, 2012, from <http://bellwethereducation.org/wp-content/uploads/2011/08/State-Teacher-Leg-Comparison.pdf>.
- Bill & Melinda Gates Foundation (BMGF). (2010). *Working with teachers to develop fair and reliable measures of effective teaching*. Seattle, WA: Bill and Melinda Gates Foundation. Retrieved September 10, 2012, from <http://www.gatesfoundation.org/highschools/Documents/met-framing-paper.pdf>.
- Bill & Melinda Gates Foundation (BMGF). (2012). *Gathering feedback for teaching*. Seattle, WA: Bill and Melinda Gates Foundation. Retrieved September 10, 2012 from http://metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf.
- Buddin, R. (2011). *Measuring teacher and school effectiveness at improving student achievement in Los Angeles elementary schools*. Santa Monica, CA: RAND Corporation. Retrieved September 10, 2012, from http://mpra.ub.uni-muenchen.de/31963/1/MPRA_paper_31963.pdf.
- Burling, K. (2012). *Evaluating teachers and principals: Developing fair, valid, and reliable systems*. Pearson Center for Educator Effectiveness. Retrieved September 10, 2012, from <http://www.state.nj.us/education/EE4NJ/resources/roadmap.pdf>.
- Colmers, J.M. (2007). *Public reporting and transparency*. Commonwealth Fund. Retrieved September 10, 2012, from http://www.commonwealthfund.org/~media/Files/Publications/Fund%20Report/2007/Feb/Public%20Reporting%20and%20Transparency/Colmers_pubreportingtransparency_988%20pdf.pdf.
- Corcoran, S.P. (2010). *Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice*. Providence, RI: Annenberg Institute for School Reform. Retrieved September 10, 2012, from <http://annenberginstitute.org/sites/default/files/product/211/files/valueAddedReport.pdf>.
- Educational Testing Service (ETS). (2011). *Graduate Record Examination [GRE] Guide to the use of scores*. Princeton, NJ: ETS. Retrieved September 12, 2012, from http://www.ets.org/s/gre/pdf/gre_guide.pdf.
- Epstein, D. & Miller, R. (2011). *Subtraction by distraction: Publishing value-added estimates of teachers by name hinders education reform*. Washington, DC: Center for American Progress. Retrieved September 10, 2012, from http://www.americanprogress.org/issues/2011/11/subtraction_by_distraction.html.

- Fensterwald, J. (2012, April). *Suit: Evaluation law requires student data*. San Jose, CA: Silicon Valley Education Foundation. Retrieved September 12, 2012, from <http://toped.svefoundation.org/2012/04/06/surprise-eval-law-requires-student-data/>.
- Forte Fast, E.F., and ASR SCASS. (2002). *A guide to effective accountability reporting: Designing public reports that effectively communicate accountability, assessment, and other quantitative education indicators in an easily understood format*. Washington, DC: CCSSO. Retrieved September 12, 2012, from <http://programs.ccsso.org/content/pdfs/GEAR.pdf>
- Gallagher, C., Rabinowitz, S., and Yeagley, P. (2011). *Key considerations when measuring teacher effectiveness: A framework for validating teachers' professional practices (AACC Report)*. San Francisco and Los Angeles, CA: Assessment and Accountability Comprehensive Center. Retrieved September 12, 2012, from http://www.michigan.gov/documents/mde/AACC_report_Key_Considerations_When_Measuring_Teacher_Effectiveness_352994_7.pdf.
- Glazerman, S. Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D. O., & Whitehurst, G. J. (2011). *Passing muster: Evaluating teacher evaluation systems*. Washington D.C.: Brown Center on Education Policy at Brookings. Retrieved September 18, 2012, from http://www.brookings.edu/~media/research/files/reports/2011/4/26%20evaluating%20teachers/0426_evaluating_teachers
- Gonen, Y. (2012, February 24). NYC makes internal ratings of 18,000 public school teachers available. *New York Post*. Retrieved September 12, 2012, from http://www.nypost.com/p/news/local/nyc_makes_internal_ratings_of_public_4nzYXTN1L4LOXU17G2YktO/0.
- Goodman, D.P. & Hambleton, R.K. (2004). Students test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*, 17(2), 145-220.
- Haladyna, T.M. & Kramer, G.A. (2004). The validity of subscores for a credentialing test. *Evaluation & The Health Professions*, 27(4), 349–368.
- Jaeger, R. M. (2003). *NAEP validity studies: Reporting the results of the National Assessment of Educational Progress, NCES 2003-11*. Washington, DC: USED National Center for Education Statistics. Retrieved September 12, 2012, from http://www.air.org/files/Jaeger_Reporting_NAEP.pdf
- Koretz, D.M. & Diebert, E. (1993). *Interpretations of National Assessment of Educational Progress (NAEP) anchor points and achievements levels by the print media in 1991*. Santa Monica, CA: RAND. Retrieved September 12, 2012, from http://130.154.3.8/content/dam/rand/pubs/monograph_reports/2007/MR385.pdf.
- The Los Angeles Times (n.d.). *LA Times releases teacher ratings*. Retrieved September 18, 2012, from <http://projects.latimes.com/value-added/>.
- National Comprehensive Center for Teacher Quality (TQ Center). (2011). *A practical guide to designing comprehensive teacher evaluation systems*. Retrieved September 12, 2012, from <http://www.lauragoe.com/LauraGoe/practicalGuideEvalSystems.pdf>.
- National Comprehensive Center for Teacher Quality (TQ Center). (2012). *Database on state teacher evaluation policies*. Washington, DC: Author. Retrieved September 12, 2012, from <http://resource.tqsource.org/stateevaldb/UseResource.aspx>.

- National Council on Teacher Quality (NCTQ). (2010). *State of the states: Trends and early lessons on teacher evaluation and effectiveness policies*. Retrieved September 12, 2012, from http://www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf.
- National Council on Teacher Quality (NCTQ). (2011). *State of the states: Trends and early lessons on teacher evaluation and effectiveness policies*. Washington, DC: Author. Retrieved September 12, 2012, from http://www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf.
- National Freedom of Information Coalition. (2012, April). State freedom of information laws [Web-based database]. Columbia, MO: NFOIC. Retrieved September 12, 2012, from <http://www.nfoic.org/state-freedom-of-information-laws>.
- New Teacher Project (2010a). *Teacher evaluation 2.0*. Brooklyn, NY: Author. Retrieved September 12, 2012, from <http://tntp.org/assets/documents/Teacher-Evaluation-Oct10F.pdf?files/Teacher-Evaluation-Oct10F.pdf>.
- New Teacher Project (2010b). *The widget effect: Our national failure to acknowledge and act on differences in teacher evaluation*. Brooklyn, NY: Author. Retrieved September 12, 2012, from <http://widgeteffect.org/downloads/TheWidgetEffect.pdf>.
- Rockoff, J.E., Staiger, D.O., Kane, T.J., & Taylor, E.S. (2010). *Information and employee evaluation: Evidence from and randomized intervention in public schools*. Cambridge, MA: National Bureau of Economic Research. Retrieved September 12, 2012, from <http://www.nber.org/papers/w16240>.
- Ryan, J. M. (2006). Practices, issues, and trends in student test score reporting. In S.M. Downing & T.M. Haladyna (Eds.), *Handbook of test development*. Mahwah, NJ: Erlbaum.
- Sawchuk, S. (2012, March). Access to teacher evaluations divides advocates. *Education Week*, 31(26), 1-18. Retrieved September 12, 2012, from http://www.edweek.org/ew/articles/2012/03/28/26evaluation_ep.h31.html.
- Sinharay, S., Puhan, G., and Haberman, S.J. (2011). *An NCME instructional module on subscores. Educational Measurement: Issues and Practice*, 30(3), 29-40. Retrieved September 12, 2012, from <http://ncme.org/linkservid/4638C280-1320-5CAE-6E1CAF90DFEF99E4/showMeta/0/>.
- U.S. Department of Education (ED). (n.d.). *Information quality guidelines*. Washington, DC: Author. Retrieved September 12, 2012, from <http://www2.ed.gov/policy/gen/guid/iq/infoqualguide.pdf>.
- U.S. Department of Education (ED). (2010). *Overview information: Race to the top fund, notice inviting applications for new awards for fiscal year (FY) 2010*. Washington, DC: Author. Retrieved September 12, 2012, from <http://www2.ed.gov/legislation/FedRegister/announcements/2010-2/041410a.pdf>.
- U.S. Department of Education Office of Elementary and Secondary Education. (2007). *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001*. Washington, D.C: Author.
- Wainer, H., Hambleton, R. K., & Meara, K. (1999). Alternative displays for communicating NAEP results: A redesign and validity study. *Journal of Educational Measurement*, 36(4), 301–335.
- Zenisky, A. & Hambleton, R. (2012). Developing test score reports that work: The process and best practices for effective communication. *Educational Measurement: Issues and Practice*, 31(2), 21-26.

Appendix A: Sample Teacher Effectiveness Reports from Selected Jurisdictions

Teacher-level Reports

- A. Draft Colorado Department of Education Rubric for Evaluating Teachers**
Checklist of expected skills as delineated throughout Colorado’s five proficiency levels and accompanying summative rating form.
http://www.cde.state.co.us/EducatorEffectiveness/downloads/Teacher%20Rubric/DRAFT%20Rubric_only_%20for%20Evaluating%20CO_rev%205_01_12.pdf
- B. Draft New York City Department of Education Teacher Data Report (old system)**
Guidance for teachers in using New York City’s teacher reports. Includes percentile ranks compared with citywide estimates, peer-teacher estimates, and among certain subpopulations. Includes interpretative guidelines and information about measures.
- C. Sample New York City Department of Education Teacher Data Report (new system)**
Guidance for teachers in using New York City’s teacher reports. Includes percentile ranks, results with subpopulations, guides to use and interpretation.
<http://schools.nyc.gov/Teachers/TeacherDevelopment/TeacherDataToolkit/default.htm>
- D. Draft North Carolina Department of Public Instruction Teacher Evaluation Summary Sheet**
Provides teacher with overall criterion rating with individual results for all six standards that compose the state’s proposed evaluation system.
- E. Tennessee Value-Added Assessment System (TVASS) Report**
Guidance for teachers in using Tennessee’s online educator evaluation reporting system. Includes score, results and judgments, guide to interpretation, and guide to use.
[http://www.tn.gov/education/assessment/doc/TVAAS TE Report Help 000.PDF](http://www.tn.gov/education/assessment/doc/TVAAS_TE_Report_Help_000.PDF)

School-level Reports

- A. New York City Department of Education Teacher Data Reports Frequently Asked Questions (FAQs)**
Good example of guidance about understanding and using the data, as well as a graphic representation of school-level value-added scores with numbers, percentages, and proficiency categories.
<http://schools.nyc.gov/Teachers/TeacherDevelopment/TeacherDataToolkit/FAQ/default.htm>
- B. North Carolina Department of Public Instruction Example of District/School-level Teacher Effectiveness Report**
Publicly available report that provides a cover letter explaining the teacher evaluation standards, a description of each proficiency level, and cautions about interpreting the data.
<http://www.ncpublicschools.org/docs/recruitment/effectiveness/010.pdf>

Public Reports

- C. The Los Angeles Times: Teacher Ratings Database for the Los Angeles Unified School District**
Good example of public reporting, including guidance about use, measures, and links to FAQs and researcher’s methodology.
<http://www.leginfo.ca.gov/cgi-bin/displaycode?section=edc&group=44001-45000&file=44660-44665> and <http://toped.svefoundation.org/2012/04/06/surprise-eval-law-requires-student-data/> and

<http://projects.latimes.com/value-added/> and http://mpa.ub.uni-muenchen.de/31963/1/MPRA_paper_31963.pdf

D. The *New York Post*: Teacher Effectiveness Rating Database

Provides a different example of public reporting—no guidance about use, measures, or methodology. Range of associated error is included, however.

http://www.nypost.com/p/news/local/nyc_makes_internal_ratings_of_public_4nzYXTN1L4LQXU17G2YktO/0

E. The *New York Times*: Teacher Effectiveness Rating Database

Public report of New York City’s teacher effectiveness ratings. Uses same data as the *New York Post*, but provides FAQs and information about the measures used.

<http://www.nytimes.com/schoolbook/2012/02/24/teacher-data-reports-are-released?scp=1&sq=teacher%20data%20reports%20United%20Federation%20of%20Teachers&st=ceht>

Links to Comprehensive State or District Systems with Reporting as Key Element

Delaware Department of Education’s Appraisal System II Materials:

<http://www.doe.k12.de.us/csa/dpasii/default.shtml>

District of Columbia Public Schools Guidebooks for Teacher Evaluation:

[http://www.dc.gov/DCPS/In+the+Classroom/Ensuring+Teacher+Success/IMPACT+\(Performance+Assessment\)/IMPACT+Guidebooks](http://www.dc.gov/DCPS/In+the+Classroom/Ensuring+Teacher+Success/IMPACT+(Performance+Assessment)/IMPACT+Guidebooks)

Hillsborough County Public Schools (Florida) Empowering Effective Teachers (home page for website):

<http://communication.sdhc.k12.fl.us/empoweringteachers/>

Indiana RISE Teacher Evaluation System: <http://www.riseindiana.org/>

Ohio Department of Education’s Plans for Teacher-Level Reporting:

<http://www.ode.state.oh.us/GD/Templates/Pages/ODE/OEDetail.aspx?page=3&TopicRelationID=1230&ContentID=125739>

Pennsylvania’s Value-Added Assessment System (PVAAS): <https://pvaas.sas.com/>

Public Schools of North Carolina, Educator Effectiveness Database:

<http://www.ncpublicschools.org/recruitment/effectiveness/a-e>

Appendix B. Summary of Guidance on Reporting From Seminal Resources

Guidelines from Professional Organizations

1. *Code of Ethics* (American Educational Research Association [AERA], 2011)
2. *Standards for Educational and Psychological Testing* (American Psychological Association [APA], AERA, National Council on Measurement in Education [NCME], 1999; also draft version posted for review and comment, 2011)

Resource	Summary Comments
1. <i>Code of Ethics</i>	<p>These resources (seminal 1992 article and 2011 update) focus on the ethical considerations of reporting. AERA, above all, stresses the importance of providing comprehensive information about data findings and appropriate interpretation. The report states:</p> <p style="padding-left: 40px;">Educational research reports to the public should be written straightforwardly to communicate the practical significance for policy, including limits in effectiveness, and in generalizability to situations, problems, and contexts. In writing for or communicating with non-researchers, educational researchers must take care not to misrepresent the practical or policy implications of their research or the research of others. (p. 1)</p> <p>Some of the more interesting considerations come from its guidelines around confidentiality and consent. Educational researchers are well-aware of confidentiality and consent issues; human subjects’ participation in research at certain levels requires informed consent and assurances that care will be taken place to protect the confidentiality of the participants. Measuring educator effectiveness is a form of educational research as administrators and policymakers will be likely utilizing reported data for decision making purposes at the individual, school, district, and state levels. Yet we know that in some jurisdictions educator effectiveness data can and will be made public. How do confidentiality and consent intersect with the public’s right to information? These questions will largely be governed by state- or district-level policies regarding release of information. SEAs and LEAs will need to become literate in their respective state’s legislative policies around these issues in order to remain prepared for the possibility of releasing sensitive data. Agencies can mitigate the potential consequences of these mandates by providing transparency and guidance around how the data were measured, and what inferences can and cannot be supported by them.</p>
2. <i>Standards for Educational and Psychological Testing</i>	<p>The Standards remain an important point of reference for reporting guidelines as they detail critical validity, interpretability, and defensibility issues germane to any system. Much of their guidance centers on providing the adequate information, context, and interpretation for reports of results from measures of any type. Critical questions for report developers are provided here, by key report feature.</p> <p>Statement of purpose: Have you clearly described the ways in which educator evaluation data are intended to be used?</p>

Resource	Summary Comments
	<p>Measures used/criteria against which educator was judged: Have you fully and accurately informed policymakers of the measures used and described the technical characteristics of these measures?</p> <p>How score was calculated: Are you reporting multiple scores or a composite score? Where composite scores are developed, have you included the basis and rationale for arriving at the composite? Has it guarded against spurious sources of dependency among measures?</p> <p>Score, results, judgments: Where measures of educator effectiveness involve human judgment (such as is the case with observations) have you established scoring rubrics and specified criteria for scoring? Where change or gain scores are used as part of a teacher’s evaluation data, have those scores and their technical qualities been defined and reported?</p> <p>Guide to interpretation: Are educator evaluation results reported in a way that enables accurate interpretation by all stakeholders? Are score reports accompanied by clear statements about the degree of measurement error associated with each score or classification level and information on how to interpret scores? If reporting results by group, do you give relevant contextual information to explain group differences (where possible) that enable meaningful interpretation of these differences? Have you developed and presented a rationale for each recommended interpretation and use of teacher evaluation data, together with a comprehensive summary of evidence and theory bearing on the intended use or interpretation? Have you considered and offered collateral information that may lead to alternative explanations for the educator’s performance? If validity for some common or likely interpretation has not been investigated, or if the interpretation is inconsistent with available evidence, have you cautioned potential users about making unsupported interpretations? Do you provide and explain any supplemental information that will minimize possible misinterpretations of the data?</p> <p>Guide to use: Are educator evaluation results reported in a way that decreases the likelihood of inappropriate decisions or actions based on findings? Have you identified and minimized potential unintended consequences of reporting teacher evaluation data? Have you considered the privacy of educators examined (unless a disclosure of private information is agreed upon, or is specifically authorized by law).</p>

Guidelines from Operational Testing Programs

3. *Graduate Record Examination (GRE) Guide to the Use of Scores* (Educational Testing Service [ETS], 2011)

4. *NAEP Validity Studies: Reporting the Results of the National Assessment of Educational Progress* (National Center for Education Statistics [NCES], 2003)

Resource	Summary Comments
<p>3. <i>GRE Guide to the Use of Scores</i></p>	<p>GRE score reporting policies have been adopted by the GRE Board to encourage the appropriate use of GRE scores and to protect the right of individuals to control the distribution of their own score reports. . . . Dissemination of score records should be kept at a minimum, and all staff who have access to them should be explicitly advised of the confidential nature of the scores. . . . The GRE Program recognizes the right of institutions as well as individuals to privacy with regard to information supplied by and about them. (p. 13-14)</p> <p>The GRE is a widely researched measurement tool. Strategies for effective reporting of scores and indicators of level of performance have been studied extensively. ETS’s (2011) recommendations include ensuring that an interpretive guide accompanies any score report. The guide should describe the content (knowledge, skills, abilities) that was assessed; how and by whom scores are intended to be used (including common errors, such as assuming that small score differences are meaningful, comparing scores from different subject tests, or using the scores for a purpose not supported by validity evidence); how scores are reported (e.g., percentile ranks and score ranges); and how report developers can remain alert to cases of alleged misuse. The need for consideration of information from multiple sources during decision making—not just a single score—is stressed.</p>
<p>4. <i>NAEP Validity Studies: Reporting the Results of the National Assessment of Educational Progress</i></p>	<p>National Center for Educational Statistics (NCES) has been reporting results from the National Assessment of Educational Progress (NAEP) for well over a decade. SEAs are encouraged to attend to lessons learned from NAEP. This document provides recommendations that are relevant in this context. NCES’ experiences with determining how and to what extent to report results are instructive because their experiences mirror those challenges faced by jurisdictions in reporting teacher effectiveness: there are many different audiences with different interests, requiring multiple reporting strategies. According to the author, “It is clear that no single report on NAEP results will meet the needs of its entire constituency.” To this end, NCES developed multiple strategies for reporting results, including</p> <ul style="list-style-type: none"> • Report Cards for policymakers; • Update Reports for parents and members of the public; • Instructional Reports for teachers; • State Reports for SEAs and chief state school officers; and • Focus Reports, for policy analysts and researchers. <p>Though there was clearly a concerted effort that was put forward to meet the needs and interests of its different</p>

Resource	Summary Comments
	<p>populations, the author acknowledges that little is known about the utility of these reports for their intended audiences.</p> <p>As a first step for approaching reporting results, NCES emphasizes the extreme importance of “fully and accurately” documenting the methods through which data were collected, the results of those methods, and the validity and reliability of those results (Jaeger, 2003). NCES frames its guidance on reporting strategies by considering the research questions of interest, the audiences to whom these questions should be addressed, and strategies through which the questions should be pursued. These first two steps in the process are guided by five important clarifying questions, all aimed at tailoring the report to be maximally useful to the target audience(s):</p> <ul style="list-style-type: none"> • What do various audiences find to be of interest? • What do they find to be useful? • What do they understand? • What could they validly interpret? • Among alternatives, what do they prefer? <p>What is compelling about NCES’s guiding questions is that it is clear that the researchers were thinking proactively in terms of what might be wanted, needed, and of interest to the audience receiving and using the results. Researchers, in their quest to present findings with as little bias as possible, may avoid guiding readers in interpreting findings. One example cited in the NCES article was that of a 1996 NAEP report in which the data within one of the tables indicated that there was a four-point change in the average 4th grade mathematics score in Minnesota from 1992 to 1996. The footnote added, “the change since 1992 in average scale score is significant at a 5% level of significance” (p. 19). What use can lay audiences make of this piece of information? Should stakeholders celebrate a four-point change in average mathematics scores or worry about inadequate growth? While the report writers did provide additional information about the validity of Minnesota’s test scores on the score report, that information appears to have been inaccessible to the average state stakeholder.</p> <p>NCES also provides guidance about the range of possibilities in reporting quantitative and qualitative data, both in terms of what numbers or categories are included, as well as how the chosen numbers/categories are visually depicted. Which method or methods are chosen should be determined by the answers to the questions above. For the purposes of reporting teacher effectiveness, understanding the characteristics of the target audience is critical. For example, the principal responsible for teacher evaluations will likely find it useful to have a teacher’s individual scores on various measures of effectiveness, the teacher’s overall rating as compared to her peers, and, finally, the teacher’s performance level in relation to set standards for performance (e.g., ineffective, effective, or highly effective). Given that the principal has input into hiring, firing, promotion, and/or tenure decisions, he benefits from having as much information as possible about that teacher’s strengths and limitations so that areas of growth can be celebrated and targets for improvement set. In contrast, parents may not need access to the same wide range of data.</p>

Resource	Summary Comments
	<p>Finally, the author raises critical questions, such as the following:</p> <ul style="list-style-type: none"> • Which strategies should be pursued to disseminate the chosen levels of information to the intended audiences? • What vehicles are accessible to various audiences? • Which vehicles are regularly used? • What types of information can feasibly be disseminated through various vehicles? <p>Clearly multiple platforms for reporting results are warranted to ensure that different stakeholder groups have access to these data at the right level of detail.</p>

Federal Guidelines

5. *Information Quality Guidelines* (U.S. Department of Education [ED], n.d.)

6. *Standards and Assessments Peer Review Guidance* (ED Office of Elementary and Secondary Education, 2009)

Resource	Summary Comments
<p>5. <i>Information Quality Guidelines</i></p>	<p>The purpose of the report was to outline policies and procedures for “reviewing and substantiating the quality of information it disseminates” (p. 2). According to the report, the guidelines seek to evaluate information quality through three factors: utility, objectivity, and integrity.</p> <ul style="list-style-type: none"> • Utility is defined as the extent to which the information disseminated is useful to its intended audience. To this end, the report suggests that all aggregate data should be “carefully described and documented.” Though there is not further clarification available as to what would constitute an adequate description and/or documentation of the data, we can imagine the practical significance for educator accountability purposes. For example, if a district were reporting categorical percentages of effective and ineffective teachers, it might be useful to have a narrative description explaining the strength and significance of the percentages, as well as an indication of how the educator effectiveness percentages for this reporting period compare to former reporting periods, or how those same percentages stack up against statewide data. Such information might help provide context and insight into the meaning of reported percentages for policymakers and other audiences. • Objectivity is the accuracy, reliability, and unbiased nature of (reported) information. The report cautions that disseminated information should identify all sources of information, should include a reason for providing the information, its potential uses, cautions about “inappropriate extractions or conclusions,” descriptions of any statistical techniques or mathematical operations applied to the data, and the identification of other possible sources of potentially corroborating or conflicting information. To continue the example from above, in reporting educator effectiveness at the district level, an agency might meet ED’s guidelines for objectivity by providing a statement of purpose as to why the data are being reported, what measures of teacher effectiveness were included, and how the final rating was calculated; any limitations of the data provided (with respect to accuracy and validity); and competing factors that might explain the rating. • Integrity is security or protection of the data from unauthorized access or revision. ED details statutory and administrative guidelines aimed at safeguarding this principle including: Privacy Act, Freedom of Information Act, Federal Policy for the Protection of Human Subjects, and Family Educational Rights and Privacy Act. These and other legislative laws and policies serve to protect sensitive data, though the report does not specify in what ways these laws protect individuals. It is interesting to note that in both cases where individual-educator-level data were released to the public, a Freedom of Information Act was used to uphold the public’s right to have access to those data. However, since findings from only one of the district’s measures of teacher effectiveness was reported (student growth estimates), the reporters violated the guidelines of objectivity and utility-not

Resource	Summary Comments
	<p>enough context and information about the data was provided to support valid interpretations and inferences.</p>
<p>6. <i>Standards and Assessments Peer Review Guidance</i></p>	<p>This resource offers guidance to states in reporting findings from No Child Left Behind (NCLB) requirements regarding reporting of information about Highly Qualified Teachers (HQT). These guidelines are especially timely for states wishing to report measures of educator effectiveness. Peer Review Guidance stresses the importance of providing context and interpretive guidance to support parents and policymakers in making use of the data. States are encouraged to utilize a number of formats such as manuals, bulletins, reports, and websites to report results to the public.</p> <p>States are also encouraged to report subscores for different domains (i.e., to report more than a single composite rating and to clearly indicate the pieces that compose a composite). This includes clearly specifying the weight of each subscore (i.e., the value each contributed in relation to the other pieces).</p>

Guidelines from Researchers

7. *Public Reporting and Transparency* (Colmers, 2007)

8. *Student Test Score Reports and Interpretive Guides: Review of Current Practices and Suggestions for Future Research* (Goodman & Hambleton, 2004)

9. *Guidelines for Effective Score Reporting* (Aschbacher & Herman, 1991)

Resource	Summary Comments
<p>7. <i>Public Reporting and Transparency</i></p>	<p>This respected author in the field of healthcare recommends practices in public reporting that have relevance in this context. Colmers (2007) addresses the “controversiality” of public reporting, citing the importance of monitoring how stakeholders use reported information in making decisions. The author lists a potential benefit of public reporting as helping stakeholders make informed choices.</p>
<p>8. <i>Student Test Score Reports and Interpretive Guides: Review of Current Practices and Suggestions for Future Research</i></p>	<p>Goodman & Hambleton (2004) conducted a review of states’ reports of results from annual end-of-grade assessments. They concluded that many of the student score reports in their sample had promising features, such as reporting information in alternate forms (i.e., narrative, numeric, and graphic), having different reports for different audiences, personalizing reports, and interpretative guides. However, they cite a number of weaknesses that require further attention and research including</p> <ol style="list-style-type: none"> 1. Reporting excessive amounts of information, such as many types of overall scores, but omitting essential pieces of information, such as the purpose of the test and information about how test results will and should be used 2. Information regarding the precision of test scores was not provided 3. The use of statistical jargon 4. Key terms were not always defined in the reports or interpretative guides, leaving interpretations up to users and inviting inaccurate interpretations 5. Efforts to report large amounts of information in such a small amount of space resulted in reports that appeared dense, cluttered, and difficult to read <p>They also provide general principles for reporting large-scale assessment results. These include the following:</p> <ol style="list-style-type: none"> 1. Make the report readable, concise, and visually attractive; 2. Keep the presentation clear, simple, and uncluttered; 3. Do not try to do too much with a data display; 4. Include text to support and improve the interpretation of charts and tables; 5. Minimize the use of statistical jargon; 6. Include a glossary of key terms; 7. Use bar charts to facilitate comparisons; 8. Group data in meaningful ways; 9. Use boxes or graphics to highlight main findings;

Resource	Summary Comments
	<p>10. Avoid the use of decimals;</p> <p>11. Use color in a purposeful manner (given the potential for misuse, however, the general use of color was not universally recommended);</p> <p>12. Pilot the reports with members of the intended audience; and</p> <p>13. Create specially designed reports for different audiences.</p>
<p>9. <i>Guidelines for Effective Score Reporting</i></p>	<p>Although decisions about the content of a report may be guided by political as well as technical considerations, decisions about how to <i>present</i> the information should be informed by research and experience (p. 3).</p> <p>In this document, Aschbacher and Herman (1991) tackle the issue of how to display data for efficient, effective reporting. Though the authors acknowledge the importance of content and quality of data, they are primarily concerned with the finished product and how agencies can best match reporting forms to their intended audiences. They present a summary checklist for effective score reporting, shown below (p.13):</p> <ol style="list-style-type: none"> 1. Know the audience and the purpose: <ul style="list-style-type: none"> -Use their expectations to enhance credibility and utility of the report. -Use their expertise and background knowledge to guide explanations and details. -Use language appropriate to the audience; avoid jargon. 2. Keep it simple: <ul style="list-style-type: none"> -Present information in a small number of categories. -Use summaries. -Be straightforward. 3. Be clear, accurate, comprehensive, and balanced: <ul style="list-style-type: none"> -Interpret statistics clearly and provide explanations. -Provide as rich a picture as appropriate for the audience and purpose. -Apportion space according to the importance of the ideas. 4. Use techniques to capture and focus the reader's attention: <ul style="list-style-type: none"> -Use visual techniques (e.g., graphics, colors). -Use effective headings. -Use negative wording for special effect. 5. Suit format to purpose: <ul style="list-style-type: none"> -Use numeric and adjectival descriptions of data. -Be consistent. -Select graphic format to suit purpose. -Follow guidelines for effective graphs and tables. <p>Aschbacher and Herman (1991) provide examples to underscore their points, as well as specific guidance as to which forms of presenting data are best for which purposes. The report, overall, serves as a useful reference for an SEA seeking to develop a reporting framework.</p>

Appendix C: Key Features of Educator Effectiveness Reports¹⁰

Key Feature	Low-stakes Report	Medium-stakes Report	High-stakes Report
Statement of Purpose/ Target Audience	For one-on-one professional development; information primarily shared between educator and supervisor.	To guide school improvement efforts; identify effective interventions or initiatives; support hiring or promotion decision making; primarily for school administrators.	To measure educator effectiveness for accountability purposes; to support decision-making about termination; primarily for SEA and LEA administrators, as well as policymakers; in some states, for parents.
Measures	Observations, classroom or school-based instruments.	District benchmark assessments; state assessments, observations.	State assessments; district benchmark assessments administered in a standardized fashion; observations conducted with standardized rubric.
Scoring	Judgment such as “Meets Expectations.”	Results derived from rubric; test score.	Test score; performance level; result derived from rubric.
Criteria	Norm referenced; district and/or state teaching standards.	Criterion and/or norm referenced; district and/or state teaching standards.	Criterion referenced; district and/or state performance standards.
Level of Support Needed for Interpretation	Minimally technical.	May include technical and descriptive information.	Depends on audience, but frequently highly technical.
Guidelines for Reasonable Use	Not intended for public access; low security risk.	Promote formative use; acknowledge limits to generalizability; aggregate results may be released to researchers and/or the general public.	State guidance recommended; include cautions about limitations of model or system; educator-level results may be released to parents and/or the general public.

¹⁰ For more information about “the stakes” associated with reporting information about teacher effectiveness, please see Chapter 2 of this paper.

Appendix D. Legislative and Policy Considerations that May Affect Reporting Practices¹¹

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
Alabama	Yes	Al. Code §36-25A-1 et seq.	No	No	No
Alaska	Yes, but teacher must consent.	A.S. §09.25.110 et seq.	No	Yes	No
Arizona	Yes	§39-121 et seq.	No	No	No
Arkansas	Yes, but only where there is “compelling public interest” where the evaluation led to suspension or termination.	Ark. Code Ann. §25-19-101 et seq.	No	Yes	No
California	No	Calif. Public Records Act: Gov’t Code §6250-6268	No	No	No
Colorado	Yes	C.R.S. 24-72-201 et seq.	No	Yes, under SB 191. First state to require reductions based on effectiveness rather than seniority. After completing the pilot phase, the state will require evaluation results to be used in decisions regarding a teacher’s status [CO16, p. 138; CO10, Section 5, (10)(a)(IV)(B)]. The State Council for Educator Effectiveness has recommended additional policy changes needed to allow evaluation results to be used in decisions over dismissal, compensation, and teacher recognition [CO16, pp. 154166]. Final recommendations are to be	The State Council for Educator Evaluation recommended that “[t]he state should develop and adopt statutory provisions to provide appropriate and timely protections regarding the use and reporting of educator evaluation data” [CO16, p. 158]. The Council recommended that the state pass new statutory provisions to protect the use and reporting of educator evaluation results [CO16, p. 158]. In 2009, Colorado passed H.B. 091065, which created the educator identifier system and pilot program, which was tasked with creating built in protections to

¹¹ Information in this table was synthesized from the following resources: Bellwether Education Partners, 2011; Fensterwald, 2012; TQ Center, 2012; National Freedom of Information Coalition, April 2012; Sawchuk, 2012; and from publicly accessible state websites.

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
				provided by the Council by the end of the pilot phase in 2015. [CO16, pp. 8492].	ensure educator confidentiality [CO8, 2268.5102]. The resulting pilot program has been incorporated in the larger evaluation system pilot for 20112012 [CO9, p. 1].
Connecticut	Yes, but teacher must consent.	Conn. Gen. Stat. § 1-200 et seq.	No	No	No
Delaware	No	29 Del. C. §10001 et seq. Exempt: Personnel, medical, and student files; trade, investigative and intelligence documents; charitable donations; collective bargaining; and pending lawsuits.	No	Yes, under Race to the Top requirements, LEAs must use evaluation results as a primary factor in distribution, promotion, dismissal, nonrenewal, and tenure decisions [DE25; DE7, secs. 1270 & 1273; DE2, sec. 106A (7.0)]. In addition, evaluation results are included in qualifications for continuing licensure [DE3, sec. 1511 (3.23.3)]. Teachers may be terminated by the LEA for demonstrating a pattern of ineffective teaching [DE7, secs. 1270 & 1273; DE2, sec.106A (7.0)].	No
District of Columbia	Yes	Freedom of Information Act -- §§ 2-531 et seq.	No	Yes, requires for dismissal, step increases and holds, and performance bonuses [DC6; DC15, p. 106]. Teachers rated ineffective or who receive any combination of ineffective and minimally effective for 2 consecutive years are subject to removal. Teachers who receive a highly effective rating for two consecutive years are eligible for compensation bonuses and salary increases of up to \$27,000. The type of school and subject taught	DCPS releases aggregate numbers on the number of teachers rated highly effective and those rated highly effective for two years in a row; the number of teachers who received pay bonuses; and the number of teachers who received separation notices as a result of consistently low evaluation results or failure to comply with licensure requirements [DC6]. Results are released to the general public at the end of each year via a press release [DCPS personal

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
				determines the size of the bonus and the increase in compensation. [DC6; DC17; DC18].	communication]. Teachers may appeal their scores or ratings through an appeals process. Access to IMPACT data is restricted to certain members within the DCPS community [DCPS personal communication].
Florida	Yes, since 1983 Florida has allowed the public to access parts of teacher evaluative data the year after the review was conducted.	Fla. Stat. sec. 119.01 et seq.	Yes, as of 2011-12 school year-SB 736 requires each school district to annually report to the parents of any student assigned to a teacher who has received two consecutive “unsatisfactory” ratings, two “unsatisfactory” ratings in the last three years, or three consecutive “needs improvement” ratings.	Yes, and reductions are based on effectiveness rather than seniority Yes, requires for dismissal, renewal, and compensation [FL17 (4)(b)(2), FL18 (2) (c)]. The LEA’s instructional personnel and school administrator compensation system awards salary increases in relation to sustained student performance [FL18 (2)(c)].	Yes, SEA issues annual report on the percentage of teachers receiving each performance rating, disaggregated by school and district
Georgia	Yes	O.C.G.A. §50-18-70	No	Yes, the state plans to link human resources decisions to the evaluation system as a part of its Race to the Top grant [GA5,p. 1].	No
Hawaii	No	Haw. Rev. Stat. §91-1 et seq. Exempt: Medical, psychiatric, or psychological information; criminal investigations; social services or welfare benefits information; personnel files; fitness to be granted a license; and personal	No	Yes	No

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
		recommendations and evaluations.			
Idaho	Yes, but teacher must consent.	Idaho Code §9-337 et seq.	No	No	No
Illinois	No, districts are precluded from reporting results of individual teachers.	5 I.L.C.S. 140 et seq. Exemptions: Numerous exemptions including personnel information, disciplinary actions and vulnerability assessments.	No, districts are precluded from doing so.	Yes, teachers who receive an unsatisfactory rating and fail to improve and/or teachers who receive two unsatisfactory ratings within seven school terms. Effectiveness rather than seniority is a primary consideration in reductions.	No, SB 315 prohibits public disclosure of teacher evaluations.
Indiana	No	§§5-14-3-1 et seq. Exempt: Trade secrets; university research; and certain educational and medical records. Discretionary exemptions include: Personnel files; diaries; journals; and personal notes.	Yes, district must inform parents if their child is assigned to an “ineffective” teacher two consecutive years. State also requires that no student be instructed for two consecutive years by teachers who have been rated ineffective in the previous year of teaching. If the LEA is unable to comply with this state rule, the school is obligated to notify the parents of each affected student that their child has been placed with a teacher who was rated ineffective in the previous year [IN4, sec. 202811.57(ad)].	Yes, law defines “incompetence” as grounds for dismissal, as either an ineffective rating on two consecutive performance evaluations, or an ineffective or needs improvement rating three out of five years. Yes, evaluation results are to be considered in compensation [IN3, sec. 202891(c)(2)], dismissal [IN2, sec.20287.51(e)(4)(AC)], probationary versus professional status [IN1,sec. 202867.5], and reduction in force decisions [IN2, sec. 20287.51(d)].	Yes, SEA annually reports to the public on teacher evaluation results for each school, district, and teacher preparation program [IN4, sec. 202811.59(a)]. Each LEA reports to the SEA the results of performance evaluations for the certificated employees, including the number of staff rated in different categories; no personally identifiable information is submitted, however [IN4, sec. 202811.5 (b)(1–2)]. Each year, the SEA reports the results of the staff performance evaluations to the State Board of Education and the public via state website. These results will show the aggregate of certificated employees of each school and school corporation; and the aggregate of graduates of each teacher preparation program in Indiana [IDOE, private communication].

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
Iowa	No	Iowa Code §22.1 et seq. State access law has 65 exemptions including personal information of public school students, hospital and medical records, trade secrets, peace officers' investigative reports, and attorney work related to litigation.	No	Yes, requires for continuing licensure, contract renewal, and promotion [IA18, 284.7(2)]. For beginning teachers, evaluation results impact movement along the licensure levels in the state and contract renewal [IA18, 284.7(1)]. For career teachers, Iowa legislation calls for evaluation results to be included in decisions on whether a teacher advances up the proposed career ladder and may be included in termination decisions [IA18, 284.7(2)]. Iowa conducted several pilot programs on pay for performance systems for teachers; these systems have not, however, been rolled out statewide [IA9].	The state has a statewide longitudinal data system (EdInsight), with data starting in 2004 and with unique student and teacher identifiers. No policy plans have been created, however, to link students to individual teachers for the purpose of summative evaluation [IA11,C4;IA12]. The SEA states that it retains a team that works with LEAs on data validation and quality and also includes a series of "interdependent checks" through the Electronic Access System for Iowa Education Records (EASIER) [IA11,C3].
Kansas	No	K.S.A. 45-215 et seq.	No	No	No
Kentucky	Yes, but only at the discretion of a third party (records custodian, school district official, or state official).	K.R.S. 61.872 et seq.	No	Yes, dismissal [KY1, sec. 156.557 and sec. 5 (a)(c)].	No
Louisiana	Yes, but teacher must consent.	La.R.S. 44:1 et seq.	No	Yes, required for dismissal and certification decisions [LA10, secs. 3886(A) and 3902 (C)(2)(a)(3)]. LEAs maintain control over personnel decisions such as hiring and pay. Educators who are rated ineffective will be placed on an intensive remediation program. Disciplinary action is taken if a teacher is rated ineffective two years	Each local school board has been tasked with assisting in the development of systems "necessary for the rapid transmission of evaluation information and reports to teachers and administrators and for maintenance of the confidentiality of such information" [LA10, sec. 3904 (A)(4)]. The superintendent will make available to the public "the data that

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
				in a row. In addition, educators who are rated ineffective for three years during their re-certification period will not be recertified by the state unless the local school board makes an appeal [LA5].	may be useful for conducting statistical analyses and evaluation of education personnel but shall not reveal information pertaining to the evaluation report of a particular employee." Public information may include school-level student growth data [LA10, sec. 3884 (C)]. SEA has a teacher identifier system with the ability to match teachers to students [LA12, p. C2].
Maine	Yes	M.R.S.A. Sec. 402 et seq.	No	No	No
Maryland	No	Md. State Gov't Code Ann. §§10-611 et seq. Md. State Gov't Code Ann. §§10-611 et seq. Exempt: Information that invades individual privacy; trade secrets; public policy development memos; and investigative materials.	No	No for the pilot programs. Yes for the fully implemented system. The state prohibits the use of evaluation results in human resource decisions during the two pilot years (20112012 for the seven school systems and 20122013 for the entire state pilot) [MD16; MD18]. The state specifies that, after the conclusion of the pilot, the results of professional evaluations must be considered in tenure decisions [MD3, sec. 6202(b)(3) (ii)(1)]. The state is currently working on model compensation systems, but final details have not been released.	The state is collecting and linking teacher and student data using the longitudinal data system with unique student identifiers [MSDE, private communication].At the database level, the personally identified information for students and teacher information is encrypted. Access to the database is controlled through Oracle Database Vault. Within the external reporting applications there are a login authorization, an application authorization, and authentication based on the application role defined for the individual accessing the application [MSDE, private communication].
Massachusetts	No	G.L. c. 66, sec. 10. Exempt: Information that would invade individual privacy; trade secrets; public policy development	No	Yes, evaluation results may be used in decisions for tenure, compensation, promotions, and dismissals [MA7, slide 6; MA1, sec.	The regulations specify that all information that may be linked to an individual educator shall be considered personnel information and shall not be subject to disclosure

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
		memos; and investigative materials.		35.08.07].	under the public records law [MA1, sec. 35.11(6)]. The state uses unique identifiers for both teachers and individual students to protect sensitive personal information [MA4, pp. 5758]. LEAs are required to submit individual-level educator evaluation data to the SEA; however, any educator evaluation information able to be linked to an individual educator is considered "personnel information" and "shall not be subject to disclosure under the public records law" [MA1, sec. 35.11(6)].
Michigan	Yes MCL 380.1249 http://www.legislature.mi.gov/(S(auijs52ewxripfve2iojx5um2))/mileg.aspx?page=GetObject&objectname=mcl-380-1249	Mich. Comp. Laws Ann. §15.231 et seq	Yes, law will not go into effect until the 2015-2016 school year; teacher must have two or more unsatisfactory evaluations.	Yes See House Bills 4625, 4626, 4627, and 4628: http://www.legislature.mi.gov/(S(nsd cwh551ndrxv55yywxb055))/mileg.aspx?page=getobject&objectname=2011-HB-4625	Yes, public reporting required at aggregate level by schools using effectiveness labels, as required by the State Fiscal Stabilization Fund (personal communication from Carla Olivares, MDE, 2012). Michigan Council for Educator Effectiveness has replaced Governor's Council for Educator Effectiveness. See April 2012 progress report: http://www.michigan.gov/documents/mde/Interim_Progress_Report_MCE_E_383698_7.PDF
Minnesota	No	Minn. Statutes 13.01 et seq. Exempt: Juvenile court records; and some personnel information.	No	No	A pending bill in the state legislature mandates that school report cards that are publicly posted include the number of teachers in each performance category.

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
Mississippi	Yes, but only at the discretion of a third party (records custodian, school district official, or state official).	Miss. Code Ann 25-61-1 et seq.	No	No	No
Missouri	No	Mo. Code §610.023 et seq. Closed: Legal actions; leasing, purchase or sale of real estate; personnel matters; the state militia; health examinations; testing materials; negotiations with employees; and sealed bids.	No	No	No
Montana	Yes, but only at the discretion of a third party (records custodian, school district official, or state official).	Montana Code 2-6-101 et seq.	No	No	No
Nebraska	Yes, but only at the discretion of a third party (records custodian, school district official, or state official).	Nebraska Statutes §84-712 et seq.	No	No	No
Nevada	Yes	N.R.S. 239 et seq.	No	Yes	No
New Hampshire	No	R.S.A. Ch. 01-A et seq. Exempt: Investigative files; parole and pardon board records; student records; and certain commercial/	No	No	No

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
		financial information.			
New Jersey	Yes	N.J.S.A. 47:1A-1 et seq.	No	No	No
New Mexico	Yes	14-2-1 NMSA 1978 et seq.	No	No	No
New York	Yes	NY Pub. Off. Law sec. 84 et seq.	No	Yes, required. The evaluation results will be a significant factor in employment decisions, including promotion, retention, tenure determinations, termination, and supplemental compensation, as well as teacher and principal professional development [NY5, section A]. The LEA will consider annual professional performance reviews as a significant factor at the time of employment decisions [NY5, section N6(c)]. In August 2011, the State Supreme Court invalidated portions of this provision, and the SEA is currently appealing the decision [NY5, R].	No
North Carolina	Yes	G.S. §132-1	No	Yes, for continuing licensure, dismissal, promotion, retention, tenure, and compensation [NC11, pp. 111, 132, 133, 139, 140]. Teacher evaluation data are used to identify ineffective teachers in need of remediation and possible dismissal [NC11, pp. 111, 132].	School and LEA aggregated evaluation results are shared with the public on an annual basis [NC11, p. 138]: the proportion of teachers whose students demonstrate expected growth or exceeds expected growth; the proportion of teachers at each proficiency level; and the proportion of teachers who moved beyond the developing “rating within the established time period” [NC11, p. 138]. Evaluation results are shared with the community via

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
					various information products in response to legislative and other public data information requests are published on a public website [NC11, p. 88].
North Dakota	Yes	N.D.C.C. §44-04-18 et seq.	No	No	No
Ohio	Yes, but only at the discretion of a third party (records custodian, school district official, or state official).	Ohio Rev. Code sec. 149.43 et seq.	No	Yes. Ohio HB 153 (2011) includes provisions for the use of performance evaluation results in teacher compensation [OH14, p. 5]. Decisions over termination or progress toward continuing licensure must include reference to teacher evaluation results [OH16, p. 3].	Ohio HB 290 (2009) removes legislative restrictions that previously prevented sharing unique student identifiers with Institutions of Higher Education, preventing complete data linkages. The bill also allows the state to enter into agreements to use the data for research and analysis of program effectiveness and provides teachers with access to the data [OH10, Section C1]. Ohio also is working in partnership with the Center for Educational Leadership and Technology in collaboration with five other states to determine best practices for student teacher data linkages [OH1; OH10, pp. C23].
Oklahoma	Yes, but only at the discretion of a third party (records custodian, school district official, or state official).	OK Stat. tit. 51, §24A.1 et seq.	No	Yes	No
Oregon	Yes, but only at the discretion of a third party (records custodian, school district official, or state	O.R.S. 192.410 et seq.	No	No	No

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
	official).				
Pennsylvania	No	65 Pa. Stat. Section 66.1 et seq. Exempt: Most investigative records; birth and death records; and accident reports. Also see: Office of Open Records	No	Yes	No
Rhode Island	No	R.I. Gen. Laws §§38-2-1 Exempt: Presumption of disclosure with few exemptions.	No. However, the state's Race to the Top application states that "No child will be taught by a teacher who has been rated ineffective for two consecutive years" [RI12, sec. D13]. The state has a commitment to ensure that a student who was assigned to a teacher rated as ineffective in a given year must be assigned the following year to a teacher rated effective or highly effective. After two consecutive years of ineffective ratings, the LEA must dismiss the teacher [RI12, section D39].	Yes, requires LEAs to use evaluation results in decisions for teacher distribution, promotion, dismissal, renewal, tenure, and certification. LEAs are required to report information on educator evaluation and dismissals each year to ensure compliance. In their Race to the Top application, the state said it is encouraging LEAs to adopt performance-based compensation systems by 2015 [RI1, p. 35; RI12, secs. D25, D27, D30]. Starting in 2015, the SEA has committed to ensuring that the granting or renewal of full certification will depend on a teacher achieving acceptable effectiveness ratings [RI12, sec. D29].	Yes. In the 2012-2013 school year, the SEA will make the Educator Performance and Support System (EPSS) the primary platform for teachers and evaluators to collect and manage data for the Student Learning, Professional Practice, and Professional Responsibility portions of the evaluation. EPSS is an online platform that will enable teachers and evaluators to engage in two-way communication, schedule observations and conferences, and provide tools for self-assessment and observations [RI19, p. 15].
South Carolina	Yes	S.C. Code Ann. §30-4-10	No	No	No
South Dakota	No	S.D.C.L. 1-27-1 et seq. Exempt: Savings and loan association reports; school records; juvenile court records; adoption records;	No	No	No

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
		hospital licensing and inspection information; and medical research information.			
Tennessee	Yes	Tenn. Code Ann. 10-7-503 et seq. While value-added scores have been calculated via TVAAS for many years, these data are shared only between a principal and a teacher as one component of the evaluation process; no teacher's overall rating has been released to the public. However, TDE has cautioned that they would be legally obligated to release teacher ratings if an Open Records request were filed.	No	Yes, evaluations demonstrating "below expectations" or "significantly below expectations" are grounds for dismissal. The state requires that annual evaluation be a factor in personnel decisions, including promotion, retention, tenure, and compensation [TN2, slide 3]. The First to the Top Act requires that evaluation outcomes be used for personnel decisions that may include recruitment, promotion, compensation, and dismissal. All these personnel policies are made at the district level and are not determined by the state. Tenure policies are part of the law, however, including new eligibility requirements [TN23]. [TN15, p. 1].	The state senate of Tennessee recently voted (SB 1447) to prohibit teacher evaluation data from being shared with parents and the public after announcing in March that these data would be made public.
Texas	No	Tex. Rev. Civ. Stat. Ann. art. 6252-17a Closed: Attorney consultations; real estate transactions; hearings of the ethics commission; personnel and individual student matters; and emergency medical service quality reviews, etc.	No	Yes	No

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
Utah	No	63-2-101 et seq.	August 2012 update: State Board of Education voted to encourage principals to share classroom level achievement data with parents who request it	No	No
Vermont	No	1 V.S.A. §§315 et seq. Exempt: Twenty exemptions including: Personnel files; criminal investigation records; tax documents; and location of historical/ archaeological sites.	No	No	No
Virginia	Yes	Va. Code Secs. 2.2-3704 - 3706	No	No	No
Washington	Yes, but only at the discretion of a third party (records custodian, school district official, or state official).	Wash. Rev. Code § 42.56 et seq.	No	Yes [WA5, sec. 202(4)(ab)] Under SB 6696, if a teacher's performance is "not judged satisfactory" according to district evaluation criteria and does not demonstrate necessary improvement within a probationary period, such a situation "shall constitute grounds for finding probable cause" to discharge the teacher from his or her current assignment [WA5, sec. 202(4)(ab)].	Not specified, but state law, enacted by SB 6696, requires all districts to report to the state annually on the number of staff members who were rated in each performance category [WA5, sec. 201(2)(a)]. The state de-identifies all data using a unique personal identifier [WA13, p. C2] and uses a secure, data encryption process when schools submit data to the state data warehouse [WA3, p. 18].
West Virginia	Yes	W.Va. Code § 29B-1-1 et seq.	No	No	No
Wisconsin	No	Wis. Stat. 19.31 et seq.	No	No	No

State	Disclosure of Teacher Evaluation Information Allowed (e.g., Sunshine Laws)	Freedom of Information Acts	Parental Notification Required if Child Has Ineffective Teacher	Dismissal of Ineffective Teachers Allowed	Public Reporting Mandated at the State Level
		Exempt: Certain investigative records, computer programs and trade secrets; public library circulation records are confidential.			
Wyoming	No	Wyo. Stat. § 16-4-201 et seq. Exempt: Some law enforcement investigation records; testing materials; details of state institutions' research projects; labor negotiations; school board and university student disciplinary records; and medical records of publicly funded hospitals.	No	No	No