

Creating a Better System:
Recommendations for a Systemic Approach to Improving Educator Effectiveness



Submitted to the Governor's Task Force on Educator Evaluation
By EQUATE
February 28, 2011
Briefed (February 7, 2011)

PURPOSE

Governor Chris Christie's Executive Order Number 42 directed the *Task Force on Educator Effectiveness* (the Task Force) to explore educator evaluation models and to recommend a statewide evaluation system "that will inform decisions about various school personnel policies, including professional development, promotion, compensation, merit-based bonuses, tenure, and reductions in force and separations." The Executive Order also stipulated that recommendations must include student achievement measures that would comprise 50 percent or more of an educator's evaluation and other demonstrated practices that would comprise the balance of the evaluation.

Recognizing this Executive Order's potential for meaningful improvement in our school system, and the need for investment, expertise and experience on the part of those charged with formulating such an important set of recommendations, we formed *Educators for the Quality Assessment of Teacher Effectiveness* (EQuATE). Our purpose is to shape and inform the work of the Governor's Task Force. EQuATE draws membership from the research community, parents, boards of education, school leaders, policymakers and teachers who have demonstrated considerable success in improving large complex systems in the public and private sectors (See *Appendix 1*).

After the battles are fought in Trenton, new regulations are promulgated, and policy wonks retreat to safer shores, EQuATE members will still be here trying to improve outcomes for *all* students in New Jersey. We seek to ensure that the State's solution is complementary to the initiatives already underway in our successful districts and schools and that our policymakers recognize that the "one best system" is that which allows each district to leverage its own experiences and tested best practices.

In preparing this brief, EQuATE reviewed the existing literature on student outcomes *that matter*, domains of teaching and leadership practice, methods of evaluating practice, and systems for the evaluation and improvement of teaching and leadership. Our recommendations flow from our review of the literature and experience in teaching, managing, leading and governing in schools and in private industry. We believe that our recommendations will take a strong statewide school system to the next level of excellence: a school system that offers excellence and equity to *all* students and a commitment to the continuous improvement of our professional practice.

EXECUTIVE SUMMARY

Governor Chris Christie issued Executive Order Number 42 on September 28, 2010, which called for the creation of a *Task Force on Educator Effectiveness* (the Task Force). The Task Force is responsible for exploring educator evaluation models and recommending a statewide teacher evaluation system.

Educators for the Quality Assessment of Teacher Effectiveness (EQuATE) was founded in response to Executive Order Number 42 and seeks to shape and inform the work of the Task Force based on the experience of its members. EQuATE membership included researchers, parents, representatives from boards of education, school leaders, policymakers and teachers (See *Appendix 1*).

Based on a thorough analysis of existing literature, evaluation programs, and its members' own experiences, EQuATE is recommending the following:

1. Appoint a “guiding coalition” to develop an aligned “Department of Education – Local Education Authority” (DOE-LEA) system for continuous improvement of teaching and school leadership through an inclusive process by December 2011.
2. Develop a balanced teacher evaluation framework and process. The DOE-LEA evaluation system should:
 - Empower teachers and school leaders to customize, adopt and implement a process and framework that is LEA-specific;
 - Reduce the weight given to standardized test-based measures of student achievement;
 - Select pupil progress indicators with wisdom;
 - Incorporate all domains of professional practice into the evaluation framework;
 - Develop guidelines, standards, processes and training around the proper use of data;
 - Design for transformative change rather than simply technical change; and
 - Conduct a cost-benefit analysis.
3. Complete a pilot of the DOE-LEA process and framework by July 2015.
4. Provide LEAs with a set of criteria by which they might opt-out of the proposed statewide system based on performance and locally developed educator improvement systems. One alternative should be a locally negotiated Peer Assistance and Review (PAR) process designed to identify and assist under-performing teachers before making a professional judgment regarding their fitness for continued service.

RECOMMENDATIONS AND ANALYSIS

RECOMMENDATION 1: Appoint a “guiding coalition” to develop an aligned DOE-LEA system for continuous improvement of teaching and school leadership through an inclusive process by December 2011.

This objective—developing a balanced teacher evaluation system—is new territory for everyone. There are no off-the-shelf, well-proven programs, software, or models available to be grafted onto New Jersey (See *Appendix 4*). Trial and error and readjustment should be the expectation. We recommend that the DOE appoint a “guiding coalition” with influence in key stakeholder groups as well as the knowledge, skills and experience to craft a vision and strategy for a system to improve the quality of teaching and school leadership statewide. The DOE lacks the knowledge of the culture of each school district in which the improvement of professional practice/evaluation processes reside. Moreover, it lacks the constitutional authority to impose such a prescriptive solution on 600+ Boards of Education (See *Appendix 2*). Therefore, it is essential for the coalition to form a DOE-LEA partnership and galvanize stakeholders around the importance of this work.

The coalition would use qualitative and quantitative data to reveal the DOE’s problem of practice for which a statewide evaluation framework is the presumptive solution. Student achievement data should be a part of this analysis; however, for research-based reasons explained elsewhere in this document, inferences drawn largely from student achievement data are too tenuous a basis from which to draw conclusions on the state of educator effectiveness in New Jersey. Nothing less than a systemic audit of educator practices benchmarked against established professional standards would suffice as a baseline set of measures. The coalition would use these same metrics to evaluate progress in achieving the policy intervention’s goals. For the DOE to conduct such an audit on a regular basis would be bureaucracy-gone-amok, therefore, it is strongly recommended that the DOE partner with LEAs in collecting local data that will be of use to the LEAs in their school improvement processes.

RECOMMENDATION 2: Develop a balanced teacher evaluation framework and process.

The DOE-LEA evaluation system should:

- *Empower teachers and school leaders to customize, adopt and implement a process and framework that is standards-driven but LEA-specific.* Each LEA, like each charter school, has unique needs, rooted in culture, and the solution that each devises will be somewhat different from the next. However, each LEA will strive to achieve standards of professional practice on which all professionals can agree. As long as an LEA is meeting the standards for a “thorough and efficient” education (T&E), it should be permitted to choose the system that works best for its community—including an alternative process that will be described below. The coalition, however, should be charged with gaining consensus on standards and principles of professional practice through well established organizations like the National Board for Professional Teaching Standards.
- *Reduce the weight given to standardized-test-based measures of student achievement.* Research studies show that the teacher’s effect on value-added scores, based on these kinds of tests, accounts for only 3-4 percent of the variation.¹ Fully 90 percent of the variation in VAMs is attributable to student characteristics and the interaction of learning/test-taking styles with the instruments used to measure achievement; it’s not the teacher. To ascribe a weight to this

¹ Chiang, Hanley and Schochet, Peter (2010). *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains (NCEE 2010-4004)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences.

measure that exceeds its explanatory power would be malpractice at best. Attribution issues aside, the margin of error associated with the estimates of teacher effectiveness is so large that it adds little value to the evaluation of teachers. In a recent analysis of VAM data from New York City, the margin of error was as high as 28 percent.² This meant that a teacher who was ranked in the 43rd percentile was statistically indistinguishable from a teacher ranked in the 15th or 71st percentiles.³ The chances of misidentifying a high or low performing elementary teacher, using 3 years of error-free value-added data is about one chance in four.⁴ The probability of misidentification only increases with fewer years of data, fewer students and imperfect data (see *Appendix 3* for a fuller explanation of statistical issues). In addition, research is clear that focusing on tested achievement in the manner suggested by the Executive Order severely limits the kind of learning that is assessed and leads to the kind of teaching and learning that is least desirable.⁵ A recent analysis of the New York State test found that 50 percent of a student’s score was attributable to 15 percent of the State’s content standards.⁶

- *Select pupil progress indicators with wisdom.* Arguably, the first and most important decision that the coalition will make concerns which student outcomes matter.⁷ We believe that the “student outcomes *that matter*” should include standardized and local achievement measures but other variables, such as LEA-developed common assessments (across all subjects), educational attainment and “non-cognitive” or socio-emotional development measures.⁸ These additional measures are not only important factors in predicting future outcomes, but they better represent the effects that great educators can have on young people and mitigate the distortions that a single measure or single group of multi-collinear measures (see *Appendix 3*) can have on both what is taught and how it is delivered. Most importantly, these outcomes—not test scores—are correlated to things that truly matter: participation in civic life, individual earnings and economic growth.
- *Incorporate all domains of professional practice into the evaluation framework.* There are multiple domains of professional practice that bear on student outcomes, and their indicators vary by content area and grade level.⁹ To the extent that student achievement measures are used in educator evaluation, those indicators should be used as part of a holistic assessment and in a discerning fashion. They should be used along with a broader set of qualitative and quantitative measures including evidence-based self-reflection, feedback from instructional rounds or lesson study, portfolio peer review, supervisor observation/evaluation, and stakeholder feedback. These

² Jacob, Brian and Lefgren, Lars. (2008). “Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education.” *Journal of Labor Economics* 26:101-136.

³ Corcoran, Sean (2010). *Can Teachers be Evaluated with Student 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, Brown University, p.21.

⁴ Chiang and Schochet (2010), *op cit.*, p. 35.

⁵ Baker, Eva, et al. (2010). *Problems with the Use of Student Test Scores to Evaluate Teachers.* Economic Policy Institute Briefing Paper #278. Washington, DC: Economic Policy Institute.

⁶ Corcoran (2010), *op. cit.*, p. 16.

⁷ Test scores are the standard metric against which teaching effectiveness is gauged; however, test scores have not been shown to have any predicative validity other than of other test scores. Economists measure the effectiveness based on outcomes that contribute to worker productivity (i.e., worker earnings) or to standards of living (i.e., per capita economic growth). The critical student outcome that factors into these models is not achievement but educational attainment. See Mankiw, N. Gregory, Romer, David and Weil, David N. (1992). “A Contribution to the Empirics of Economics Growth.” *Quarterly Journal of Economics*, 107, May, 407 – 437; Barro, Robert J. and Jong-Wha Lee (2001). “International Data on Educational Attainment: Updates and Implications.” *Oxford Economic Papers*; and Barro, Robert J. (1991). “Economic Growth in a Cross-Section of Countries.” *Quarterly Journal of Economics*, 106, May, 407 – 443. Another student outcome that has recently been found to have an impact of worker productivity is “non-cognitive” development. See Heckman, James and Krueger, Alan. 2003. *Inequality in America: What Role for Human Capital Policies.* Cambridge: MIT Press, 92-93, 314.

⁸ The literature on the role of non-cognitive measures in influencing lifetime outcomes is broad. Economic analyses may be found in the work of Nobel-prize winner, James Heckman. See Heckman, James and Rubinstein, Yona. (2001). “The Importance of Noncognitive Skills: Lessons from the GED Testing Program.” *American Economic Review*. 91(2): 145-149.

⁹ Danielson, Charlotte. *Enhancing Professional Practice: A Framework for Teaching, 2nd edition.* Alexandria, VA: ASCD, 2007. See also, Kane, Thomas J. and Cantrell, Steven. *Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project.* Seattle: Gates Foundation, 2010

indicators evince practices that bear on student outcomes, though they may not show up in a test score or value-added measure due to aforementioned reasons. The inclusion of indicators of these domains would preclude some of the undesirable consequences that derive from the weighty focus on test-score-based achievement.¹⁰ New Jersey's highly effective practitioners deserve a more sophisticated tool to facilitate the evaluation and improvement of practice.

- *Develop guidelines, standards, processes and training around the proper use of data.* Data on all these varying indicators requires deep statistical understanding and experience to interpret properly. One need only consider how quickly policymakers and media (e.g., *LA Times*) embraced VAMs and their presupposed “meaning” without consideration for statistical concerns noted above. Raw evaluation results, like VAMs, should not be available to the general public; they should be used only by well-trained professionals. In our view, the purpose of employing pupil progress indicators, or any other data, is to inform and improve professional practice. It is neither to sort nor screen, retain nor fire, reward nor punish educators (or students). The statistics lack the power for such determinations, and it would be malpractice to use them to any of these ends. It is the whole of an educator's contributions, judged by a competent supervisor or professional review board that should be the basis for employment decisions.

The misuse of data can lead to more narrowly focused curricula, pedagogically unsound teaching practices, counter-productive school procedures (e.g., tracking students, firing misidentified teachers, etc.), and higher drop-out rates (but, ironically, higher average test scores). Training on the use of data—specifically, mining them for meaning at the intersection of cognition, language acquisition, learning disabilities and culture/poverty—should result in more precise interventions for educators and students and more constructive school policies concerning differentiation of instruction, retention and tracking. We recommend drawing upon the work of the National Board of Professional Teaching Standards (NBPTS) for these guidelines.¹¹

- *Design for transformative change rather than simply technical change.* The coalition must balance the public's interest in accountability with a learning organization's need for purpose, autonomy, personal mastery, fair dealing and actionable feedback to improve professional practice (as defined for our purposes as the NBPTS standards). A process that engages educators in improving their practice through feedback from multiple sources (i.e., state tests, local assessments, 360-feedback, learning teams), as opposed to solely external judgments of practice by supervisors or by computers crunching student test scores, is more likely to transform practice and improve student outcomes than a compliance-driven system. A system in which feedback is welcomed and due process ensured, would speed the evolutionary process. Changing an evaluation framework is mere technical change. It tinkers at the margin and will pass with the next-best-thing. What will ultimately transform educational practice will be educator ownership of and belief in the process surrounding the evaluation framework. It is this kind of adaptive and transformative change that we are seeking.
- *Conduct a cost-benefit analysis.* New Jersey's Race-to-the Top (RTTT) application included requests for funding technology infrastructure, technology support, and building district capacity to support the use of value-added models. Although we could not locate New Jersey's budget figure for these items, using Massachusetts's RTTT budget as a guide, we estimated the upfront cost of this innovation at \$50 million to the State alone.¹² The ongoing cost to maintain the

¹⁰ Campbell, Donald. (1979). “Assessing the Impact of Planned Social Change.” *Evaluation and Program Planning*. Vol. 2, Issue 1, pp. 67-90.

¹¹ See National Board for Professional Teaching Standards at http://www.nbpts.org/for_candidates/the_portfolio#doc. See also Universal Design for Learning at <http://cast.org/udl/index.html>.

¹² Massachusetts's RTTT application included \$23.2 million to transform their existing data systems, \$4.9 million in data support and \$6.8 million in capacity building at the LEA level. Since NJ is larger, we scaled up costs proportionately.

evaluation system, to ensure the accuracy of data input and to offset additional LEA support staff costs would significantly increase this present value. Given the limited return on this investment noted above (as compared with the judgment of supervisors) and the economic times that have resulted in dramatic reductions in state aid for constitutionally prescribed services, the initial analysis would suggest that the state should commit to this course if and only if it receives grant support.

RECOMMENDATION 3: Complete a pilot of the process and framework by July 2015.

Conducting a pilot of the evaluation framework and process is essential for a number of reasons. Most importantly, it will provide data to evince the DOE's theory of action. After reviewing the results from the pilot, the DOE would make a more evidence-based decision to increase the scale of the project or to abandon it. Another benefit of piloting the evaluation framework and process is to get a sense of the technology infrastructure, staffing, training, and time-value costs involved in taking the project to scale. Finally, a pilot program provides the opportunity to resolve unanticipated implementation issues and to allow the DOE's staff or consultants to prepare documentation that can be used in training subsequent cohorts of adopters.

We believe that 2015-16 is an appropriate target for evaluating a pilot program for a number of practical reasons. Currently, LEAs are in the midst of developing curricula around the newly minted (2009/10) NJ Core Curriculum Content Standards (NJCCCS). These curricula in math, language arts and science are scheduled for implementation in September 2011. Other subjects' curricular revisions follow in 2012-13. Revised curricula should be in place for math, language arts and science for the 2012-13 school year; revisions would be based on feedback from field tests of the new State tests in the spring of 2012. The test results in this transition period would not be comparable: one could not make legitimate comparisons of educator efficacy using VAMs or percentile growth measures until the State tests were fully vetted. Therefore, we recommend the following timeline:

- From 2010-11 through 2013-14, develop and field test valid, reliable, embedded, LEA-based, common assessments (plural) around important curricular specifications and concomitant local evaluation processes and frameworks aligned to professional standards (e.g., NBPTS standards). The coalition would be helpful in facilitating the formation of consortia and directing the DOE to provide necessary fidelity-checking, coordination and support.
- In 2011-12, pilot a beta-version of the educator evaluation framework on a random sample of educators. Collect teacher and student characteristic data. Although the 2012 State test is likely a field test only, selected clusters might be used in a limited VAM analysis in 2013.
- In 2012-13, fully field-test (due notice testing) standardized State assessments based on new NJCCCS and adjust LEA curricula based on results. Collect a second round of educator evaluation framework and student characteristic data. This will generate the first set of value-added data, based on the prior year's data collection, for selected clusters only. This data will be imprecise and unstable and should not be used for any purpose other than vetting the framework controlling for student characteristics. This preliminary look at the data will likely affirm concerns about stability and precision of estimates. We recommend assessing school-level attribution using this data.
- In 2013-14, roll-out vetted standardized State assessments beginning with lowest grade levels so that students are only tested on material that has been taught (*Debra P. v. Turlington* 654 F2d 1079). Alternatively, one could administer tests in Grades 4, 8 and 11 as long as the only high stakes test was in Grade 11. Analyze the impact of the student data on the educator evaluation frameworks, by teacher/leader sub-group, using 2012-2014 field-test data. This analysis would be based on 2 years worth of complete VAM data, therefore, the percentage of false-positives would exceed 25 percent. It should be used for internal studies only. At this juncture, the data

may suggest whether the frameworks are resulting in the kinds of improvement expected—both in student outcomes and educator practice. As in 2012-13, we recommend assessing school-level attribution using this data, too.

- In 2014-15, collect and analyze student and teacher data to determine effectiveness of the process and instruments. Note that although we will have three years of data at this juncture, the stability and precision of the teacher-attribution results will likely be lacking, however, school-attribution analyses may bear fruit. An analysis of the pilot would ensue followed by a decision to scale up, modify or shutdown the initiative.

Although we have proposed an extremely aggressive timeline in order to be responsive to the Governor’s Executive Order, we recognize that it will require a staffing up the DOE and active participation by a consortium of LEAs that has never been seen before. In order to meet the expectations contained in the Governor’s order, valid and reliable assessments will need to be developed for every content area at every grade level. In the 30 or so years of testing in this State, that scope of work has never been tried or achieved. In the end, New Jersey students will be the most highly tested students in the State, and it is inconceivable how many days will be lost to secure-State testing.¹³

RECOMMENDATION 4: Provide LEAs with a set of criteria by which they might opt-out of the proposed statewide educator evaluation system based on district performance and locally developed educator improvement systems. One alternative should be a locally negotiated Peer Assistance and Review (PAR) process designed to identify and assist under-performing teachers before making a professional judgment regarding their fitness for continued service.

Most New Jersey districts are doing well by conventional measures. Students are receiving a through and efficient education. The community is engaged. The local board of education is high-functioning. Teachers and school leaders meet and exceed professional expectations. The need for State intervention, therefore, is neither needed nor welcomed. An appropriate response to the Governor’s Executive Order would allow successful¹⁴ districts to be exempted from yet another unnecessary and over-reaching mandate. One size does not fit all.

As an alternative to the Statewide Educator Evaluation System, we recommend piloting a locally negotiated Peer Assistance and Review Program (PAR) modeled on Montgomery County, Maryland’s Professional Growth System (PGS). This alternative to the lengthy evaluation framework development process is likely to have a greater impact on improving professional practice and student outcomes, in shorter order, and for the longer term. We expect that districts that possess highly collaborative cultures would be able to advance this work as early as the 2011-12 school year.

The PAR program would have three components: the PAR Panel, Consulting Teachers (CTs) and PAR Pair. CTs would be “distinguished practitioners” who provide direct instructional support to teachers and collect data through formal observations. They would report monthly on the progress of identified teachers to the PAR Pair, one distinguished teacher and one school leader assigned to oversee the work of CTs. The CT would write a final summative report at the conclusion of the period of support to be provided to the PAR Panel, a county or consortium-sponsored structure comprised of LEAs’ most distinguished educators. The PAR Panel would consist of equal numbers of teachers and school leaders,

¹³ Secure State testing currently requires 19-administrator days for secure test receipt and storage, secure test administration, secure test packaging and shipping...in 2-3 subjects at selected grade levels. It is typical for a school to terminate instruction during testing weeks due to space limitations created by the need for special testing accommodations for students with special needs as well as classroom accommodations for all general education students. In order for all teachers to be measured by VAMs or growth scores, 6-7 additional tests will need to be developed and administered for existing tested grade-levels and 3-6 additional grade-levels will require test development and implementation in all 9 core content areas.

¹⁴ “Successful” is a term in need of definition; however, a number of State tools already exist to measure “success” (e.g., QSAC) and may be a useful point of departure.

recommended jointly by their respective professional associations and LEA superintendents. Based on the data and information gathered through the program, the PAR Panel would make recommendations to the LEA superintendent in March for provisional teachers and in May for tenured teachers regarding contract renewal, recommendation for a second year in PAR, or contract termination.

The PAR process has costs, but the costs are minimal when compared to a data infrastructure, and the benefits are greater. First, the PAR process addresses a real concern: how do we improve professional practice for under-performing educators in short order. Second, the time and money spent training consulting teachers or PAR Panels further enhances professional practice for those served by the process and for those who serve in the process. Third, to the extent that regional institutions of higher education assist in the PAR process—through documentation studies, provision of facilities and staff training—a feedback loop informs pre-service training programs while providing objectivity and credibility to the entire process.

Based on the NBPTS standards, the Teacher PGS would have six clear standards for the evaluation of all classroom-based teachers, including ESL and special education, at all levels, as well as music, art, and physical education at the elementary level. Broadly, the standards would include:

- Standard I: Teachers are committed to students and their learning.
- Standard II: Teachers know the subjects they teach and how to teach those subjects to students.
- Standard III: Teachers are responsible for establishing and managing student learning in a positive learning environment.
- Standard IV: Teachers continually assess student progress, analyze the results, and adapt instruction to improve student achievement.
- Standard V: Teachers are committed to continuous improvement and professional development.
- Standard VI: Teachers exhibit a high degree of professionalism.

Each professional standard is clarified by performance criteria and descriptive examples of observable teaching behaviors in NBPTS documents.¹⁵

A waiver to existing tenure laws would need to be granted to LEAs that participate in this program. This waiver would serve as an incentive to districts that might be skeptical of the new processes' efficacy.¹⁶ The Teacher PGS model also has other advantages building on the considerable prior work of the national education organizations, progressive LEAs and the DOE. It also empowers professional educators for broad-based action by providing an alternative to a state-imposed, one-size-fits-all solution. Furthermore, the PGS model provides employees with a fair yet deliberate process for professional judgment and increases the opportunities for strategic partnerships.

¹⁵ See *Montgomery County Teacher Level Professional Growth Handbook* at http://www.montgomeryschoolsmd.org/departments/development/documents/TeacherPGS_handbook.pdf

¹⁶ For outcome data on the PAR process, see the Harvard Graduate School of Education's PAR website: <http://www.gse.harvard.edu/~ngt/par/resources/outcome.html>. Over 8 years in Montgomery County, 191 tenured teachers of 9,371 teaching staff members (20%) were referred to the PAR process. Of those referred 20% were dismissed after one year, another 15% resigned or retired, and 37% eventually returned to the classroom.

Table of Abbreviations

Educators for the Quality Assessment of Teacher Effectiveness	EQuATE
Task Force on Educator Effectiveness	Task Force
National Assessment of Educational Progress	NAEP
New Jersey Education Association	NJEA
Department of Education	DOE
Local Education Agency	LEA
Pre-kindergarten through Grade 12	PK - 12
Kindergarten through Grade 12	K - 12
National Board of Professional Teaching Standards	NBPTS
Race-to-the-Top	RTTT
Thorough and Efficient	T&E
Value-Added Measure	VAM
Assessment of Knowledge and Skills	ASK
High School Proficiency Assessment	HSPA
National Board of Professional Teaching Standards	NBPTS

Appendix 1

Vitae of Educators for Quality Assessment of Teacher Effectiveness (EQuATE)

Liz Cresci – Principal, Jonathan Dayton High School, Springfield Elizabeth Cresci is in her sixth year at Jonathan Dayton. Throughout her 28 years in education, she has served as English teacher, Supervisor of Humanities, Assistant Principal and Principal. Jonathan Dayton has been recognized by Newsweek as one of the top high schools in the United States. Under her leadership at Highland Park High School, the school was honored with the School Bell Award from NJN in 2005. She holds a bachelor's degree from New York University and master's degrees from New York University and Rutgers University. Ms. Cresci has been an active member of the New Jersey Principals and Supervisors Association and will join its Board of Directors in July 2011.

Gail Davis – New York City Superintendent of Schools (retired) Gail has long experience as a teacher and administrator in an urban school setting. Most recently, Gail has worked as a facilitator for the *New Jersey Network of Superintendents*, a professional learning community of 16 superintendents from around the New Jersey who meet monthly to improve their professional practice.

Michelle Fine – Distinguished Professor of Social Psychology, City University of New York Prior to this appointment, she taught at the University of Pennsylvania for twelve years. Her research focuses on youth in schools, communities and prisons, developed through critical feminist theory and method. For information about her research or the work of the CUNY Graduate Center Public Science Project, link to <http://web.gc.cuny.edu/psychology> or <http://web.gc.cuny.edu/che/start.htm>.

Drew Gitomer - ETS Distinguished Researcher (soon-to-be-Rutgers Professor) Drew's research interests include policy and evaluation issues related to teacher education, licensure, induction and professional development. His studies have focused on enhancing the validity base for *Praxis*[™] teacher licensure examinations and *National Board for Professional Teaching Standards*[®] assessments for the advanced certification of teachers. Drew's research has also focused on the design of assessments, particularly those that support improvement of instruction. He is currently co-directing the Cognitively Based Assessments for Learning Project, CBAL, an effort designed to transform current K-12 assessment practices. He has most recently been involved with the Gates Foundation work on teacher evaluation. Drew earned a Ph.D. in cognitive psychology from the University of Pittsburgh.

Robert Goodman – Director, NJ Center for Teaching and Learning & Science Teacher, Bergen County Technical High School – Teterboro An MIT physics major with a Rutgers doctorate in science education, Bob retired from a twenty year career in the private sector that included serving as President and CEO of Harman-Kardon, JBL Consumer Products and Onkyo International. Bob was named the 2006 NJ State Teacher of the Year and has been a leader in science education.

John Hart – former Chief of Staff, NJ DOE. John worked in the private sector as an executive for Virtua Health System and as a senior researcher for PHMC Corporation for over a decade before serving as Deputy Chief of Staff for the NJ Department of Human Services and Associate Director for the Walter Rand Institute for Public Affairs at Rutgers University. He served as Chief of Staff to Commissioner Lucille Davy. John is a graduate of Rutgers University.

Bernie Josefsberg – Superintendent of Schools, Leonia Bernie has served in some of the nation's highest performing schools and school districts as a teacher and administrator. Bernie did his doctoral dissertation on teacher evaluation. Bernie serves on the Board of the *Garden State Coalition of Schools* and Executive Committee of the *NJ Association of School Administrator*.

Stan Karp – Director, Secondary Reform Project, Education Law Center Stan was a public school teacher in Paterson, NJ for 30 years and the founding lead teacher of the Communications Academy, a small school reform project. He has served on several advisory committees for NJDOE including the Workgroup on K-12 Education Improvement, and the Secondary Advisory Group. He is an editor of the

school reform journal *Rethinking Schools* and a co-editor of several books including *Rethinking Our Classrooms: Teaching for Equity and Justice*.

Ellen Lawrence – Elementary Special Education Teacher, Orchard Hill Elementary School Ellen is a long-time educator and respected member of the OHES community. Ellen has served on a Learning Team at OHES for the past two years and is an Association Vice President.

Earl Kim – Superintendent of Schools, Montgomery Township Following a stint in the Marine Corps, Earl began his career as an “alternate route” teacher at Trenton Central High School. He took time out to complete a master’s degree from Princeton and was one of the first “alternate route” school administrators in the New Jersey. In 2001-02, Earl was named NJ Principal of the Year. Earl serves on the board of the Foundation Academy Charter School in Trenton.

Gordon MacInnes – Fellow, The Century Foundation in New York Gordon has had a distinguished career in public service. Most recently, he was a lecturer at the Woodrow Wilson School at Princeton University. He also served as the Assistant Commissioner of the New Jersey Department of Education and was a member of the New Jersey State Senate and General Assembly. Gordon also directed the New Jersey Network and was the first director of the Fund for New Jersey.

John Quattrocchi – President, Verona Board of Education John is the president of the Verona Board of Education and has served on the board for nine years. John is a Managing Director at Barclays Capital.

Jim Reed – Middle School Teacher, Montgomery Middle School Jim is a math teacher at Montgomery’s Upper Middle School. Prior to serving as a teacher, Jim was the Director of AT&T’s School of Business and Technology as well as Leadership Academy. He was also a Human Resources manager and software developer.

Willa Spicer – former Deputy Commissioner, NJ DOE Willa has worked at every level of education from teacher to school board member in a variety of settings. Her awards and recognitions are too numerous to list. She is a graduate of Wellesley (BA), Harvard (MA) and Rutgers (ABD).

Russ Walsh – Director of Human Resources, Montgomery Township School District Russ is a long-time educator whose first love is helping students to become fluent readers. He has been a middle school teacher, elementary reading specialist, K-12 Director of Language Arts, lecturer at the undergraduate and graduate levels, and is past president of the International Reading Association. His experiences have ranged from urban to suburban settings.

Rich Wertz – Teacher, Verona High School Rich is a teacher and association leader in a suburban school district. Prior to entering the teaching profession through the Alternate Route, Rich was a Managing Director at Goldman Sachs. Rich graduated from Princeton with a degree in Electrical Engineering & Computer Science and did graduate work in mathematics at The Courant Institute.

Appendix 2 Organizational Analysis

A traditional approach to political and organizational analysis would involve identifying the major actors and their interests in this policy proposal and its implementation. This analysis goes beyond that. It considers the roles of some of the largest stakeholders in the New Jersey public education system at the policy level but also dives into the culture of LEAs and more importantly schools and classrooms. This novel idea, that State-level policy initiatives should consider implications in schools and classrooms is an outgrowth of Harvard's *Public Education Leadership Project* (PELP) framework.

State Board of Education (State BOE) – The State BOE plays a critical role in the implementation of the Task Force's recommendations. Any recommendations that require administrative code changes (e.g., educator evaluation guidelines and seniority regulations) require State BOE approval. The State BOE would have to agree with the substance of the Commissioner's recommendations, his process for achieving policy goals and the scope of authority that the State BOE exercises in LEA affairs. In light of the questionable benefit and significant cost of the growth model or value-added proposal both to the department and to LEAs, the State BOE might require a detailed implementation plan: costs to LEAs as well as the State, an evaluation plan for the pilot program, and a decision point to scale up the project after successful completion of the pilot. As mentioned elsewhere, given the significant start-up costs and difficulty fulfilling constitutional levels of school funding, the project should only move forward as grant-funding (e.g., RTTT) allows. In addition, the State BOE should be cautious about over-reaching its constitutional role and usurping the authority of LEA Boards of Education (Local BOEs) in the instructional improvement process (viz., evaluation of staff). Finally, the State BOE should heavily consider the capacity of the DOE to effect, manage and drive such widespread change without stakeholder involvement. Without LEA partnerships, the initiative will likely not succeed and at considerable opportunity and direct cost.

State Legislature – The legislature plays another critical policymaking role to the extent that the EQuATE initiative requires statutory reform (viz., granting tenure law waivers for districts participating in PAR process). While the legislature may meet significant opposition from unions in any initiative to remove tenure protections, the PAR proposal neutralizes that opposition by requiring that the process be collectively bargained and by making waivers contingent on having a mutually agreed upon alternative to due process. Unions should view the PAR alternative as an opportunity to elevate the profession, like other professions, by institutionalizing standards of practice and "policing one's own".

Another way in which the legislature plays a role is in determining funding priorities and appropriating funds for certain initiatives (e.g., merit pay, VAM infrastructure...). Prior to authorization of long-term and large scale spending plans for data systems and staff, the legislature has an interest in seeing detailed implementation plans with established appropriations control points (critical path milestones) in the process. For example, before releasing funds for a pilot project, the appropriate legislative committee might require a report on the research-backed outcomes that such an expensive program would deliver and in what timeframe. If the theoretical benefits do not justify the cost, then the decision is apparent.

Local Boards of Education – Local BOEs are empowered to set policy and regulations surrounding the evaluation of staff according to N.J.A.C. 6A:32. Over time, however, code has stipulated more and more of what must be included in evaluations of staff members from non-tenured teachers to superintendents. This encroachment on the authority of local BOEs has been part of a longer trend that was noted as early as 1970 in a *Seton Hall Law Review* article, "Schools and School Districts – The Subtle Move Toward Total State Control."¹⁷ In the "shared power" structure created by the NJ constitution,¹⁸ Local BOEs

¹⁷ "Schools and School Districts – The Subtle Move toward Total State Control," 2 *Seton Hall Law Review* 174 (1970).

retain the authority to decide the process, standards and instruments that are used in teacher evaluation. As the democratically elected education policymakers for their municipality, and the most responsive elected officials to the educational demands of their school communities, decisions that affect the kinds of teachers who are recruited, employed and retained ought to be subject to the rules crafted by this institution.

Political Powerbrokers – Powerbrokers stand to gain influence under the reforms introduced by the Governor’s Task Force. To the extent that tenure protections are relaxed and the role of local BOEs diminished, political leaders in particular will have greater influence over hiring and firing decisions within schools. In part, it was this kind of influence that led to boards, ethics laws, tenure and seniority protections. As recently as the 1980s, the role of powerbrokers in influencing firing and retention decisions was cataloged in the policy journal, *New Jersey Reporter*. For the good of all stakeholders—including political leaders—we propose a system of evaluation that ensures adherence to explicit professional standards.

Unions – The New Jersey Education Association (NJEA) and New Jersey Federation of Teachers (AFT) have significant influence on the alternatives that are considered by legislators. The NJEA has openly opposed the use of VAMs in educator evaluations for all of the reasons noted in this paper. Although it has not endorsed any specific PAR proposal, a National Education Association affiliate is the originator of the PAR model used in this paper. The AFT has a longer history of collaborative labor-management practices, PAR panels being among them. Its training centers serve all stakeholders in preparing them for collaborative reform efforts. We believe that given the reliance of PAR on collective bargaining authorization, the history of PAR models in both unions and the tradition of collaboration between local affiliates and Local BOEs, PAR panels would be an improvement on the system of tenure that currently exists.

Families – Families play a role in improving our education system and our professional practice as educators. The inexpert use of student achievement data will likely lead to mistakenly demonizing students and families as well as teachers. Low test scores, even today, are frequently used to reinforce public, educator or self misperceptions of the value and capacity of our youth. The untested capacity and abilities of many of our urban, suburban and rural young people eludes the psychometrician’s and test-focused educator’s eye. Families have been drawn into this policy debate as “external factors”, when they really should be an integral part of our improvement process.

Evaluation Systems and Organizations – The culture of an organization is a key element to its success or failure. An evaluation system is critical component of that culture. If we examine successful organizations closely, we find that their cultures may differ, as may the evaluation systems that promote and sustain those cultures.

Evaluation systems serve to affirm, on a periodic basis, the successful operation of other mechanisms which promote the mission and principles of the organization on a day-to-day basis. The culture of one high-performing school might not be successfully employed at another school due to environmental differences. We need to guard against a one-size-fits-all mentality in striving toward building more successful schools. We need to give schools the discretion to create a culture and an inherent evaluation system that works best in the community it serves.

Autonomy of Schools and Districts – We need to consider that many of our most successful schools exercise a high degree of autonomy. This autonomy includes discretion in recruiting, retaining and promoting staff in a way that best serves the children of its community. This autonomy extends to

¹⁸ Implied powers derive from Constitution of the State of New Jersey, Article 4, Section 7, Paragraph 11

assigning teachers to classes so as to maximize the value of instruction. While the attempts to “*better equip school districts* to improve their performance” and “*inform decisions* about various school policies including professional development, promotion, compensation, merit-based bonuses, tenure, and reductions-in-force and separations” are laudable goals, any concomitant prescription of mandated personnel action—reassignment, termination, promotion or merit pay—will undermine this autonomy of local schools and local school districts and the benefits which flow from that autonomy.

Effects of Incentives on Collaboration – The competitive effect of the introduction of educator-effect metrics would likely undermine collaboration between teachers and disrupt vertical integration of instruction. There is a public misperception that schools excel because of individual, “star” educators. This misperception discounts the value of vertical integration of instruction from grade to grade, the tenure of individual teachers in their teaching assignments, and the judgment of administrators in allocating teachers to teaching assignments so as to maximize achievement. It discounts the notion of teamwork, which is a key element of any successful school.

Continuity of Instruction and the “T-Word” – In our experiences as teachers, we have found considerable support for the research findings around “continuity of instruction”. Some of the biggest prior knowledge gaps we have found in our students have been a result of the “that was the year we had 3 math teachers” phenomenon. Having students move from one grade to another and receive a consistent learning experience depends on administrators “getting the right athletes in the right positions” and giving their teachers the time to make the “curricular capital” investment necessary to master their teaching assignment.

Though the word “tenure” has lost its original meaning and has taken on a meaning in the political vocabulary of “legal protection for underperforming educators”, “tenure” in an individual teaching assignment has a large impact on the quality of instruction. We’ll call this “tenure-in-prep”.

Inherent in the drive to use teacher-effect metrics to identify underperforming teachers, is the assumption that there is an extremely qualified replacement waiting in the wings, with the content knowledge and tenure-in-prep necessary to right the ship. This is a poor assumption. It is useful to consider the extreme circumstance of a new, different teacher being assigned to a teaching assignment each year because of a school’s inability to reach student achievement benchmarks in that area of instruction.

Prescribing evaluation consequences for a teacher or in a formulaic way risks replacing “continuity of instruction” with “volatility of instruction” and diminishes the future benefits of tenure-in-prep investments made by teachers and their administrators in prior years.

Incentives, Disincentives and Educator Behavior – The mention of “promotion, compensation, merit-based bonuses” in the Executive Order warrants a few questions:

- Would these be financially *material*? If I am faced with an uncertain incentive of \$2,500 to improve math scores and a certain incentive of \$5,000 for coaching JV basketball or \$5,000 for working one shift per week at The Home Depot, which do I chose?
- Would these incentives be *perpetual* (unlikely)?
- Will the structure and *balance* of incentives and disincentives fundamentally change the way young people value a career in education. The sanctions that may potentially be proposed for underperformance (i.e., non-renewal, tenure expiration, RIF and separations) have a perpetual financial impact. Coupling non-perpetual incentives and perpetual disincentives will significantly devalue the expected lifetime earnings of an educator and reduce the quality of young people entering the profession.

- Would incentives recognized *sustained improvement* or be based on 1-year achievement? As the financial industry struggles to tear itself away from 1-year compensation incentives, is education marching toward them? Would educators “game” the system by “sandbagging” certain years, so as to exhibit outperformance in the next year?
- Would incentives truly change educator behavior?
 - For master teachers, merit pay is “economic rent”. Before merit pay, they did their job with aplomb for \$X. After merit pay, doing nothing differently, they reap an economic rent \$R on top of their optimal wage \$X. In economic terms, the system lost efficiency when this rent was paid.
 - Will outperforming educators perform even better in the presence of incentives?
 - Will mediocre educators outperform in the presence of incentives?
 - Will underperforming educators become mediocre in the presence of incentives?
 - Will the presence of steeper disincentives discourage educators from taking “classroom risk” (i.e., employing innovative curricula, technology, teaching, and assessment tools)?
 - Would an administrator hire a teacher who only outperformed in the presence of incentives? It almost sounds like a teacher holding test scores captive for ransom.

Our society’s social contract with teachers has been, “You will never get rich, but you will never be poor.” When a young person enters the teaching profession, he cedes significant financial upside but also gains significant protection against downside. In options-pricing parlance, he trades a series of long-dated call options on his salary for a series of long-dated put options on his salary.

This protection, along with the intrinsic motivation to help young people, sustains teaching as a profession. Young people do not enter the teaching profession to take speculative financial risk on their own performance. We do not want teachers who are motivated by such incentives but who maximize their utility on altruism over profit. Finally, we do not want the world of education to be distorted by the grotesque behaviors of many in the financial sector. Greed is not good when it comes to nurturing of young people.

Deleterious Effects of a Precise, but Inaccurate Framework – The Executive Order speaks of a more *precise* framework. We should be shooting higher in establishing a framework which is precise (statistically reliable) and *accurate* (educationally valid). A precise, but inaccurate framework may damage our schools more than it improves them.

Appendix 3 Statistical Analysis

Statisticians Agree that VAM and Growth Models are not Ready for Policy Implementation – Experts agree that using VAMs in the manner suggested in Executive Order No. 42 would be unsupportable due to a wealth of statistical issues. The National Research Council and National Academy of Science write that “VAM estimates of teacher effectiveness should not be used to make operational decisions because such estimates are too unstable to be considered fair or reliable.”¹⁹ Furthermore, the RAND Corporation writes, “The research base is currently insufficient to support the use of VAM for high-stakes decisions about individual teachers or schools.”²⁰

Henry Braun²¹ further elaborates on why the seemingly straightforward approach suffers from potential limitations in implementation. He does welcome the possibility of including a quantitative component to teacher evaluations but cautions that “VAMs should *not* be used as the sole or principal basis for making consequential decisions about teachers (concerning salaries, promotions and sanctions, for example).” Specifically, he notes the following shortcomings of VAMs:

- Classroom placement and teacher assignment is not random. Students will not have the same inherent characteristics (i.e. parental support, motivation, study habits, interpersonal dynamics and other relevant characteristics). Furthermore, students are typically assigned to teachers in a non-random way. Therefore, because of this nonrandom assignment, it is difficult to statistically or anecdotally disentangle intrinsic student characteristics that drive performance from teacher effectiveness. Teachers who have students with higher potential for achievement may obtain a high score regardless of their effectiveness.
- School and district level differences influence chance of success. Student progress can also be influenced by the physical condition of the school and the resources available, as well as school policies and school level implementation of district policies. All of these features are out of control of the teacher and may result higher performance scores for teachers in schools or districts with more resources.

Both of these issues makes is difficult to interpret teacher performance as the sole cause of a statistical increase in students’ performance. Other potential drawbacks to VAMs outlined by Braun include: the nature of the test scores that serve as the raw material for VAMs; the amount of information available to estimate each teacher’s effectiveness; and the treatment of missing data.

Principals are Already Able to Identify Effective Teachers – Principals and other heads of schools may be in a better position to identify performance, given their familiarity with constraints on their individual school and on particular teachers. Jacob and Lefgren²² find that principals are able to identify the best and worst teachers (top and bottom 10-20%) when it comes to producing student achievement gains but are not as good at discerning among the broad middle. This compares favorably with the ability of VAMs to identify only the top and bottom 15% with confidence. VAMs address the wrong problem: the problem is not identification of less than effective teachers or teaching but what to do with the matter once it is revealed. The main benefit of VAMs is that they can be used to affirm a principal’s or PAR

¹⁹ Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education, National Academy of Sciences, “Letter Report to the U.S. Department of Fund.” October 5, 2009, p.10. http://books.nap.edu/openbook.php?record_id=12780&page=1.

²⁰ MacCaffrey, D.F., *et al.* (2003). “Evaluating Value-Added Models for Teacher Accountability.” Santa Monica: RAND Corporation.

²¹ Braun, Henry I. (2005). “Using Student Progress To Evaluate Teachers: A Primer on Value-Added Models.” Princeton: Educational Testing Service.

²² Jacob, Brian and Lefgren, Lars (2008). “Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education.” *Journal of Labor Economics* 26:101-136.

Panel's judgment. VAMs, to the extent that they are used, should not drive or supplant informed judgments of teacher effectiveness.

Granularity of Attribution should be the School not the Individual – While the evaluation of all individual educators and all educational entities has merit, we suggest that investment in evaluation *at the school level, rather than the level of individual educators*, has greater value to the general public. School-level performance statistics are more stable and more easily controlled for differences between school environments. Educator-effect measures, as they are based on smaller datasets, are less stable and more difficult to statistically isolate in the context of other random and non-random effects on student performance. Even sub-group performance, with sufficient n-size, can be statistically stable and worthy of public recognition. New Jersey already has the tools in place to make these kinds of distinctions and allocations.

Do the stockholders or board of directors of a public corporation interfere with evaluation or compensation policies at any level below the most senior officers of the corporation? Such an action would be viewed as extreme micromanagement. Would potential homebuyer make a decision on buying a home in a community based on an outstanding evaluation or unsatisfactory evaluation of an individual teacher? It's unlikely. When a child graduates from high school, his education is product of approximately 100 individual teachers and about 10 administrators. School effects are more stable, more practicable, and more consonant with the manner in which the public considers its schools.

Defining “Improved Educator Performance” – The definition of “improved teacher and leader performance” relies on statistical controls. As mentioned in the policy brief, external factors must be controlled in order to properly attribute teacher effects. How would a VAM measure control for:

- Summer Learning/Achievement
- Student Attendance
- Specific Learning Disabilities
- School Quality including Peer Effects
- Non-Random Student Placement – stuffing a district, school or class with high or low performers
- Demographic Factors (e.g., maternal education, health, wealth/poverty...)

How would “improved teacher and leader performance” be defined?

- Would statistical significance be based on “partialling out” “student factors” to define “improvement”? If not, are we promoting a system where the largest component of educator evaluation is chance?
- Would a VAM be used to recognize only one-year changes on prior performance or sustained performance above expected performance?
- Would educators at a local level retain discretion in ascribing meaning to changes in student performance metrics or differentials of student performance metrics versus peer groups?

Appendix 4 - Comparison of Selected Teacher Evaluation Systems

The attached Teacher Evaluation Systems Comparison Matrix looks at the evaluation designs in five states and the District of Columbia. Each of these designs has been looked to as models as other states seek to develop teacher evaluation systems. In completing the matrix, EQuATE looked at several aspects of teacher evaluation systems that research indicates are central to a program that improves student learning.

- Inclusive process that involves all stakeholders
- Empowerment of teachers and school leaders
- A broad perspective on the domains of teaching including student achievement, instructional competence, content knowledge, classroom environment and professionalism
- A thoughtful process that is designed to yield useful information for teacher professional development and employment decisions
- Costs in relation to the benefit realized by the process

Strengths and weaknesses were identified using these criteria. It is hoped that the matrix can help the Task Force borrow from the strengths of some of the systems and avoid some of the pitfalls of these designs.

State/District	Domains of Teaching Addressed	Evaluation Process	Strengths	Weaknesses
Washington, DC	<p>Student Achievement</p> <p>Teacher Instructional Expertise</p> <p>Commitment to School Community</p> <p>School Value Added</p> <p>Core Professionalism (attendance, team player)</p>	<p>A teacher Impact Score is determined using the following:</p> <p>Student achievement is determined using a VAM statistical model (Weighted 50%)</p> <p>Instructional expertise determined through administrator and teaching expert observation (Weighted 35%)</p> <p>Commitment to School Community determined by score on rubric (Weighted 10%)</p> <p>School Value Added is a statistical determination of how the school scored overall in student achievement (Weighted 5%)</p> <p>Core Professionalism determined through score on rubric (Points may be added or subtracted from Impact Score)</p> <p>VAM not applicable to 80% of teachers in non-tested disciplines. For these teachers, 75% is based on observations by principals and “master teachers” evaluation is based</p>	<p>Captures all domains of professional practice that influence student outcomes</p> <p>Well-developed rubrics for various components</p> <p>Requires careful supervision of staff</p> <p>Framework phased-in over time.</p>	<p>Does not capture all important student outcomes</p> <p>Does not ensure validity and reliability of measures of professional practice; VAM weighted too heavily</p> <p>In EQuATE’s view, cost of technology infrastructure and associated technology and data support staff does not justify the benefit</p> <p>Requires an additional layer of supervision (Instructional Experts)</p> <p>Process did not include teachers in design of evaluation</p> <p>Framework does not empower teachers and school leaders. Only 2% of teacher evaluation is dedicated to collaboration with colleagues</p> <p>Framework does not evaluate all staff</p> <p>Links teacher performance to school performance instead of team performance</p>
Tennessee Value Added Assessment System (TVAAS)	<p>Student Achievement</p>	<p>Student Achievement is a determined using a VAM model based on student scores on a statewide achievement test in grades 3-8 and end of course tests in grades 9-12</p>	<p>Provides a snapshot of teacher impact on student achievement</p>	<p>Ignores all but one domain of teaching and learning</p> <p>Does not ensure validity and reliability of measures of professional practice; VAM weighted too heavily</p> <p>In EQuATE’s view, cost of technology</p>

State/District	Domains of Teaching Addressed	Evaluation Process	Strengths	Weaknesses
				<p>infrastructure and associated technology and data support staff does not justify the benefit</p> <p>Is limited to the evaluation of teachers in tested areas only</p> <p>Process did not include teachers in design of evaluation</p> <p>Framework does not empower teachers and school leaders</p>
Delaware	<p>Student Achievement</p> <p>Classroom Environment</p> <p>Instruction</p> <p>Professional Responsibilities</p> <p>Planning and Preparation</p>	<p>Student Achievement is identified through standardized test measures, including the State assessment, MAP and DIBELS.</p> <p>Student Achievement is also measured through common assessments designed by groups of teachers</p> <p>Other domains are assessed using the observation process guided by Danielson's domains.</p>	<p>Uses multiple test measures to determine student achievement</p> <p>Encourages collaboration through the use of common assessments as a student achievement measure</p>	<p>While giving a nod to the complexity of skilled instruction, seems focused on student achievement as the driver of making judgments about teacher effectiveness.</p> <p>How much the other domains count is not clearly stated.</p> <p>Does not ensure validity and reliability of measures of professional practice.</p>
California (PACT)	<p>Student Achievement</p> <p>Instruction</p> <p>Planning and Preparation</p> <p>Classroom Environment</p>	<p>Student Assessment accomplished through a variety of standardized tests, benchmark assessments and student artifacts</p> <p>Other domains assessed through supervisor observation, expert evaluations, teacher self-assessment, teacher portfolios, videos</p>	<p>Aims at providing information on the complexities of teaching</p> <p>Potential to provide rich information about what is happening in the classroom</p> <p>Provides both formative and summative information to teachers</p>	<p>Costly</p> <p>Difficult to insure validity and reliability across school districts</p> <p>Does not address weighting of various components of teaching</p>
Montgomery County, MD	Standards based on National Board Standards	Teachers are assessed through formal evaluation.	Developed collaboratively by the school district with teacher and administrator associations	Heavily weighted on formal observations and evaluations
Peer Assistance and Review (PAR)	<p>Commitment to students</p> <p>Content Knowledge</p> <p>Learning Environment</p>	<p>Those judged seriously at risk are referred to PAR panel</p> <p>Consulting Teachers (CTs) are assigned to work with these teachers</p>	<p>Based on nationally recognized standards of teacher practice</p> <p>Multi-faceted</p>	<p>Makes limited use of teacher teams, instructional rounds, lesson study</p>

State/District	Domains of Teaching Addressed	Evaluation Process	Strengths	Weaknesses
	<p>Assessment of Student Progress</p> <p>Commitment to continued professional growth</p> <p>High Degree of Professionalism</p>	<p>One teacher and one principal from the PAR Panel oversee the work of the CTs.</p> <p>CT prepares final summative report</p> <p>PAR Panel makes recommendations to the superintendent regarding contract renewal, additional year of PAR, or contract termination.</p>	<p>Focused on professional development of teachers</p> <p>Job –embedded professional learning</p>	
Harrison, Colorado	<p>Student Achievement</p> <p>Professional Performance including</p> <p>Comprehensive approach to instruction</p> <p>Use of student assessment data to guide instruction</p> <p>Quality instruction</p> <p>Culture conducive to student well-being</p> <p>Personal professional growth</p> <p>Meets responsibilities to students, parents, public</p>	<p>E&R is a pay for performance plan</p> <p>Combined Achievement and Performance Scores are used to determine placement on salary guide</p> <p>Teachers are graded as novice, progressing, proficient, exemplary and master based on the combined measure and paid accordingly</p>	<p>Rubrics have been developed for effective assessment of professional performance</p> <p>Uses a variety of observation constructs including walk-throughs, formal observations using a clinical model, etc to evaluate professional performance</p> <p>Administrators spend a great deal of time observing instruction which raises validity and reliability of judgments</p> <p>PLCs are in place to use student achievement data to improve instruction.</p> <p>Highly developed and systematic program of incentives</p>	<p>Assumes that monetary incentives will Improve teacher performance</p> <p>Like other financial incentive plans, may foment an unhealthy competitive rather than cooperative culture</p> <p>Unclear about the inclusion of teachers in the design</p>

State/District	Domains of Teaching Addressed	Evaluation Process	Strengths	Weaknesses
Washington, DC	<p>Student Achievement</p> <p>Teacher Instructional Expertise</p> <p>Commitment to School Community</p> <p>School Value Added</p> <p>Core Professionalism (attendance, team player)</p>	<p>A teacher Impact Score is determined using the following:</p> <p>Student achievement is determined using a VAM statistical model (Weighted 50%)</p> <p>Instructional expertise determined through administrator and teaching expert observation (Weighted 35%)</p> <p>Commitment to School Community determined by score on rubric (Weighted 10%)</p> <p>School Value Added is a statistical determination of how the school scored overall in student achievement (Weighted 5%)</p> <p>Core Professionalism determined through score on rubric (Points may be added or subtracted from Impact Score)</p> <p>VAM not applicable to 80% of teachers in non-tested disciplines. For these teachers, 75% is based on observations by principals and “master teachers” evaluation is based</p>	<p>Captures all domains of professional practice that influence student outcomes</p> <p>Well-developed rubrics for various components</p> <p>Requires careful supervision of staff</p> <p>Framework phased-in over time.</p>	<p>Does not capture all important student outcomes</p> <p>Does not ensure validity and reliability of measures of professional practice; VAM weighted too heavily</p> <p>In EQuATE’s view, cost of technology infrastructure and associated technology and data support staff does not justify the benefit</p> <p>Requires an additional layer of supervision (Instructional Experts)</p> <p>Process did not include teachers in design of evaluation</p> <p>Framework does not empower teachers and school leaders. Only 2% of teacher evaluation is dedicated to collaboration with colleagues</p> <p>Framework does not evaluate all staff</p> <p>Links teacher performance to school performance instead of team performance</p>
Tennessee Value Added Assessment System (TVAAS)	<p>Student Achievement</p>	<p>Student Achievement is a determined using a VAM model based on student scores on a statewide achievement test in grades 3-8 and end of course tests in grades 9-12</p>	<p>Provides a snapshot of teacher impact on student achievement</p>	<p>Ignores all but one domain of teaching and learning</p> <p>Does not ensure validity and reliability of measures of professional practice; VAM weighted too heavily</p> <p>In EQuATE’s view, cost of technology infrastructure and associated technology and data support staff does not justify the benefit</p> <p>Is limited to the evaluation of teachers in tested areas only</p>

State/District	Domains of Teaching Addressed	Evaluation Process	Strengths	Weaknesses
				<p>Process did not include teachers in design of evaluation</p> <p>Framework does not empower teachers and school leaders</p>
Delaware	<p>Student Achievement</p> <p>Classroom Environment</p> <p>Instruction</p> <p>Professional Responsibilities</p> <p>Planning and Preparation</p>	<p>Student Achievement is identified through standardized test measures, including the State assessment, MAP and DIBELS.</p> <p>Student Achievement is also measured through common assessments designed by groups of teachers</p> <p>Other domains are assessed using the observation process guided by Danielson's domains.</p>	<p>Uses multiple test measures to determine student achievement</p> <p>Encourages collaboration through the use of common assessments as a student achievement measure</p>	<p>While giving a nod to the complexity of skilled instruction, seems focused on student achievement as the driver of making judgments about teacher effectiveness.</p> <p>How much the other domains count is not clearly stated.</p> <p>Does not ensure validity and reliability of measures of professional practice.</p>
California (PACT)	<p>Student Achievement</p> <p>Instruction</p> <p>Planning and Preparation</p> <p>Classroom Environment</p>	<p>Student Assessment accomplished through a variety of standardized tests, benchmark assessments and student artifacts</p> <p>Other domains assessed through supervisor observation, expert evaluations, teacher self-assessment, teacher portfolios, videos</p>	<p>Aims at providing information on the complexities of teaching</p> <p>Potential to provide rich information about what is happening in the classroom</p> <p>Provides both formative and summative information to teachers</p>	<p>Costly</p> <p>Difficult to insure validity and reliability across school districts</p> <p>Does not address weighting of various components of teaching</p>
Montgomery County, MD Peer Assistance and Review (PAR)	<p>Standards based on National Board Standards</p> <p>Commitment to students</p> <p>Content Knowledge</p> <p>Learning Environment</p> <p>Assessment of Student Progress</p> <p>Commitment to continued professional growth</p>	<p>Teachers are assessed through formal evaluation.</p> <p>Those judged seriously at risk are referred to PAR panel</p> <p>Consulting Teachers (CTs) are assigned to work with these teachers</p> <p>One teacher and one principal from the PAR Panel oversee the work of the CTs.</p> <p>CT prepares final summative report</p>	<p>Developed collaboratively by the school district with teacher and administrator associations</p> <p>Based on nationally recognized standards of teacher practice</p> <p>Multi-faceted</p> <p>Focused on professional development of teachers</p> <p>Job –embedded professional learning</p>	<p>Heavily weighted on formal observations and evaluations</p> <p>Makes limited use of teacher teams, instructional rounds, lesson study</p>

State/District	Domains of Teaching Addressed	Evaluation Process	Strengths	Weaknesses
	High Degree of Professionalism	PAR Panel makes recommendations to the superintendent regarding contract renewal, additional year of PAR, or contract termination.		
Harrison, Colorado The Effectiveness and Results Plan (E & R)	<p>Student Achievement</p> <p>Professional Performance including</p> <p>Comprehensive approach to instruction</p> <p>Use of student assessment data to guide instruction</p> <p>Quality instruction</p> <p>Culture conducive to student well-being</p> <p>Personal professional growth</p> <p>Meets responsibilities to students, parents, public</p>	<p>E&R is a pay for performance plan</p> <p>Combined Achievement and Performance Scores are used to determine placement on salary guide</p> <p>Teachers are graded as novice, progressing, proficient, exemplary and master based on the combined measure and paid accordingly</p>	<p>Rubrics have been developed for effective assessment of professional performance</p> <p>Uses a variety of observation constructs including walk-throughs, formal observations using a clinical model, etc to evaluate professional performance</p> <p>Administrators spend a great deal of time observing instruction which raises validity and reliability of judgments</p> <p>PLCs are in place to use student achievement data to improve instruction.</p> <p>Highly developed and systematic program of incentives</p>	<p>Assumes that monetary incentives will Improve teacher performance</p> <p>Like other financial incentive plans, may foment an unhealthy competitive rather than cooperative culture</p> <p>Unclear about the inclusion of teachers in the design</p>