Building a Breakthrough Framework
for Educator Evaluation in the Commonwealth

Submitted by the Massachusetts Task Force on the Evaluation of Teachers and Administrators
March, 2011

MA Task Force on the Evaluation of Teachers and Administrators

Members
Patty Barrett, Principal, Andover Public Schools
Anna Bradfield, Dean, Bridgewater State University
Henry Braun, Professor, Boston College
MaryAnn Byrnes, President, MA Council for Exceptional Children
Mary Czajkowski, Superintendent, Agawam Public Schools
John D'Auria, President, Working Group for Educator Excellence
Christine Evans, Past President, MA School Counselor Association
Lisa Famularo, Research Director, Rennie Center for Education Research and Policy
Michael Flynn, Teacher, Southampton Public Schools, former MA Teacher of the Year
Tom Fortmann, Retired Engineer/Executive, former ESE Board Member
Robert Fraser, MA Association of School Personnel Administrators
Jon Fullerton, Center for Teacher Effectiveness, Harvard University
Tom Gosnell, President, American Federation of Teachers MA
Amanda Green, MA Association of Special Education Parent Advisory Councils
at the Federation for Children with Special Needs
Orin Gutlerner, Founding Director, MATCH Charter Public High School
Linda Hayes, Assistant Director, MA Secondary School Administrators' Association
Nadya Higgins, Executive Director, MA Elementary School Principals Assoc.
Caitlin Hollister, Teacher, Boston Public Schools
Elsie Huang, Principal, Boston Preparatory Charter School
Pamela Hunter, Principal, Southwick-Tolland Regional High School
Neelia Jackson, Teacher, MA Mathematics Association of Teacher Educators
Carla Jentz, Executive Director, MA Administrators for Special Education
Glenn Koocher, Executive Director, MA Association of School Committees
Jim Lynch, MA Association of Vocational Administrators
Joam Marmolejos, Student, Chelsea High School
Seth Moeller, Director, Talent Management, Fidelity Investments
Constance Moore, Vice President, MA Art Education Association
Linda Noonan, Executive Director, MA Business Alliance for Education
Floris Wilma Ortiz-Marrero, Teacher, Amherst Public Schools, MA Teacher of the Year, 2011
Elizabeth Pauley, Senior Program Officer, The Boston Foundation
Steve Rivkin, Professor, Amherst College, School Committee Member, Amherst Public Schools
Beth Schiavino-Narvaez, Chief Academic Officer, Springfield Public Schools
Tom Scott, Executive Director, MA Association of School Superintendents
Norm Shacochis, Vice-President, MA Council for the Social Studies
Jesse Solomon, Director, Boston Teacher Residency (Resigned February 13, 2011)
Paula Squires, VP for Human Resources, Baystate Health
Mary Ann Stewart, President, Massachusetts State PTA
Paul Toner, President, MA Teachers Association
Shakera Walker, Teacher, Boston Public Schools
Martin West, Professor, Harvard University

Alternates
Joseph Casey, Superintendent, Melrose Public Schools, MA Association of School Superintendents
Phil Flaherty, Assistant Director, MA Secondary School Administrators Association
Dan Murphy, Director of Education Policy and Programs, American Federation of Teachers MA
Kathleen Robey, Board Member, MA Association of School Committees
Kathie Skinner, Director, Center for Policy and Practice, MA Teachers Association

Table of Contents

We will know that our work is complete when every student can say with confidence:
"I am challenged and engaged in school, and I see how what I'm learning connects with the real world. I know what I'm good at, I know what I need to work on, and I know where to go for support. I am on track to go to college, get a job that I'm great at, and keep learning."

And every teacher can say:
"I know how to reach, motivate, support, and engage every student in my classroom. I receive honest, useful feedback from my peers and principal, recognition when I succeed, and support when I do not. All of my students have the ability to go college, and I know that it's my job to prepare them so they have that choice."

Massachusetts *Race to the Top* application

"Effective administrators create a climate where every teacher is going to thrive. The main focus is on student learning: that is a given. But the learning of students occurs in direct proportion to the high expectations and supportiveness of the professional culture of the school. Inquiry, intellectual risk taking, and mistakes are expected, valued, and recycled into learning. The job of the administrator and leader is to create a climate that fosters serious, ongoing adult and student learning. This is the standard against which we should be evaluating all leaders."

Task Force Member
and former Administrator

"Current evaluation practices in the state are wobbly, at best. We are often stuck in place, unable to move beyond simple compliance with procedures. The Task Force and the Board of Education have a chance to break this logjam. We can create a more ambitious, focused and growth-oriented framework. I am hoping for a breakthrough."

Task Force Member, former Teacher and Principal

Executive Summary

The Massachusetts Task Force on the Evaluation of Teachers and Administrators is pleased to present its recommendations to the Board of Elementary and Secondary Education and Commissioner Mitchell Chester.

The Challenge

National and statewide evidence is clear – educator evaluation does not currently serve students, educators or society well. In its present state, educator evaluation in Massachusetts is not achieving its purposes of promoting student learning and growth, providing educators with adequate feedback for improvement, professional growth, and leadership, and ensuring educator effectiveness and overall system accountability.

The Task Force concludes that current educator evaluation practice in Massachusetts:
- Rarely includes student outcomes as a factor in evaluation
- Often fails to differentiate meaningfully between levels of educator effectiveness
- Fails to identify variation in effectiveness within schools and districts
- Rarely singles out excellence among educators
- Does not address issues of capacity, or "do-ability"
- Fails to calibrate ratings, allowing inconsistent practices across the state
- Fails to ensure educator input or continuous improvement
- Is often under-resourced or not taken seriously

Simply put, poor evaluation practices are a missed opportunity for promoting better leading, better teaching, better learning, and better schools.

The Opportunity

Despite these problems, the Commonwealth is poised for change, and it is the judgment of the Task Force that a breakthrough is both needed and achievable. By developing the proposed Framework and applying Race to the Top resources to the challenge, Massachusetts can transform educator evaluation from an inconsistently applied compliance mechanism into a statewide catalyst for educator development and continuous professional growth. The framework that the Task Force proposes is intended to support, develop and retain the highly effective educators our children need to learn, grow and achieve.

Evaluation Framework: Key Design Features

The use of multiple measures of student learning, growth and achievement as a significant factor in all educator evaluations is a core feature of the framework. In addition to this core recommendation, the Task Force proposes that a new evaluation framework include the following key design features:

4 Standards with Indicators for all Educators[1]

For Administrators	For Teachers
Curriculum, Instruction and Assessment	Curriculum, Instruction and Assessment
Management and Operations	Teaching All Students
Family and Community Partnerships	Family and Community Engagement
Professional Culture	Professional Culture

3 Categories of Evidence

Three categories of evidence will be used in every district's educator evaluation system:

- *Multiple measures of student learning, growth and achievement*, including
  - Progress toward learning targets
  - MCAS growth measures in comparison to comparable schools, based on appropriate school-level demographics, where applicable, and
  - Measures of learning comparable across grade or subject district-wide
- *Judgments based on observation and artifacts of professional practice*, using a DESE-approved observation system
- *Collection of additional evidence relevant to one or more Standards*, documenting fulfillment of other areas of professional responsibilities and growth as well as contributions to the school community and the professional culture.

4 Performance Ratings that apply to all educators, across the state

- Exemplary: Practice is consistently, significantly above proficiency on the Standard or overall
- Proficient: Practice demonstrates skilled performance on the Standard or overall
- Needs Improvement: Practice demonstrates lack of proficiency on the Standard or overall
- Unsatisfactory: Practice demonstrates lack of competence on the Standard or overall

5-Step Evaluation Cycle

Self-Reflection and Self-Assessment. Two core principles emerged from the deliberations of the Task Force: that educators a) engage in on-going improvement of their own professional practice, and b) take responsibility for their students' learning, growth and achievement. The evaluation process begins with educators reflecting on and assessing their professional practice, and analyzing the learning, growth, and achievement of their students.

Goal Setting and Development of a Plan. Each educator meets with his or her evaluator to: a) review self-reflections and self-assessments, b) jointly analyze students' learning, growth and achievement, and c) develop the educator's goals and Plan. Goals encompass both practice and student learning, growth and achievement.

Implementation of the Plan. Educator and evaluator collect evidence using the three categories of evidence. Educators receive professional development and support needed to

---

[1] Though they are referred to as Draft Standards and Indicators in this document, the Task Force recognizes that these elements of the Framework may eventually be characterized by ESE as the Revised Principles of Effective Teaching and Administrative Leadership.   To see the draft Standards, go to Appendices I and J.

be successful with their plans, such as additional observation with feedback, release time to observe another educator's practice, or peer review and/or assistance.

Formative Assessment/Evaluation. Formative Assessments allow the evaluator and educator to check in on the educator's progress toward goals, and performance on the Standards. They can include feedback based on observations and walkthroughs (announced and unannounced), educator/evaluator review of student learning, growth and achievement data, instructional rounds, and other sources.

Summative Evaluation. The evaluator assesses the educator's a) performance against the Standards, b) progress made on student learning, growth and achievement goals, and c) progress made on the professional practice goals, and determines overall ratings using the 4-point rating scale and evidence collected from three designated categories of evidence. Summative Evaluations lead to personnel decisions consistent with the provisions of current statute.

4 Paths and 4 Plans differentiated by career stage and performance:
- For teachers without Professional Teaching Status and Administrators in their first three years: Developing Teacher Plans and Developing Administrator Plans
- For experienced Teachers and Administrators rated Proficient or Exemplary: Self-Directed Growth Plans
- For experienced Educators rated Needs Improvement: Directed Growth Plans
- For experienced Educators rated Unsatisfactory: Improvement Plans

Implementation
Every member of the Task Force agrees: effective implementation of the framework is essential. Without it, very little will change. ESE must be willing and able to guide, support and monitor effective implementation at the district and school level. ESE has to put an unprecedented amount of time, thought and resources into this effort. Recommended ESE roles include:
- Fostering local stakeholder engagement in the new framework
- Developing rubrics that clearly illustrate what Standards and Indicators look like
- Developing a model system for districts to adopt or adapt
- Establishing statewide expectations for evaluator knowledge and skill
- Helping districts to develop valid assessments of student learning and growth
- Provide high quality training for all educators involved in evaluation
- Periodically review and revise the Framework based on lessons from the field

Conclusion
The members of the Task Force are clear: educator evaluation in Massachusetts is poised for large-scale transformation, and the work ahead, while sweeping in scope, is both necessary and within the grasp of public educators. The Task Force membership believes that it has made headway on this work, and looks now to both ESE and local districts to pick up the challenge. Working together, the educators and stakeholders of the Commonwealth have the opportunity to make Massachusetts a national leader in the re-invention of educator evaluation.

Introduction

This report presents the recommendations of the statewide Massachusetts Task Force on Educator Evaluation to the Board of Elementary and Secondary Education (BESE), which formally charged the Task Force to:

> "…recommend…a revised set of regulations and principles ("evaluation framework") consistent with the Board's mission statement: "To strengthen the Commonwealth's public education system so that every student is prepared to succeed in postsecondary education, compete in the global economy, and understand the rights and responsibilities of American citizens."
> (See Appendix A for text of the BESE motion.)

In August 2010, Commissioner of Elementary and Secondary Education Dr. Mitchell Chester convened a 40-person Task Force to accomplish this charge. The Task Force included a broad cross-section of stakeholders, representing diverse viewpoints, expertise and perspectives from the leadership of statewide organizations of teachers, principals, superintendents, school committees, and parent organizations. The Task Force also included practicing classroom teachers and administrators, representatives of subject matter associations, special educators and special education administrators, higher education representatives, vocational educators, a student representative, business representatives, and several at-large members with expertise in areas relevant to performance management, psychometrics, economics and statistics.[2]

The Task Force met regularly from August 2010 through March 2011 to develop its recommendations to the Commissioner and BESE. The Task Force created a set of working groups on three subjects: teacher evaluation, administrator evaluation, and cycles of improvement and professional growth. The working groups' recommendations were advisory to the Task Force, which made all final decisions on the recommendations contained in this report. (See Appendix B for a list of Task Force working groups and membership).

Department of Elementary and Secondary Education (ESE) staff and consultants supported and facilitated the Task Force and its working groups, and Associate Commissioner for Educator Policy David Haselkorn served as the Task Force staff director. ESE staff and Task Force members reviewed and considered a wide range of research and opinion on topics related to educator evaluation, performance measurement and human capital development, and studied the approaches of other states and districts. Leading evaluation experts made presentations to the Task Force and its working groups on a variety of issues. (See Appendices C, D & E for lists of staff/consultants, studies reviewed, and presenters.)

This report contains the recommendations of the Task Force, as well as an overview of many of the key issues the Task Force has grappled with in the course of its deliberations. In all work, and in this text, the Task Force consistently used the team "educator" to denote both teachers and administrators.

---

[2] The final membership of the Task Force settled at 39. One member resigned due to other responsibilities.

Educator Evaluation: The National Perspective

Educator evaluation is the focus of intense national discussion and debate. This interest is due, in part, to growing recognition that the single most important school-based factor in strengthening students' educational achievement is the quality and effectiveness of the educators who teach in and lead the schools (Sanders & Rivers 1996; Barber & Mourshed 2007; Rivkin, Hanushek, & Kain 2005; Darling-Hammond & Bransford 2005; Leithwood. Louis & Wahlstrom 2004). This sharpened focus also stems from a series of reports and studies critical of the current status of educator evaluation across the nation and in Massachusetts (The New Teacher Project 2009; Donaldson 2009; The National Council on Teacher Quality 2010). Among the most prominent concerns these studies raise are that current educator evaluation policies and practices:

- Do not provide educators with adequate feedback for improvement
- Lack sufficient connection to goals of student learning and growth
- Fail to differentiate levels of educator effectiveness
- Fail to identify variability in educator effectiveness within schools and across districts

These failures are particularly significant, because they make it hard for schools and districts to capitalize on the knowledge and skills of highly effective educators, promote professional growth and continuous learning, and value and reward excellence. Likewise, they prevent the identification and active support of teachers and administrators who have the potential to become highly effective. Finally, they may inhibit the removal of the small percentage of persistently poor performing educators who fail to make progress, despite being provided reasonable time and support for improvement. Simply put, poor evaluation practices are a missed opportunity for promoting better leading, better teaching, better learning, and better schools.

Educator Evaluation in Massachusetts

The federal government's Race to the Top (RTTT) funding competition made the overhaul of educator evaluation one of its central objectives. RTTT required participating states to have or develop policies that differentiate educator performance by at least three levels and use student learning and growth as a significant factor in educator evaluation. The federal School Improvement Grant program, which focuses on high need schools, requires similar policies.[3]

In May 2010, BESE charged the Task Force to recommend an evaluation framework that:
1. Provides all educators with honest, fair, and improvement-oriented feedback annually
2. Treats educators differently based on their career stage
3. Rates performance on at least three different levels
4. Uses student growth as a significant factor in evaluation
5. Gives districts the flexibility to consider measures of effectiveness beyond those required
6. Establishes a Continuous Improvement Plan for every educator
7. Links comprehensive evaluation to decisions about tenure, career advancement, compensation for additional roles and responsibilities, demotion and dismissal

---

[3] Massachusetts' work on educator evaluation did not begin with Race to the Top. For a summary of the foundation for Massachusetts' recent policy work on educator evaluation, see Appendix F.

In Massachusetts, educator evaluation is governed by a combination of state statutory provisions, state regulatory requirements, and performance standards determined at the local level. This overlapping system of governance allows districts to design evaluation systems that respond to local needs and conditions, subject to state requirements and collective bargaining. The intricacies of this structure place constraints on the creation of a single statewide system for educator evaluation.

The current Regulations for the Evaluation of Teachers and Administrators have remained unchanged since they were first adopted in 1995 in the wake of Massachusetts' landmark educational reforms (An Act Establishing the Education Reform Act of 1993. 1993 Mass. Acts 159. 16 June 1993.). They include a set of Principles for Effective Teaching and Administrative Leadership that serve as "best practice" guidelines for districts to use in establishing their own systems of evaluation.[4]

Of all the charges to the Task Force made by the Board of Elementary and Secondary Education, by far the most important, central and challenging was the BESE requirement to make student learning and growth "a significant factor" in educator evaluation. While current regulations *allow* for student academic achievement to be taken into account in educator evaluation, they do not require that it be used. Few districts in the Commonwealth formally use student learning, growth or achievement in a substantive way in educator evaluation. As a consequence, the knowledge and tools to do so are at a rudimentary level in most districts across the Commonwealth.

> "More than anything, evaluation systems should be recognizing, developing and promoting the most talented and successful educators. We need an approach to evaluation that is all about celebrating excellence, and ensuring that those who excel also thrive in their workplaces, and stay in education. The better we get at developing and rewarding excellence, the better we will get at building schools that succeed for all students."
>
> *~ Representative of Business Leaders & Task Force Member*

Task Force Perspective

In assessing the impact and efficacy of current evaluation policies and practices, while Task Force members expressed a wide variety of views, there was near universal agreement that:

- In its present state, educator evaluation in Massachusetts is not achieving its intended aims: providing educators with adequate feedback for improvement and serving as an important accountability tool to ensure educator effectiveness that supports student learning and growth.

---

[4] The specific performance standards of these systems are established by school committees, subject to collective bargaining.

- A breakthrough is both needed and achievable – to transform educator evaluation from an inconsistently applied compliance mechanism into a statewide catalyst for educator development and continuous professional growth that will provide the highly effective educators our children need to learn, grow and achieve.

Task Force members cite considerable variability statewide in the quality of educator evaluation. They find that the current educator evaluation practice in Massachusetts:

- Fails to identify excellence among educators
- Does not address issues of capacity, or "do-ability"
- Lacks a strong statewide set of common calibration practices, or a way of ensuring that Proficient or Exemplary mean the same thing in two different districts in the state
- Contains too many Standards and indicators
- Puts limited focus on ensuring educator input into the process
- Can contribute to a culture of apathy, mistrust and cynicism
- Relies on training without providing it
- Does not encourage reflective thinking, by either evaluator or educator
- Lacks a focus on continuous improvement

> "Student learning and growth are about more than numbers. Making strong connections with ALL the diverse learners who are my students, motivating them, making sure they really understand, raising their expectations of themselves, collaborating with their families – these are hard to measure, but they are essential to my success as a teacher. We need a common understanding that students' academic growth and progress is not a linear equation, and we need an evaluation strategy that honors this complexity."
>
> *~ Teacher of the Year and Task Force Member*

The Task Force identifies multiple factors that contribute to this variation in quality:
- Common two-scale rating systems fail to identify excellence or achieve accountability
- Lack of resources to support effective implementation
- Inadequate training for evaluators on use of data, ratings, etc.
- No uniform statewide system for calibration or "inter-rater reliability"
- Inadequate time for supervisors to conduct thoughtful evaluations
- Excessive supervisory workloads
- Competing demands on supervisor attention

Notwithstanding these differences, a strong majority of the members were in agreement on the recommendations that follow, and stressed their belief in the need for a breakthrough in educator evaluation in the Commonwealth.

Spirited Discussions within the Task Force

Task Force members wrestled with hard questions, and registered a range of views on them.

Prioritizing the Use of Multiple Measures of Student Learning, Growth and Achievement
While firm in support of the use of multiple measures of student learning, growth and
achievement in educator evaluation, the Task Force did not chart a simple path to that goal. The
majority of Task Force members reject approaches to weighting student learning and growth in a
way that could mechanistically over-ride the professional judgment of trained evaluators and
supervisors, or create an over-reliance on one set of assessments. The use of three categories of
evidence and the assessment of educator progress toward both student learning and professional
practice goals is the alternative worked out by the Task Force.

"Loose" vs. "Tight" The tension between local control and statewide Standards – what the
Task Force came to term the "loose-tight" question – is keenly felt. On the one hand, both
teachers and administrators on the Task Force want a substantial measure of freedom to set a
locally appropriate agenda, and to preserve the bargaining and decision-making rights reserved
to them in current statute. On the other hand, almost all Task Force members agree that the lack
of statewide consistency, comparability, and calibration are major flaws in the current
framework.

Giving Student Outcomes a Significant Role in Evaluation Most members of the Task Force
believe that student outcomes should play a significant role in educator evaluation, but should
not be the primary yardstick. A few Task Force members believe that student learning and
growth, broadly defined, is the most important factor by which an educator's effectiveness must
be measured. Many felt that the inclusion of student outcomes in the framework is in itself a
significant development.

A Range of Views on the Use of Statewide Testing in Evaluation All professional
associations and unions have been consistent in their opposition to basing high stakes
employment decisions for educators on the results of statewide tests developed to assess student
learning, such as MCAS. Other Task Force members argue that statewide tests are more valid
and reliable psychometrically and are tied more closely to state curriculum frameworks than
many district and classroom based assessments or commercially available tests. In their view, it
would be a grave error *not* to use their results in educator performance assessment, among other
multiple measures. However, since more than 80 percent of the state's teachers do not work in
fields or grades assessed by MCAS, and are therefore unaffected by MCAS growth measures,
there is consensus that multiple additional measures of student learning and growth are essential
to the success of a statewide evaluation framework.

A General Concern about Equity Some believe that it is inherently unfair to hold educators
accountable for student learning and growth until there is a level playing field of equitable
resources and adequate school conditions. Many on the Task Force believe that disparities in
conditions or resources must be taken into account during the evaluation process.

Evaluation Framework Recommendations

Values that Inform Effective Evaluation

These core beliefs inform the Task Force's specific recommendations to BESE:

- Student learning, growth and achievement are the primary goals of public education
- Student learning, growth and achievement extend beyond academic progress and include other developmental factors – social and emotional well-being, civic learning and engagement
- Educator effectiveness and student learning, growth and achievement are inextricably linked
- Educator expertise is the foundation of educator effectiveness
- Leadership, school climate and culture are essential elements for supporting the learning and growth of both students and adults
- Evaluation alone cannot guarantee that all educators are effective, but it is an important lever for change
- Changing evaluation practices in schools can require a significant culture shift. Evaluation is often perceived as an obligatory exercise that offers educators limited feedback, does not affect professional growth, fails to distinguish variability in performance, and is ineffective as an accountability tool
- Educator evaluation should be the occasion for data-informed self-assessment and reflection by the educator and improvement-focused collaborative inquiry with their supervisor and, potentially, their peers
- Adequate resources and time are necessary ingredients to meaningful inquiry, evaluation, and improvement

> **On Social and Emotional Growth**
>
> "We have to achieve academic growth in ways we haven't before – this is critical. But every time I had a parent in my office, it was not because their child was not learning the concepts; it was because, in the eyes of the parent, the teacher was not making their child feel safe, accepted and valued. We cannot divorce social and emotional learning from academic learning, and we need to hold all educators accountable for them both."
>
> *~ Former Principal and*
> *Superintendent &*
> *Task Force Member*

This framework is designed to create the conditions for realizing these principles.

Student Learning, Growth and Achievement:
A Significant Factor

Nothing was more central to the Task Force's deliberations than its thorough consideration of the proper role of student learning, growth and achievement data in educator evaluation. This work was spurred both by Task Force members' desire to incorporate student learning outcomes as a factor in the assessment of educator effectiveness and by the requirements of the BESE motion and the Race to the Top funding obligations that this inclusion be "a significant factor" in the new evaluation framework.

Task Force members engaged in a series of discussions, references to research, interactions with experts, and planning and working group meetings – punctuated by vigorous debate. Much of the debate centered around the breadth or narrowness of the definition of "student growth" and the means by which student outcome data would be incorporated into evaluation practices. In the end, the Task Force arrived at three critical conclusions:

1. For purposes of educator evaluation, it is critical to adopt an inclusive definition of student learning, growth and achievement, one that recognizes and assesses the wide range of learning experiences students must have in order to succeed academically

2. In order to achieve accuracy, promote professional growth and ensure accountability, *multiple measures* of student learning, growth and achievement must be considered and used in educator evaluation

3. For student learning to play a significant role in educator evaluation, educator's annual goals must incorporate student learning, growth and achievement

Evaluation Framework: Key Design Features

The Framework has five key design features, which are detailed below. Taken together, they constitute the critical ingredients that every evaluator and educator will use or experience.

4	3	4	5	4
Statewide Standards and Indicators	Categories of Evidence	Statewide Performance Ratings	-Step Evaluation Cycle	Paths **4** Plans

Statewide Standards and Core Indicators for Effective Administrative Leadership and Teaching

The Task Force believes it is critical to develop and adopt a common statewide understanding about what effective teaching and administrative leadership looks like. To this end, it has proposed a set of Draft Standards and Indicators for Effective Teaching and Administrative Leadership (see Appendices I and J).[5] These proposed new Standards and core Indicators streamline the current Principles adopted by BESE in 1995.

The Task Force and its working groups reviewed current work under development at DESE, on new performance indicators for leadership licenses, developed in alignment with relevant national standards. Similarly, in developing its proposed Standards and Indicators for Effective Teaching, the Task Force engaged in an extensive comparison of relevant state and national standards and proposals, including those proposed by the High-Expertise Teaching Project, convened by ESE and its partners over the past two years.

These Standards and Indicators provide multiple functions that the Task Force values. They signal and prioritize the promotion of student learning, growth, and achievement as the primary work of

"We do not choose lightly, or without thorough debate, to include Family and Community Engagement as one of only four Standards for the evaluation of all teachers and administrators. Our choice is based on thirty years of national research demonstrating that school-family partnerships are crucial to student achievement, and responds directly to the public's keen interest in ensuring students' academic success. The research is unambiguous: when teachers and administrators engage with families, student achievement rises."

~ Parent and Family Advocate
& Task Force Member

---

[5] The draft Standards and Indicators for Effective Administrative Leadership and Effective Teaching were reviewed by the Task Force and Working Groups on multiple occasions. There was general agreement on the Standards identified for each group of educators, but due to a lack of time, the Task Force was unable to approve the core Indicators for each of the Standards. The Task Force strongly encourages ESE to use the proposed drafts included in the Appendices to complete this work, in consultation with external stakeholders as appropriate.

educators at all levels of education. They serve as the spine of the new evaluation framework, and will do so in the evaluation systems that districts adopt. Beyond the core Standards and Indicators that the Task Force recommends be adopted statewide, there is a keen appreciation that educators and school committees in different communities across the state may want to supplement these essentials with others that they deem critical to success locally.

Draft Statewide Standards for Effective Administrative Leadership			
Curriculum, Instruction, and Assessment	Management and Operations	Family and Community Engagement	Professional Culture
The education leader promotes the learning and growth of all students and the success of all staff by cultivating a shared vision that makes powerful teaching and learning the central focus of schooling.	The education leader promotes the learning and growth of all staff by ensuring a safe, efficient and effective learning environment.	The education leader promotes the learning and growth of all students and the success of staff through partnerships with families, community members, and other stakeholders that support the mission of the school and district.	The education leader promotes success for all students by nurturing and sustaining a school culture of professional growth, high expectations, and continuous learning for staff.

Draft Statewide Standards for Effective Teaching			
Curriculum, Instruction, and Assessment	Teaching All Students	Family and Community Engagement	Professional Culture
The teacher promotes the learning and growth of all students through planning, instructional and assessment activities that support a cycle of creating lessons focused on clear learning objectives, designing authentic and meaningful student assessments, analyzing student performance and growth, and continuously refining learning objectives.	The teacher promotes the learning and growth of all students through instructional practices that establish high expectations, create a safe and effective classroom environment, and demonstrate cultural proficiency.	The teacher promotes the learning and growth of all students through effective partnerships with families, caregivers, community members, and organizations.	The teacher promotes the learning and growth of all students through ethical, culturally proficient, skilled and collaborative practice.

Three Categories of Evidence

The Task Force calls for three categories of evidence to be used in every district's educator evaluation system to assess educator needs and performance. Subject to collective bargaining, districts are also free to include additional relevant evidence that has been shared between the educator and evaluator.

Multiple Measures of Student Learning, Growth and Achievement
- Measures of student progress and/or achievement toward student learning targets set between educator and evaluator for the academic year.
- MCAS growth measures in comparison to comparable schools, based on appropriate school-level demographics, where applicable. Other statewide measures, such as MEPA, should be used, where applicable.
- District-determined measure(s) of student learning, comparable across grade or subject district-wide.

Optional but encouraged:
- Group measure(s) aligned with designated team, grade, department, or school-level goals.

Judgments based on Observation and Artifacts of Professional Practice
All districts adopt either the DESE-designed comprehensive observation system or use a locally developed observation system that is approved by DESE. To be approved, systems need to:
- Align with the evaluation Standards and Indicators adopted by BESE
- Use the statewide rating scale adopted by BESE
- Capture meaningful and observable differences in educator performance
- Be informed by research and best practices

Observation of practice may also include other evidence of professional practice observed by evaluators in making judgments, such as lesson plans, unit plans, school improvement plans, district budgets, IEPs, redacted written evaluations, etc.

Collection of Additional Evidence Relevant to one or more Standards
All educators will compile evidence of their work that documents fulfillment of professional responsibilities, professional growth, and contributions to the school community and the professional culture. Documentation will include, at a minimum:
- Evidence of fulfillment of professional responsibilities and growth, such as self-reflection(s) and goals; classroom-based action research projects; peer collaboration; professional development; and contributions to the school community and professional culture
- Evidence of the educator's outreach to and engagement with families
- Evidence of broad based parent and student input or feedback
- For administrators, evidence of staff input or feedback

Any other artifacts or evidence to be included would be determined at the district level. These might include peer observations, evidence from school climate and culture surveys such as Mass TeLLS (for administrators), and/or other school and district surveys.

More on Multiple Measures

The Task Force has paid particular attention to the use of multiple measures of student learning, growth and achievement as a significant factor in evaluator judgments. It finds that:

- All educators are responsible for bringing to their evaluators evidence of student learning, growth, and achievement for students under their responsibility.
- All evaluators are responsible for analyzing multiple measures of student learning, growth and achievement for students under the responsibility of the educator
- Discrepancies between the evaluator's analysis of multiple measures of student learning, growth and achievement and the evaluator's ratings based on the Standards of effective teaching and administrative leadership must prompt further discussion between educator and evaluator, and further analysis and review of the data by the evaluator.
- In cases where there are significant discrepancies between evidence of student learning, growth and achievement and the performance ratings that the evaluator makes, the evaluator's supervisor must discuss and review these with the evaluator and render a judgment about the quality of the evaluator's performance as an evaluator.
- Conditions and resources needed by the educator to meet state Standards should be considered by the evaluator, where appropriate.

"Student assessment data, such as MCAS, must be a part of the teacher evaluation process, but it can never be the sole measure of what is working in our classrooms. It is important to incorporate multiple measures of teacher impact in a teacher's evaluation. As a Kindergarten teacher, I believe assessments are important, and it is my responsibility to use the data from these assessments on a regular basis to evaluate and improve my practice, and ultimately strengthen outcomes for the students in front of me."

~ Teacher and Task Force Member

Statewide Performance Rating Scale

The Task Force recommends that all school districts in Massachusetts use the following four rating categories in both the Formative and Summative stages of the evaluation cycle. Evaluators would rate educators on each Standard, on progress towards achieving the goals, and in determining an overall rating:

Statewide Performance Rating Scale	
Category	**Definition**
Exemplary	Practice is consistently and significantly above proficiency on the Standard or overall
Proficient	Practice demonstrates skilled performance on the Standard or overall
Needs Improvement	Practice demonstrates lack of proficiency on the Standard or overall
Unsatisfactory	Practice demonstrates lack of competence on the Standard or overall

5-Step Evaluation Cycle

The Task Force recommends a 5-stage cycle[6] for educator evaluation, differentiated by the educator's career stage and performance. Experienced educators with ratings of Proficient or Exemplary use a two-year cycle; in the first year they are rated against their goals; in the second year, against all four Standards. All other educators use a cycle that lasts a year, at the most.[7]

As mentioned earlier, the term "educators" refers to *both* teachers and administrators throughout this report.[8] Evaluators typically are administrators authorized to conduct evaluations, but districts might decide that evaluators could include peer reviewers as well.

---

[6] As will be explained in this section, the order of these steps may vary slightly, depending on educator's career stage and/or performance.

[7] Task Force members recognize that each district will have to determine an initial rating for experienced educators in order to initiate the new framework.

[8] The proposed framework covers a full range of teacher and administrator roles; for example, teachers would include classroom teachers, caseload educators (counselors, guidance counselors, school psychologists), special education teachers and others, and administrators would include principals, vice principals, directors of special education, department heads and others.

5-Step Evaluation Cycle
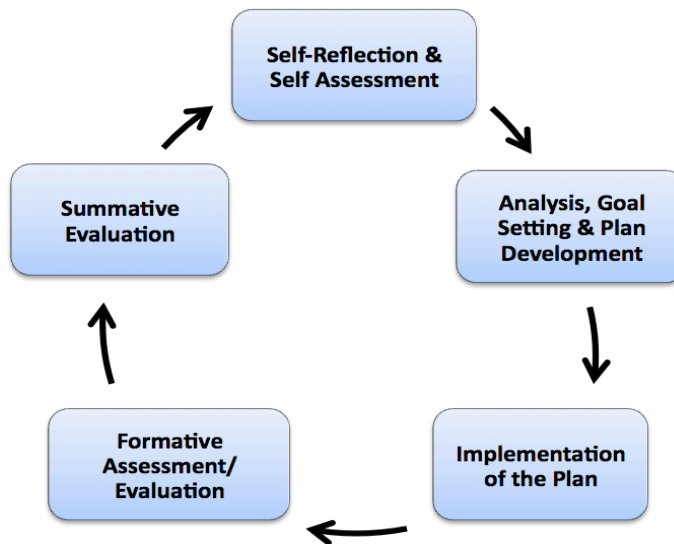
1. Self-Reflection and Self-Assessment

Two core principles emerged from the deliberations of the Task Force: that educators a) engage in on-going improvement of their own professional practice, and b) take responsibility for their students' learning, growth and achievement. The evaluation process begins with educators reflecting on and assessing their professional practice, and analyzing the learning, growth, and achievement of their students. Reflection and assessment helps the educator to identify areas of practice to develop or refine. During this stage of the Evaluation Cycle, educators summarize their reflections in a narrative that includes:

- Review of available multiple measures of student learning and growth
- Self-assessment of educator practice against the Standards
- The educator's proposed goals for the coming year, both for the improvement of practice and for the improvement of student learning and growth

2. Goal Setting and Development of a Plan

Each educator meets with his or her evaluator to: a) review the self-reflections and self-assessments, b) jointly analyze students' learning, growth and achievement, and c) develop the educator's Plan. The meeting is an opportunity for collaborative inquiry into professional practice and student outcomes. All educators, regardless of career stage or evaluation rating, engage in Plans that:

- Include goals to improve both student learning, growth, and achievement and educator professional practice
- Are aligned to Standards and Indicators
- Are informed by district and school goals

The Task Force recommends the following four Paths and four Plans:

4 Paths, 4 Plans for Educators, Differentiated by Career Stage and Performance		
Path	**Plan**	**Description** *All Plans include measurable goals, both for learning, growth, and achievement of students under the educator's responsibility, and for improvement of the educator's practice, as evaluated against the Standards.*
Administrators in first 3 Years; Teachers without PTS (Professional Teaching Status)	Developing Teacher / Administrator Plans	Educators and evaluators work together to develop a **Developing Teacher/Administrator Plan**.
Administrators / Teachers rated "Proficient" and "Exemplary"	Self-Directed Growth Plans	Educators develop a **Self-Directed Growth Plan**.
Administrators / Teachers rated "Needs Improvement"	Directed Growth Plans	Educators and evaluators work together to develop a **Directed Growth Plan** that focuses on specific areas for improvement.
Administrators / Teachers rated "Unsatisfactory"	Improvement Plans	Educators and evaluators work together to develop an **Improvement Plan** that focuses on areas in which the educator must improve during a specific time period.

3. Implementation of the Plan

The educator and evaluator collect evidence using the three categories of evidence. Educators receive professional development and support needed to be successful with their plans, such as additional observation with feedback, release time to observe another educator's practice, or peer review and/or assistance.

4. Formative Assessment /Evaluation

Formative Assessments are a means for evaluator and educator to check in on the educator's progress toward goals and performance on the Standards. Formative assessments can include feedback based on observations and walkthroughs (announced and unannounced), educator/evaluator review of student learning, growth and achievement data, instructional rounds, and many other sources. All educators receive formative assessments, but they vary by the career stage and effectiveness of the educator.

- o For educators on a *Self-Directed Growth Plan*, in the first year, evaluators base their Formative Evaluation rating on progress made towards completing the goals outlined in the Plan. In the second year, evaluators base their Summative Evaluation rating on performance against the Standards.

o   For educators on *Directed Growth, Improvement* or *Developing Plans*, evaluators complete formative assessments that take the form of Interim Reviews, completed midway through the school year. Interim Reviews are an opportunity to check on the educator's progress towards completing the goals outlined in the Plan and performance on the Standards. This progress towards both – goals in the Plan and performance against the Standards – determines a mid-year Formative Assessment.
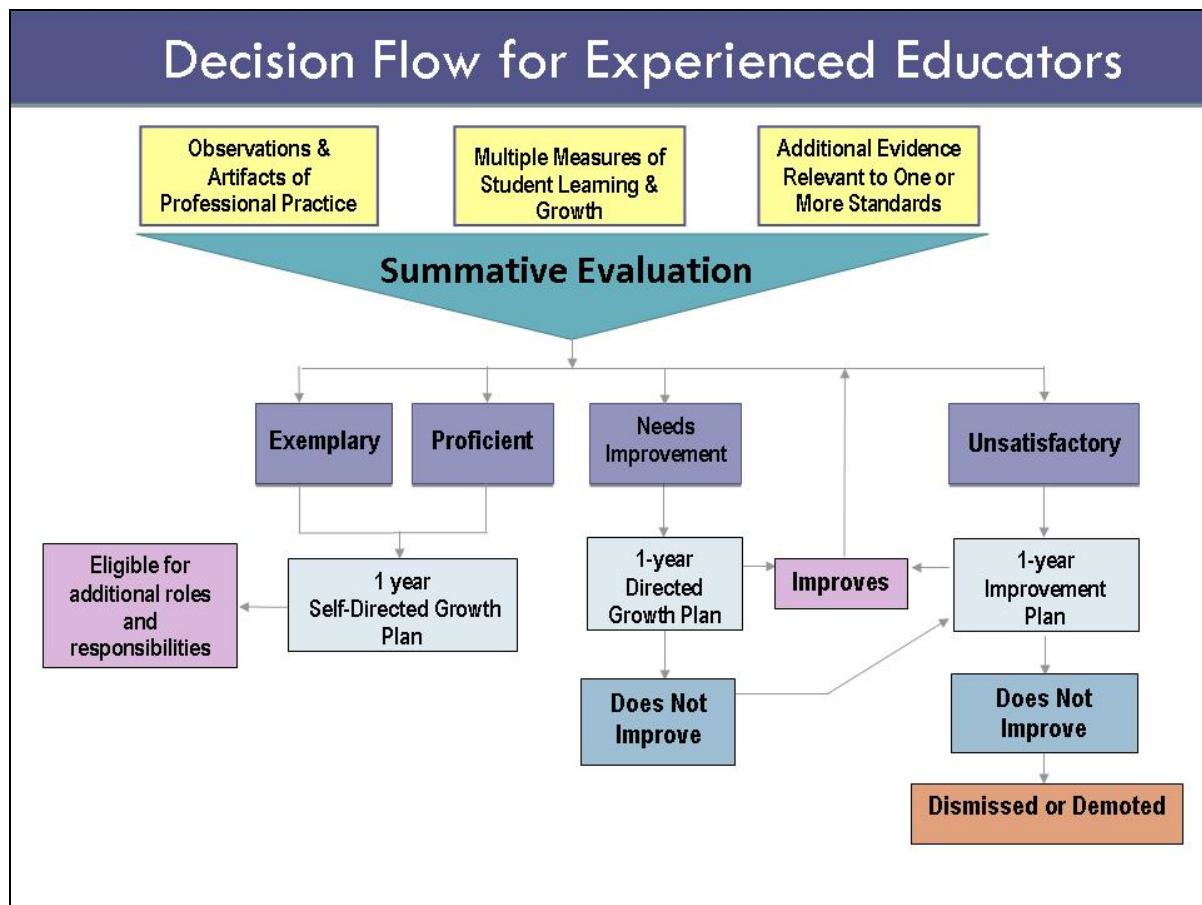
5. Summative Evaluation

Evaluators complete Summative Evaluations for educators at the conclusion of their evaluation cycle. The evaluator determines overall ratings using the 4-point state rating scale and evidence collected from three designated categories of evidence. The evaluator assesses the educator's: a) performance against the Standards, b) progress made on student learning, growth and achievement goals, and c) progress made on the professional practice goals. When there is a discrepancy between measures when determining a rating for any single standard, the evaluator uses professional judgment. When there is a discrepancy between ratings on individual Standards, or when determining an overall rating, the evaluator's professional judgment is used to arrive at a rating.

Evaluators complete Summative Evaluations within a year for educators on *Improvement, Directed Growth, or Developing Plans,* and at the end of the second year of the cycle for those on *Self-Directed Growth Plans*. When an educator on an *Improvement Plan* does not show sufficient improvement on their Summative Evaluation, the evaluator makes a personnel decision consistent with the provisions of current statute, which specify that "the results of … evaluations may be used in decisions to dismiss, demote or remove a teacher or administrator."[9] Educators rated Proficient and Exemplary can use the conclusion of the Summative Evaluation process to generate new self-reflections and self-assessments, and to jointly craft, with the evaluator, their next cycle's Self-Directed Growth Plan.

Decisions flow from the Summative Evaluation. The chart below summarizes the flow of decisions for experienced educators. For other educators – teachers without PTS, administrators with less experience – decisions can be made at end of each year.

---

[9] http://www.malegislature.gov/Laws/GeneralLaws/PartI/TitleXII/Chapter71/Section42

Decision Flow for Experienced Educators

Observations & Artifacts of Professional Practice

Multiple Measures of Student Learning & Growth

Additional Evidence Relevant to One or More Standards

Summative Evaluation

Exemplary | Proficient | Needs Improvement | Unsatisfactory

Eligible for additional roles and responsibilities

1 year Self-Directed Growth Plan

1-year Directed Growth Plan → Improves ← 1-year Improvement Plan

Does Not Improve

Does Not Improve

Dismissed or Demoted

The Task Force recommends that ratings inform key personnel decisions. Some examples: Educators who receive an Exemplary or Proficient rating may be eligible for leadership roles, such as Mentors, Coaches, and Team Leaders. Educators rated Needs Improvement who do not meet the goals of their Direct Growth Plans may be placed on an Improvement Plan for a period not to exceed a year. Educators on Improvement Plans whose ratings indicate lack of competence on the Standards and minimal progress on the Goals maybe demoted or dismissed. New teachers must be rated at the Proficient level or above on all Standards to be granted Professional Teacher Status (PTS).

The Implementation Challenge

Every member of the Task Force agrees. The evaluation practices called for in its proposed framework are vastly different from those in place in many of the Commonwealth's schools. Effective implementation of the framework is absolutely essential. Without it, very little will change. As one Task Force member said, "the framework will be just a piece of paper if schools don't have great support for implementation."

ESE must be willing and able to guide, support and monitor effective implementation at the district and school level. ESE has to put an unprecedented amount of time, thought and resources into this effort.

Task Force members are frustrated that they were unable to delve deeply into the many challenges of implementation and offer more than basic recommendations. They see the need for ESE to seek out stakeholders and others with expertise in implementing new performance management systems to provide guidance and candid feedback about its plans and progress.

ESE has $10,000,000 in Race to the Top funding to support effective district implementation of the new evaluation framework. Below are arenas and examples of the responsibilities the Task Force sees ESE assuming:

Local stakeholder engagement
- Convene school committee, union, district and school leaders together to present expectations for educator evaluation and opportunities for collaboration, networking and support
- Make clear specifically what will be required for districts to have their educator evaluation systems approved by ESE as meeting new state regulations.

Statewide Standards and Indicators
- Finalize and adopt statewide Standards and Indicators
- Develop rubrics that provide clear examples of what the Standards and Indicators look like at different levels of performance: exemplary, proficient, needs improvement, unsatisfactory

> **On Implementation**
>
> Task Force Member, in a meeting: "Whatever model for evaluation is adopted, it needs to be practical. We need to have the conversation – 'Is this doable?'"
>
> Task Force Member, in reply: "I would slightly change the question from 'Is it doable?' to 'How can we make it doable? What will it take to make it possible?'"

Models for districts to adopt or adapt
- Develop model systems for teacher and administrator self-assessment, goal setting and the other components of the educator evaluation framework – complete with rubrics, protocols, templates and exemplars
- Support and develop meaningful and reliable approaches to gathering feedback from students and families

Statewide expectations and training for evaluators
- Establish clear standards for evaluations and evaluator knowledge and skill
- Design and deploy tools for evaluator training, assessment and rater calibration
- Consider establishing guidelines for how many educators an evaluator can reasonably be expected to supervise and evaluate effectively

Student Assessment
- Help districts develop or access valid assessments of learning and growth, across all subjects and grades, that supplement the statewide assessment system

Professional development
- Provide high quality training and professional development in all relevant aspects of the evaluation framework for administrators, school committees, teachers and others involved in implementing the framework at the local level
- Make available face-to-face, distance learning, web-based and networking opportunities

Continuous Improvement
- Periodically review, refine, and revise the Evaluation Framework based on implementation lessons learned from the field, as well as local and national research

No member of the task force wants to see the work of the past seven months squandered by inadequate support at the state or local level. As more than one member of the Task Force said, transforming educator evaluation is a "heavy lift." ESE needs to commit its all to the challenging work ahead.

ESE's Model Evaluation System for Administrators and Teachers

As reported to the Task Force by Deputy Commissioner Karla Baehr, developing a model evaluation system for administrators and teachers is central to ESE's plan for assisting districts.

The Board of Elementary and Secondary Education called upon ESE to develop and disseminate a model educator evaluation system *for principals and teachers* in spring 2011 to assist districts with Level 4 schools. Starting in fall 2011, districts with Level 4 schools receiving federal school redesign grants must be using evaluation systems that "differentiate performance by at least three levels" and "use student growth as a significant factor in evaluation." Built on the framework developed by the Task Force, ESE's model can serve as the starting point for these districts now. It is being designed to support implementation in all other districts, as well (Race to the Top districts will be implementing new systems by fall 2012; all other districts, by 2013.)

ESE's Center for Targeted Assistance is working with district and union leaders from the Level 4 districts to refine and disseminate the initial version of the model system. Task Force recommendations are serving as its foundation. Its final form will be consistent with Board regulations to be adopted in 2011. Its key features are expected to include:

- Contract language describing process, timelines and collection of evidence
- A rubric for each Standard and Indicator that describes performance vividly and clearly at four levels of performance
- Templates for self-assessments and growth plans
- Guidelines for developing and using measures of student learning and growth
- Examples of ways to collect and use student, staff and parent feedback (initially for administrators)

ESE will update the model at least annually to reflect new knowledge from the field. The new evaluation framework gives us the opportunity to share a common vision of what excellent teaching and leading look like. Therefore, ESE's support for local district implementation will be built around the model system. For example, on-line training and resources for administrators and teachers will use the performance rubrics as their basis. Assessments of evaluator's knowledge and skill at assessing practice will also use the rubrics as their basis.

Conclusion

The members of the Task Force are clear: educator evaluation in Massachusetts is poised for large-scale transformation. The work ahead, while sweeping in scope, is both necessary and within our grasp.

Improving educator evaluation is an important step in a longer range effort to systematically recruit, retain, support, and reward effective educators in *all* of the Commonwealth's districts, schools, and classrooms. Ensuring that there are effective teachers and administrators in every classroom, school, and district in the Commonwealth will not be achieved by improved evaluation policies and practices alone, but it *cannot* be achieved without them.

The Task Force believes that its work provides a blueprint for all Massachusetts schools and systems for differentiating performance, identifying best practices and exemplary performers, providing better supports to educators who are struggling or need improvement, and better connecting personnel decisions and rewards to performance.

Massachusetts is on its way to ensuring that powerful educator evaluation practices are in use throughout its schools and districts. The Task Force now looks to DESE, to local districts, and to public education's many stakeholders to pick up the challenge. Working together, the educators of the Commonwealth have the opportunity to make Massachusetts a national leader in the re-invention of educator evaluation.

> **A Call for 'Reflective Practice' by the State**
>
> "As a Commonwealth, we are on a steep learning curve. There has never been more research and experimentation underway on evaluating teachers and administrators than there is right now. Many lessons will be learned in the next few years. It will help a great deal to leave the new evaluation framework open to changes based on evidence and practice. Many of us on the Task Force hope that the Commonwealth will deliberately plan for a recalibration and improvement of the new framework in future years, as the field gets wiser and more experienced at including student learning and growth data in all evaluations.
>
> *~ Task Force Member*

By holding the highest expectations for educators and students alike, by bringing the best of local practice and national research to bear on the challenges, by finding common ground and collaboratively developing a shared vision for student success and professional excellence, and by championing a robust civic and institutional investment in this work, we can achieve this goal.

References

Babu, S. and Mendro, R. (2003). *Teacher accountability: HLM-based teacher effectiveness indices in the investigation of teacher effects on student achievement in a state assessment program.* American Educational Research Association annual meeting.

Darling-Hammond, L., and Bransford, J. (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do.* San Francisco: Jossey-Bass.

Donaldson, M. L. (2009, June). *So long, Lake Wobegon? Using teacher evaluation to raise teacher quality.* Center for American Progress.

Mendro, E., Gomez, H., Anderson, M., Bembry, K. (1998). Longitudinal teacher effects on student achievement and their relation to school and project evaluation. American Educational Research Association annual meeting.

McKinsey & Company. (2007, September). *How the world's best-performing school systems come out on top.* Sir Michael Barber & Mona Mourshed.

The National Council on Teacher Quality. (2010, February). *Human capital in the Boston Public Schools: Rethinking how to attract, develop, and retain effective teachers.*

The New Teacher Project. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* Daniel Weisberg, Susan Sexton, Jennifer Mulhern, & David Keeling.

Rivkin, S.G., Hanushek, E.A., Kain, J.F. (2005). Teachers, schools, and academic achievement. *Econometrica.* Princeton, 73(2), 417-458.

Sanders, W. L. & Rivers, J. C. (1996). *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement.* Knoxville, TN: University of Tennessee – Value Added Research and Assessment Center.

The Wallace Foundation. (2004). *How leadership influences student learning.* Kenneth Leithwood, Karen Seashore Louis, & Kyla Wahlstrom.

Appendices

Appendix A: BESE Motion creating the Task Force

Appendix B: Membership of the Task Force Working Groups

Appendix C: ESE Staff to the Task Force and Consultant Team

Appendix D: Annotated Bibliography of Materials Considered by the Task Force

Appendix E: List of Presenters to the Task Force and Working Groups

Appendix F: Educator Evaluation Policy in Massachusetts

Appendix G: Essential Conditions for School Effectiveness

Appendix H: Glossary of Terms

Appendix I: Draft Standards of Effective Administrative Leadership

Appendix J: Draft Standards of Effective Teaching

Appendix A – Board Motion Creating the Task Force

Board of Elementary and Secondary Education Meeting: May 25, 2010
Policy Direction on Evaluation of Teachers and Administrators

On a motion duly made and seconded, it was:

VOTED: that the Board of Elementary and Secondary Education, in accordance with
Chapter 69, Section 1B and Chapter 71, Sections 38 of the Massachusetts General
Laws, hereby direct the Commissioner to establish a Task Force on Evaluation of
Teachers and Administrators. The task force shall review the Board's Regulations
on Evaluation of Teachers and Administrators, 603 CMR 35.00, and the
Principles of Effective Teaching and Principles of Effective Administrative
Leadership incorporated therein, and shall recommend, no later than January 31,
2011, a revised set of regulations and principles ("evaluation framework")
consistent with the Board's mission statement: "To strengthen the
Commonwealth's public education system so that every student is prepared to
succeed in postsecondary education, compete in the global economy, and
understand the rights and responsibilities of American citizens."

Further, that the Task Force on Evaluation of Teachers and Administrators shall
recommend a state evaluation framework that:

1. provides teachers and principals with honest, fair, and improvement-oriented
   feedback annually,
2. differentiates by career stage and ensures flexibility for districts to consider additional
   measures of effectiveness beyond those required in the framework,
3. establishes a two-year cycle of improvement via a formative assessment and
   summative evaluation based on a Continuous Improvement Plan for every educator.
   a. *For teachers*, the Continuous Improvement Plan will define goals for
      improving teaching performance and student performance, the
      professional development (content-based or other) to achieve these
      goals, other professional support such as coaching, and interim
      benchmarks that may include observations of teacher work, student
      work, and teacher work products.
   b. *For principals and administrators*, the Continuous Improvement Plan
      will define goals for improving administrative performance and student
      performance, the professional development to achieve these goals, other
      professional support such as coaching, and interim benchmarks that may
      include observations by supervisors and administrator work products.
4. differentiates performance by *at least* three rating categories based on student growth
   as a significant factor with other measures of effectiveness for the purpose of
   establishing the requirements of the Continuous Improvement Plan.
5. incorporates categories of appropriate data and information to be used in evaluations:
   a. Measures of student growth will include trends in the MCAS growth model
      where they apply, along with state, district, school, and/or teacher-generated

assessments that are comparable across subjects and grades, such as beginning- and end-of-year tests, performance tasks, portfolios of student work, and other student work products.

  b. Student performance will be determined through locally developed and/or publisher-created measures that assess student academic improvement and are reliable and comparable across similar subjects and/or grades in the school and/or district.

  c. Other measures of educator effectiveness might include:
       i. *For teachers:* Supervisor ratings using research-based observational tools and rubrics; evidence of content knowledge, professional skills, cultural competency, professional growth; teacher self-assessments; peer observations; additional student, classroom, team, and school measures including indicators of school culture, climate, and conditions.
       ii. *For principals and administrators:* Supervisor ratings; professional skills in such areas as strategic planning, instructional leadership, evaluation and supervision, cultural competence, human resources and development, management, external development, and micro political leadership; professional growth; principal self-assessments; peer observations; additional student, classroom, team, and school measures including indicators of school culture, climate, and conditions.

6. Links comprehensive evaluation to key personnel decisions, as permitted by law and/or as provided by contract, including:
   a. Professional teaching status (tenure),
   b. Career advancement through a teacher leadership career ladder,
   c. Compensation for additional roles and responsibilities and for hard to staff schools, and
   d. Dismissal and demotion (A teacher or principal identified as ineffective who does not make acceptable progress toward achieving the goals of his/her continuous improvement plan after at least one year of intensive support can be dismissed or demoted.)

Further, that the Task Force on Evaluation of Teachers and Administrators will include:
1. representatives from all MassPartners organizations (the state associations of superintendents, school committees, teachers, elementary and secondary school principals, and parents),
2. representatives from statewide counseling and special subject organizations, e.g., guidance, reading, arts, vocational/technical schools,
3. parents who reflect experience with children with disabilities, English language learners, and as PTO members, and
4. at least one student representative chosen by the State Student Advisory Council.

Further, that the Commissioner shall present proposed amendments to the Regulations on Evaluation of Teachers and Administrators, 603 CMR 35.00, and

the Principles of Effective Teaching and Principles of Effective Administrative Leadership to the Board for review in February 2011, in accordance with the Administrative Procedure Act.

Appendix B – Members of the Working Groups

Teacher Effectiveness
Working Group
Lisa Famularo
Robert Fraser
Jon Fullerton
Caitlin Hollister
Elsie Huang
Joam Marmolejos
Seth Moeller
Constance Moore
Dan Murphy
Floris Wilma Ortiz
Paul Toner
Shakera Walker
Martin West

Administrator Effectiveness
Working Group
Patty Barrett
Anna Bradfield
Tom Fortmann
Amanda Green
Linda Hayes
Nadya Higgins
Pamela Hunter
Carla Jentz
Elizabeth Pauley
Steve Rivkin
Kathie Skinner
Mary Ann Stewart

Continuous Improvement Plan
Working Group
Mary Czajkowski
John D'Auria
Michael Flynn
Amanda Green
Orin Gutlerner
Carla Jentz
Glenn Koocher
Jim Lynch
Dan Murphy
Linda Noonan
Tom Scott
Norm Shacochis
Kathie Skinner
Jesse Solomon

Ad Hoc
Working Group
John D'Auria
Phil Flaherty
Amanda Green
Kathie Skinner

Appendix C – ESE Staff to the Task Force and Consultants

ESE Staff
David Haselkorn, Task Force Director
Claudia Bach
Karla Baehr
Matthew Borek
Deborah Dahl
Bob Lee
Christina Lento
Elizabeth Losee
Shana Pies

Task Force Consultants
Heather Peske, TeachPlus
Emily Kalejs Qazilbash

Task Force Facilitators
Andrew Bundy, Community Matters
Lainy Fersh, Community Matters

Appendix D – Annotated Bibliography of Studies Reviewed

3X for all: Extending the reach of education's best – Emily and Bryan Hassel, Public Impact (2009)
This work argued that top-tier teachers (who achieve 3 times the learning gains) should be more available. They offered suggestions (in-person reach extension, remote reach extension, and boundless reach extension) and a clear description of the '3X economy'. In addition, the authors examined barriers to implementation, measurement issues, and possible spaces for innovation.
http://www.publicimpact.com/images/stories/3x_for_all_2010-final.pdf

AFT - A New path forward: Four approaches top quality teaching and better schools / A Continuous improvement model for teacher evaluations
A speech by Randi Weingarten from January, 2010, outlining approaches for schools to stay competitive in a global economy that requires higher level thinking and skills such as creativity and problem solving. She argues that evaluation systems must use multiple forms of data to give timely feedback, and identify effective practices, so long as there is meaningful training for evaluators. "A continuous improvement model" describes some of the necessary elements to the proposed approach.
http://www.aft.org/pdfs/americaneducator/spring2010/Weingarten.pdf
http://www.aft.org/pdfs/press/improvemodel011210.pdf

Assessing the potential of using value-added estimates of teacher job performance for making tenure decisions – Dan Goldhaber and Michael Hansen (2009)
This paper explored the possibility of using VAM as primary criteria in tenure decisions, using data from North Carolina. The authors argue that VAM may provide better indicators of teacher quality, and estimates of long-term performance, than a number of observable teacher attributes. However, the authors conclude that significant changes to tenure policy may drastically alter the composition of the workforce, so they suggest that the results be interpreted cautiously.
http://www.urban.org/UploadedPDF/1001265_Teacher_Job_Performance.pdf

Beyond basic skills: The Role of performance assessment in achieving 21st century standards of learning. Linda Darling-Hammond and Frank Adamson, Stanford Center for Opportunity Policy in Education (SCOPE) (2010).
This work places 21st century skills in the context of NCLB, and argues that new performance assessments are needed to accurately assess learning. On p. 22 there is a section entitled 'The challenges of performance assessments' that provides a useful overview of the topic, and could apply to teachers; the next section addressed how assessments affect teaching and learning
http://edpolicy.stanford.edu/pages/pubs/pub_docs/assessment/scope_pa_overview.pdf

Can teacher quality be effectively assessed? Goldhaber and Anthony (2004)
This work analyzes the relationship between National Board Certification and elementary student achievement. Estimation models used are presented, in a somewhat technical fashion. The authors find a relationship between certification and student achievement and that NBPTS certification identified effective teachers. They also find that unsuccessful candidates were

actually less effective in the year in which they applied and that effectiveness after certification is mixed.

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 – Corcoran, Annenberg Institute for School Reform (2010)
This work reviews the evidence on the shift from measures of teacher performance based on teacher inputs to one focused on student outcomes. There is a section reviewing 'What is VAM?' using practical examples. In addition, section 4 outlines conceptual and practical challenges to the use of VAM. This is a useful primer on the characteristics of VAM. The overall argument is that evaluation systems (and the corresponding human capital responses) should not be tied simply to VAM, because of the many challenges described in the paper.
http://www.annenberginstitute.org/pdf/valueAddedReport.pdf

Can you recognize an effective teacher when you recruit one? – Johan Rockoff, Brian Jacob, Thomas Kane, and Douglas Staiger (2009)
This study seeks to determine if characteristics not typically analyzed through research, or collected by administrators, can predict teacher effectiveness, as well as a comparison with the Haberman Star Teacher Evaluation PreScreener. The authors use data from an online survey of teachers in NYC, as well as a number of academic measures, such as SAT scores. "Traditional" indicators, such as a graduate degree or passing the basic competency test on the first attempt, were not statistically significant in terms of correlation with student achievement. The authors conclude that when recruiting teachers, a broad set of credentials should be used, including some that are not typically collected by school districts.
http://closup.umich.edu/files/closup-wp-11-recognize-effective-teacher.pdf

Challenges in evaluating special education teachers and English language learner specialists – Holdheide, Goe, Croft, & Reschly (NCCTQ) (2010)
Using surveys and interviews, NCCTQ attempted to determine the opinions of special education and ELL teachers regarding evaluation. Most respondents indicated that modification of evaluation is not allowed, while nearly half responded that separate evaluation processes are desired. Few respondents indicated the use of student performance data (i.e. – standardized tests) but instead used IEPs or other individualized measures. Specific issues highlighted in the paper include: using value-added models for populations whose learning trajectories are atypical, or whose classroom sizes are small and possible contain students from mobile populations; crediting specialist teachers for student learning gains; accounting for alternative assessments; and differentiated training needed to observe these classrooms. The use of multiple measures is stressed, yet even with the multiple measures there is a need for training of evaluators on specialized classrooms. Differentiated evaluation criteria may also be necessary, where appropriate. Practical examples are included throughout the paper.
http://www.tqsource.org/publications/July2010Brief.pdf

Current approaches to defining and measuring educator effectiveness (Policy Brief #4) – Brett Lane (2009)
Summarized the work of Milanowski (2009) and focused on the topic of assessing teacher performance – looking at 7 systems and categorizing on the type of information collected by

each. Contains a central focus of 'what do these systems actually measure' as an overview and there is also a connection to how information is collected.

A Decade of Boston school reform: Reflections and aspirations – Rennie Center (2006)
This work summarized a number of reports that were prepared as Superintendent Payzant left BPS. The first topic covered BPS's efforts to rebuild the human resource system – this was highlighted as an area of success to build upon. Similar summaries were done for leadership development, instructional improvement, using data to inform decision-making, high school reform, and work with students with disabilities.

Determining processes that build sustainable teacher accountability systems – Lucy Steiner (Public Impact), National Comprehensive Center for Teacher Quality (2009)
This brief explored the change processes involved in creating and implementing new evaluation systems. Their study asked select districts about: the impetus for change; the goal; stakeholder involvement; outcomes; challenges; and responses to challenges. An appendix provides details of each district's context. This document is useful to analyze lessons learned from others.
http://www.tqsource.org/publications/October2009Brief.pdf

Developing a psychometrically sound assessment of school leadership: The VAL-ED as a case study – Porter, Polikoff, Goldring, Murphy, Elliott, & May (2010)
The authors begin with the premise that leadership quality will be improved by evaluating principals on the improvement of behaviors shown to improve student learning. The article describes the iterative research process used in the development of the VAL-ED instrument. The article is somewhat technical in nature, as item analysis is detailed, in terms of overall questions of validity of the instrument, but remains accessible and is useful in understanding how VAL-ED was developed and the research questions that drove its design and revision.

Developing tools for identifying effective teaching (Powerpoint) - Thomas Kane, Gates Foundation (2008)
This provides a concise overview of preliminary work and results from the *Measures of Effective Teaching* project. This is really more of a 'what's to come' presentation than anything containing tangible products.

Devil in the details: An Analysis of state teacher dismissal laws – Bireda (American Progress) (2010)
This paper places dismissal reform in the larger context of improving human capital systems. The policies of various states are reviewed in terms of obstacles to dismissal and protections in place for teachers. Recommendations include: state laws should differentiate dismissal policies by particular performance issues; consideration of a state-run system of hearing officials; non-binding mediation sessions; clarification of vague legal language; and working collaboratively with unions to create fair and efficient processes, including peer assistance and review.

Empowering effective teachers: Readiness for reform – Gates Foundation (2010)
The work focused on 4 main conditions of readiness: shared leadership and a common vision; a data-driven culture; stakeholder engagement; and a supportive policy environment. After

describing each, a rubric is used to assess district readiness for each category (early to advanced), broken into sub-categories.
http://www.gatesfoundation.org/united-states/Documents/empowering-effective-teachers-readiness-for-reform.pdf

Empowering effective teachers: Strategies for implementing reforms – Gates Foundation (2010)
This work contained a summary of the above work and feedback from participating sites. It tended to be largely repetitive of the work above – using multiple measures, using evaluation to inform PD, making tenure meaningful, differentiating compensation (especially in priority placements) and implementing a career ladder – but is based on feedback from research sites.
http://www.gatesfoundation.org/united-states/Documents/empowering-effective-teachers-empowering-strategy.pdf

Error rates in measuring teacher and school performance based on student test score gains – Peter Schochet and Henley Chiang – Mathematica Policy Research (2010)
The research in this paper (which is rather technical in nature) addressed the concern of 'false negatives', or teachers who are average being identified as needing assistance, if only a value-added evaluation system is in place. The authors found that if three years of data were used, it is conceivable that a quarter of all teachers could receive false negatives. They suggested that multiple years of data be used. In addition, the authors stated that while value-added measures were noisy, they do have advantages – they are strong predictors of subsequent-year outcomes and are less susceptible to manipulation. The overall conclusion was to use multiple measures, and possibly multiple phases of evaluation, so that if someone is misclassified, subsequent measures would draw this out.
http://ies.ed.gov/ncee/pubs/20104004/pdf/20104004.pdf

Evaluating teachers: The Important role of value-added – Glazerman, Loeb, Goldhaber, Staiger, Raudenbush, & Whitehurst (Brookings) (2010)
This paper addresses the debate of when and how to include value-added information in teacher evaluation systems. The authors recommend including value-added measures, but not as the sole measures, especially when personnel decisions are connected to evaluations. In addition, the authors confront recent reports addressing the misclassification of teachers using value-added models by stating all systems have inherent classification error, and that the goal should be to minimize error (by using multiple measures, and multiple years of data). This is an accessible paper, in that the analysis is not overly-technical, and outlines some of the key issues; particularly, the notion that a combination of good, not perfect, performance measures is an objective worth striving toward in designing evaluation systems.
http://www.brookings.edu/~/media/Files/rc/reports/2010/1117_evaluating_teachers/1117_evaluating_teachers.pdf

Evaluating teaching with multiple measures – Goe (AFT)
This work provides an overview of the use of multiple measures in teacher evaluation systems. It is slightly repetitive of other works in the bibliography, but still reinforces the essential point that classroom observations and value-added scores on their own provide incomplete pictures of the complexity of teaching. When discussing student learning, the author also stressed that multiple

measures be used, and that decision-makers question what type of information may be obtained from different forms of student assessment.
http://www.atfunion.org/files/dmfile/EvaluatingTeachingwtihMultipleMeasures.pdf

The Evaluation of principals: What and how do states and districts assess leadership? – Goldring, Cravens, Murphy, Elliott, Carson, and Porter (2008)
The authors describe the significant challenges of leadership assessment; namely, *what* to assess (given the complexity of a principal's role) and *how* to assess (appropriate methods to make inferences on performance). The authors reviewed a variety of assessment instruments in terms of content and usage, and concluded: districts focus on a variety of areas; instruments tended to have limited focus on curriculum, instruction, connections to community, and specific accountability measures; ensuring a culture of learning and professional behavior received the most emphasis; and there is little consistency in how instruments are developed nor attention to validity and reliability.
http://peabody.vanderbilt.edu/Documents/pdf/LSI/AERA_EvaluationPrincipals.pdf

Examining district guidance to schools on teacher evaluation policies in the Midwest Region – REL-Midwest (2007)
This was a background study, to provide the context for future research. The authors analyzed district policy documents, and found that procedure was emphasized over content and assistance. Other findings included: evaluations tended to focus on beginning teachers; little to no guidance on consequences were provided; little guidance on the training of evaluators; and vague terminology. With the research literature on evaluation being thin, this study attempted to begin to fill that void.
http://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/REL_2007030.pdf

Glossary of teacher effectiveness and evaluation terms (Powerpoint) – McKinsey & Company (2010)
In addition to proposed definitions, this slideshow includes graphic conceptual maps and an example from Pittsburgh.

How leadership influences student learning – Kenneth Leithwood, Karen Seashore Louis, and Kyla Wahlstrom (2004) – Wallace Foundation
This work was a comprehensive review of research addressing leadership effects on student learning. Especially in 'challenging' schools, the effect of school leadership was significant and under-estimated. Discussed various 'fad' titles given to school leaders, and argued that some core characteristics (namely, setting a direction and influencing members to move in that direction) were more important that titles, which were often representative of different approaches rather than different skill sets. There was a section entitled 'The basics of successfully leadership' that could be a useful starting point.
http://www.wallacefoundation.org/SiteCollectionDocuments/WF/Knowledge%20Center/Attachments/PDF/ReviewofResearch-LearningFromLeadership.pdf

How should states define teacher effectiveness? (Powerpoint) - Bryan Hassel, Public Impact (2009)

The slideshow attempted to articulate 'teacher effect' as a concept by defining 'effectiveness' and 'reach' as essential components. This approaches the topic of 'effectiveness' without getting into how to measure student outcomes – there is an emphasis on including teacher behaviors. In addition, the discussion on 'reach' reinforces a career ladder and begins the conversation on the state role.
http://www.publicimpact.com/publications/PublicImpact-How_Should_States_Define_Teacher_Effectiveness.pdf

Human capital in Boston Public Schools: Rethinking how to attract, develop, and retain effective teachers – Nat'l Center on Teacher Quality (2010)
This report is a comprehensive look at the human capital policies and practices in BPS. The analysis was done using interviews, policy documents, and personnel data. A section on 'developing an effective teaching corps' covers induction, evaluation, and tenure. The report analyzed BPS's current evaluation instrument and its shortcomings, in addition to problems associated with the training and support needed by principals to complete quality evaluations. While the whole report is very broad, the section on developing an effective teaching corps may, on its own, be useful.
http://jamesliou.com/liouwp/wp-content/uploads/2010/02/nctq_boston_human_capital.pdf

White Paper - Human capital reform in education: Teacher evaluation and compensation – Rennie Center (2010)
This work provides an overview of the Massachusetts context of teacher evaluation, including a brief history, and challenges of the current system. In addition, there is an overview of evaluation tools, including strengths and weaknesses of each. 'Promising models' are described, including career differentiation, group evaluation, peer review, career ladders, and TAP. The second half of the work focuses on performance pay.

Human capital: Unions and school districts collaborating to close achievement gaps – NEA Foundation (2010)
This paper describes work done in pilot districts through the Closing the Achievement Gaps Initiative, and attempts to connect that work to human capital initiatives. Lessons learned from pilots include a sharing of best practices through networks and collaborating with community partners. The paper concludes with a discussion of new roles for unions, including with evaluation.
http://www.neafoundation.org/downloads/HumanCapital.pdf

Identifying effective classroom practices using student achievement data – Kane, Taylor, Tyler, and Wooten (2010)
The authors used data from the Cincinnati school system and found a positive relationship between TES scores and student achievement growth. The analysis also differentiated by TES rating category and found that higher ratings (from Basic to Proficient or from Proficient to Distinguished) was associated with 1/5-1/6 of a standard deviation of student achievement gains, and that certain subsets of teaching practice ("classroom environment" management and "questioning and discussion") may also yield higher achievement, although policies focusing on helping teachers improve on all 8 major categories of skills was emphasized. The authors

concluded that multiple measures of teacher effectiveness may be more predictive of future student achievement than any single measure.
http://www.nber.org/papers/w15803.pdf?new_window=1

Identifying effective teachers using performance on the job – Robert Gordon, Thomas Kane, and Douglas Staiger (The Brookings Institution) (2006)
This work examined the shift from an input-based teacher effectiveness paradigm to one focused on constant evaluation and improvement. There were five recommendations: remove the barriers to enter teaching; make tenure more difficult to achieve; provide bonuses for working in hard to staff schools; evaluate using multiple measures; and link student performance with teachers. This presented the argument of why the current input-centered system is not working.
http://www.brookings.edu/~/media/Files/rc/papers/2006/04education_gordon/200604hamilton_1.pdf

Improving instruction through effective teacher evaluation: Options for states and districts – Carrie Mathers and Michelle Oliva, National Comprehensive Center for Teacher Quality (2008)
This work provided an overview of the purpose of evaluation and research findings from Midwest schools. Then, there is an extensive description of evaluation tools (lesson plans, observations, self-assessments, portfolio, student achievement data, student work samples) with strengths and weaknesses of each. There is also a description of various process considerations in evaluation (who, how often, training, communication) with recommendations for each. There are sections with state and local policy options. This is a very comprehensive design overview; the section that is light is on using evaluations.
http://www.tqsource.org/publications/February2008Brief.pdf

Improving teacher evaluation to improve teaching quality – Liam Goldrick, NGA Center for best practices (2002)
The brief provided an overview of what policymakers should do to improve evaluation, including: defining quality; focusing on improvement; training evaluators; etc. Examples from different states are included as each topic is explored in greater detail.
http://www.nga.org/cda/files/1202IMPROVINGTEACHEVAL.pdf

Including student results in teacher evaluation – a case study in focus on teaching expertise – Research for Better Teaching, Inc. & TEACHERS21 (2005)
This essay begins with the question "Should teachers be accountable for student results?" and presumes an answer of 'Yes'. The authors outline their argument, that standardized tests should never constitute a sole judgment about teaching quality, but that various data sources should provoke questions and initiate dialogue about teaching performance. Their second thesis is that the larger thrust of improving teaching quality is developing shared responsibility for student learning. Montgomery County is highlighted as an example of how to use student data. Part I concludes with sample evaluations. Then, in Part II, the focus is on using a multiple-year evaluation cycle. In addition, training is also discussed.

Incorporating use of a performance continuum in teacher evaluation systems – Hanover Research Council (2010)

This report was the result of a survey of 21 state departments of education about challenges in the design and implementation of summative evaluations. The bulk of the report listed evaluation designs from CT and the New Teacher Project (which has differentiated levels for teachers). There were a number of 'In Practice' sections, describing how various California and Connecticut districts implemented their evaluation system, with an eye of the teacher career continuum. In addition, there was a section about rating scales used in AZ, DE, and NC. This latter work also included administrator evaluations. This is a clear, useful resource in the design and implementation stages.

Information and employee evaluation: Evidence from a randomized intervention in public schools –Rockoff, Stainer, Kane, and Taylor (2010)
This study addressed the appropriateness of using student performance data in teacher evaluations. The research was based in NYC, and sought to determine the use of subjective evaluations and the possible influence that objective data on the subjective aspects of evaluation. The authors found that when principals in the experimental group had access to teachers' value-added estimates, their subjective evaluations of teacher performance were significantly higher as well (more so in math than English), while in the control group this was not so. The authors concluded that student performance data may provide valuable information in evaluations, and speculated that the privacy of such objective data may result in low-performing teachers moving on to other schools, and that principals in the new schools could find value in such data as well.
http://papers.nber.org/papers/w16240

Investigating the links to improved student learning - Wahlstrom, Seashore Louis, Leithwood, and Anderson (2010) – Wallace Foundation
This study included survey, interviews, and classroom observations and attempted to analyze different forms of leadership and their links to student performance. This was a very large study, encompassing a broad range of topics, including usage of data, the role of parents as leadership partners, school context, the affective side of leadership, and the supports needed in schools. In the conclusion, the authors outlined three concepts that were consistent across findings: Expectations and accountability (at numerous levels); Efficacy and Support; and Engagement and stakeholder influences (making real connections with others).
http://www.wallacefoundation.org/KnowledgeCenter/KnowledgeTopics/CurrentAreasofFocus/EducationLeadership/Documents/Learning-from-Leadership-Investigating-Links-Final-Report.pdf

Leadership for learning: A Research-based model and taxonomy of behaviors – Murphy, Elliott, Goldring, and Porter (2007)
In this article, the authors outline a learning centered leadership framework and reviewed the relevant research literature to determine eight major dimensions of behavior within this model: vision for learning; instructional program; curricular program; assessment program; communities of learning; resource acquisition and use; organization culture; and social advocacy. This is a comprehensive review of the literature, placed into a framework to better understand how the various dynamics interact.

Learning about teaching: Initial findings from the Measures of Effective Teaching project – Gates Foundation (2010)

This is a progress report from the MET Project. The assumptions behind the project are outlined (when feasible, a teacher's evaluation should include achievement gains; additional measures should be related to achievement gains; and measures should include feedback on practice to support professional growth and development), and the measures are described (achievement gains, observations, reflections, content knowledge, and perceptual data [student and teacher]). Preliminary findings indicate: past VAM is among the strongest predictors of achievement gains; teachers with high VAMs tend to promote deep conceptual understanding; and student perceptions suggest that students recognize effective teaching (particularly as related to classroom management and challenging students with rigorous work). Details are provided regarding the treatment of VAM and student perception in their analysis (which follows), as well as a description of the sample.
http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf

Learning from Ohio's best teachers: A Homegrown model to improve our schools – Piet Van Lier, Policy Matter Ohio (2008)
The focus of this work was on Peer assistance and Review (PAR) models being used in Ohio. The author described components of successful PAR programs and the benefits in terms of retention, improved student achievement, and improved union-management relations. The work also included a research review, common arguments against PAR, and provides cost estimates. If a PAR model is to be used at all, this is a good resource in terms of background and practical implementation issues.
http://www.policymattersohio.org/pdf/PAR2008.pdf

The Link between teacher quality and student outcomes: A Research synthesis. Laura Goe (ETS), National Comprehensive Center for Teacher Quality (2007)
The main foci of this synthesis are: What is teacher quality, why is it important, and how can it be measured? Do certain aspects of quality matter more than others, and how important is experience? In general, the finding was that studies (focusing mostly on research since 2000) were inconclusive, and while many studies found significance in terms of various features of teacher quality, there was little practical significance demonstrated. In terms of experience, the first four or five years are very important, but then performance tended to level off. There was a review of previous research syntheses and a useful graphic (p. 9) of a framework for teacher quality, accounting for inputs, processes, and outcomes. This is a useful reference guide, as it groups research into practical categories and provides brief summaries.
http://www.tqsource.org/publications/LinkBetweenTQandStudentOutcomes.pdf

Measure for measure: The Relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores – Grossman, Loeb, Cohen, Hammerness, Wyckoff, Boyd, & Lankford (CALDER) (2010)
The authors of this study sought to determine those instructional practices that have the greatest impact on student performance and the extent to which value-added scores signal differences in instructional quality. The classroom observation tool used was a combination of CLASS and an observation protocol developed specifically for English language arts (PLATO). Teachers with higher value-added scores tended to score higher on PLATO elements, most notably the Explicit Strategy Instruction element. In addition, teachers with higher value-added scores tended to provide models of desired student work and scaffolded instruction.

Measuring teacher and leader performance: Cross-sector lessons for excellent evaluations – Kowal and Hassel (Public Impact) (2010)
This report summarizes steps to take in designing evaluation systems, as a result of cross-sector (government, nonprofit, for-profit) analyses. Recommendations include stating the purpose of evaluation, aligning objectives with organization mission, designing performance measures and standards that align with expectations, be clear about the process, and use results to drive actions. There are useful design specifications (including the "how" and "what" of evaluation) and procedural advice (how to design performance measures that are responsive to goals and objectives), as well as private sector examples throughout the paper. The final section discusses how to apply the findings to performance measures for educators. This is a useful primer on some of the essential issues of measurement currently being discussed.
http://www.publicimpact.com/images/stories/performance_measurement_2010.pdf

Measuring teacher effectiveness (Powerpoint) – Laura Goe (ETS), National Comprehensive Center for Teacher Quality (2009)
This slideshow explored common ways to measure effectiveness, and suggested a 5-point definition of effectiveness as a starting point. There was a discussion of the limitations of value-added models *on their own*, which connected to a discussion of other measures. In addition, there was a thorough discussion of validity issues regarding observation instruments and 'design consideration'.

Measuring the effectiveness of human capital program investments (Powerpoint) - Larry Stanton, Consortium for Educational Change (2010)
This presentation reviewed four interconnected elements to managing human capital: defining effectiveness; tracking progress and creating transparency; learning and making informed decisions; and establishing rewards, consequences. Some very basic, but useful, definitions of effectiveness are offered, as well as a consideration of various metrics.

Measuring principal performance: How rigorous are commonly used principal performance assessment instruments? Christopher Condon and Matthew Clifford (2010). Learning Point Associates.
This work examined a few assessments (including: Change facilitator style questionnaire; Leadership practices inventory; VAL-ED; and others) in validity and reliability terms. There was a helpful, accessible table summarizing the findings.
http://www.learningpt.org/pdfs/QSLBrief2.pdf

Methods of evaluating teacher effectiveness – Laura Goe and Andrew Croft (ETS), National Comprehensive Center for Teacher Quality (2009)
This document provided a comparison of value-added measures and classroom observations, noting advantages and disadvantages of each. Then, the authors considered surveys and 'administrative judgment' as additional methods, and consider potential purposes these four methods may serve. Finally, there were suggestions for creating a strong system, as well as links to other states' guidelines and/or components of systems (DE, FL, MN Q-Comp, NM, NC, OH, SC, TN, and WI).

http://www.ped.state.nm.us/RTTT/part2/D2-RestoPractice_EvaluatingTeacherEffectiveness-%20Resource.pdf

A New conceptual framework for analyzing the costs of performance assessment - Lawrence Picus, Frank Adamson, William Montague, and Margaret Owens (SCOPE) (2010). Like the other SCOPE work, this focused on student assessment. As a reference, the table on pp. 14-20 may be useful, as it analyzed the costs, expenditures, and benefits of performance assessment for formative and summative purposes. There were two additional works in this series (not shown here), that also focused on student performance assessments. http://edpolicy.stanford.edu/pages/pubs/pub_docs/assessment/scope_pa_picus.pdf

Opportunity at the top: How America's best teachers could close the gaps, raise the bar, and keep our nation great – Bryan and Emily Hassel, Public Impact (2010) This work argued that current policies fail to take advantage of effective teachers (top-quartile). There was a comparison of the boldest current initiatives versus alternatives that could capitalize on existing talent in schools. The authors argued that retention and dismissal alone would not reach the same number of students as building on effectiveness would. This was part of a series on creating an 'opportunity culture' in education. While this piece did not specifically refer to evaluation, it might be helpful when considering the use of evaluations. http://opportunityculture.org/images/stories/opportunity_report_web.pdf

Overview of existing teacher evaluation models - Hope Street Group (2010) This work provided some very general background information on the purpose and elements of teacher evaluation systems. In addition, the Danielson Framework, TAP, Montgomery County's Peer Advancement and Review Program, and Teach for America's framework are analyzed – for each, there was an overview, an 'in-use' section, and analysis. In addition, similar analyses of CLASS, MQI, and PLATO were conducted.

The Policy uses and "policy validity" of value-added and other teacher quality measures. Douglas Harris, TQR (2007). This work addressed the validity of value-added measures, and argued that multiple measures should be used, but that value-added measures have their purpose. This work mentioned using value-added to identify effectiveness (not mentioned in the 2009 work by Harris). The paper introduced "policy validity" as a concept, and examined the use of multiple measures as a way for 'signaling' or 'improvement' – the assumption is that each measures would have an explicit purpose, for formative and/or summative evaluations.

A Practical guide to evaluating teacher effectiveness – Olivia Little, Laura Goe, and Courtney Bell (ETS), National Comprehensive Center for Teacher Quality (2009) This document is a comprehensive overview of research on teacher effectiveness. The challenge of defining effectiveness is examined, and the 5-point definition is provided. The majority of this work is an in-depth analysis of various methods (value-added; classroom observation; principal evaluation; classroom artifacts; portfolios; self-report; and student evaluation), with each section containing a definition, research synthesis, examples, and considerations. Finally, there were recommendations for designing a system, and a table that examines methods of evaluation with

various purposes of evaluation. On its own, the table (or Appendix C, which summarizes the methods in a table) is a potentially-useful document, but overall this is a very thorough analysis. http://www.tqsource.org/publications/practicalGuide.pdf

Principal effectiveness: A new principalship to drive student achievement, teacher effectiveness, and school turnarounds - New Leaders for New Schools (2009)
This overview proposed using a three-pronged definition of principal effectiveness, based on: student outcomes; teacher effectiveness; and leadership actions. There was a description of the Urban Excellence Framework, a discussion of the principal as human capital manager, and a series of policy recommendations at the state and local levels. This is written in the context of turnaround schools, but some sections could be useful for any school system. http://www.nlns.org/documents/uef/principal_effectiveness_nlns.pdf

Principal effectiveness and leadership in an era of accountability: What research says – Jennifer King Rice (2010) – CALDER
This was a very brief review of the research literature. The major finding was that a school principal's quality affected a range of outcomes, and the quality of a principal was important to teachers. This is a very quick read. http://www.urban.org/uploadedpdf/1001370_principal_effectiveness.pdf

The Principal's role in human capital management: A Missing piece of the principal evaluation puzzle (Powerpoint) – Tony Milanowski (2010)
This was a very brief overview of human capital management. If this is to be considered as an element to principal evaluation, this could be used as a starting point.

Principles of teacher evaluation design (Powerpoint) – The New Teacher Project (2010)
This slideshow presented an overview of the current issues in teacher evaluation, human capital reform, and the use of multiple measures.

Problems with the use of student test scores to evaluate teachers – Baker, Barton, Darling-Hammond, Haertel, Ladd, Linn, Ravitch, Rothstein, Shavelson, and Shepard (EPI Briefing Paper) (2010)
This paper reviews the recent trend of states' assigning a great amount of weight to value-added measures in teacher evaluations. The authors argued that such a trend is unwise and may lead to negative consequences when these measures are tied to high-stakes decisions. The assumption throughout the work is that VAMs such as that employed in Tennessee (which is different from the model used in MA) is the focus of analysis. Nonetheless, the overall conclusion reinforces that of numerous other authors featured in the annotated bibliography: VAMs may provide some useful information, but should be used cautiously, and multiple measures should be involved in any high-stakes evaluation framework.

The Real Value of Teachers: Using New Information about Teacher Effectiveness to Close the Achievement Gap - Kevin Carey, The Education Trust (2004)
This paper made the case for the "real value" of teachers and described the evidence of the impact of individual teachers on student achievement from various districts and states. The author argued that although the data demonstrated the power of individual teachers to produce

gains in student learning, teacher policies do not reflect that understanding. The report laid out an ambitious policy agenda, premised on a review of the existing research on teacher effectiveness—often referred to as "value-added."
http://www.edtrust.org/dc/publication/the-real-value-of-teachers-using-new-information-about-teacher-effectiveness-to-close

Policy Brief: Recognizing and developing effective teaching: What policy makers should know and do. – Linda Darling-Hammond (2010)
Briefly reviewed the qualities associated with productive teachers (verbal ability, content knowledge, understanding of learning, etc.), and research on the benefits of National Board Certification. A discussion of the use of performance assessments focused on licensure, and advanced National Board Certification to a greater number of teachers. The authors briefly touched on different forms of evidence that may be included in teacher evaluation, and also addressed some limitations of value-added measures.
http://www.aacte.org/pdf/Publications/Reports_Studies/Effective%20Teaching%20-%20Linda%20Darling-Hammond.pdf

Review of teaching performance assessments for use in human capital management – Anthony Milanowski, Herbert Heneman, and Steven Kimball - SMHC (2009)
This study reviewed various teacher assessments, including Praxis III, PACT, Formative Assessment System Continuum (UC – Santa Cruz), Framework for Teaching, TAP, NBPTS, and CLASS. The author sought to make comparisons and examined past reliability and validity research. Comparisons were made of underlying competency-model content and assessment processes. Following the analysis, there were recommendations and a 'roadmap' for designing teacher assessments. The overall theme of this work was to connect performance assessment to a strategic human capital approach. While this work is long, the language is not overly-technical, and this is a very thorough examination of common assessments.
http://www.eric.ed.gov/PDFS/ED506953.pdf

Rush to judgment: Teacher evaluation in public education – Thomas Toch and Robert Rothman, Education Sector (2008)
This work began with a brief overview of the issues with current 'drive-by' evaluation systems, and suggested a model for comprehensive performance-based systems. The authors showed connections and differences between commonly referred-to models such as Danielson, TAP, BEST, National Board, and Praxis III. Peer review was also discussed, and the authors examined systems designed to 'weed our bad apples' versus improvement-based models (such as Toledo, Cincinnati, and CT). There was a brief discussion of performance pay, and a section on teacher unions' influence on evaluation systems. This is a comprehensive analysis of the pertinent challenges, and the appendix provides brief descriptions of all systems, categorized into their usages.
http://www.educationsector.org/usr_doc/RushToJudgment_ES_Jan08.pdf

Searching for effective teachers with imperfect information –Staiger and Rockoff (2010)
The authors examined the research on five typical statements regarding teacher effectiveness, including: productivity based on gains is heterogeneous; estimates of teacher effect are noisy; there is substantial improvement in the first few years; costs associated with turnover relates to

reduction in achievement and not direct hiring costs; and school leaders have little ability to select effective teachers during hiring. Next, the authors made suggestions about hiring practices and tenure decisions, based on the research review. This does not directly address evaluation, but it does touch on using evaluation for personnel decisions. The second link is a later issue of the paper, and includes a section on '5 facts about teacher effectiveness'.
http://mfi.uchicago.edu/publications/papers/searching-for-effective-teachers-with-imperfect-information.pdf
http://www.nctq.org/docs/Firing_Teachers_8046.pdf

So long, Lake Wobegon? Using teacher evaluation to raise teacher quality - Morgaen Donaldson, Center for American Progress (2009)
This work was a comprehensive review of evaluation practices, including the state of evaluations and generally-used practices and tools. In addition, the report described how changes in the education workforce and technologies were laying the groundwork for meaningful evaluation change. Then, there was an analysis of Cincinnati's evaluation system – the challenges of evaluation systems were detailed one-by-one, and each was 'answered' through TES. Using Cincinnati as a model for reform, the authors provided recommendations, similar to those found in many other reports in this bibliography.
http://www.americanprogress.org/issues/2009/06/pdf/teacher_evaluation.pdf

A State system for assessing teaching practice, performance, and effectiveness – Allan Odden, Strategic Management of Human Capital (2009)
This paper focused on the development and use of a system to assess teacher performance, extending the work to a hypothetical licensure system. The general themes could be examined without the licensure link – however, the paper does tend to focus at the state level rather than local implementation. The proposed system focuses on career milestones, more so than annual evaluations. There are design principles that could apply to the work of the Task Force, but in general this work would be more appropriate for the development of 'career milestone' assessments.

Subjective and objective evaluations of teacher effectiveness –Rockoff and Speroni (2010)
The purpose of this study was to measure the extent to which evaluation of new teachers in NYC could predict future impact on student performance. The authors found that higher subjective evaluations (observations) prior to hire or during first year of teaching correlated with future student performance. They concluded that objective and subjective measures have power in evaluation systems, especially if evaluators were sufficiently trained on observation protocol, to reduce the variation between evaluators.
http://www0.gsb.columbia.edu/faculty/jrockoff/papers/rockoff_speroni_subjective_evals_AEA_PP_final.pdf

Supporting effective teaching through teacher evaluation: A Study of teacher evaluation in five charter schools –Donaldson and Peske, Center for American Progress (2010)
The charter schools in this work reported annual evaluations, used evaluations for continuous improvement, and incorporated student performance data (although not value-added), so there is a parallel to the work of the Task Force. There was a section on the challenges faced by schools.

In general, this appears to be of most use when discussing how to use evaluations to inform decision-making.
http://www.americanprogress.org/issues/2010/03/pdf/teacher_evaluation.pdf

Taking the lead: With peer assistance and review, the teaching profession can be in teachers' hands – Jennifer Goldstein, *American Educator* (2008)
Describes the experience of a novice teacher entering a Compton elementary school, including the support she received on an informal basis from a colleague. The article then described the PAR system being used in OH and CA, and the possibility that such a system may provide the needed support for new teachers and lessen the burden on principals to be the sole evaluators in schools. The article describes differences between PAR and typical evaluations, and reviews typical problems and solutions associated with PAR. An interview with 3 people associated with PAR is also included.
http://www.aft.org/pdfs/americaneducator/fall2008/goldstein.pdf
http://www.aft.org/pdfs/americaneducator/fall2008/par.pdf

Teacher evaluation as a policy target for improved student learning: A Fifty-State review of statute and regulatory action since NCLB –Hazi and Arrendondo Rucinski (2009)
Using archival records, the authors analyzed statutes and regulations concerning teacher evaluation, including: state oversight; changes in policy; and data changes since NCLB. This was a constructive overview of various strategies that are being adopted in other states, including summary tables. While this was in the context of NCLB, it is still relevant and useful in that it provided a state-by-state comparison (although, following RTTT, some of this information will not longer be accurate).
http://epaa.asu.edu/ojs/article/viewFile/7/7

Teacher evaluation: New approaches for a new decade – Jennifer Dounay Zinth, Education Commission of the States (2010)
This work contains a thorough examination of using student achievement data in teacher evaluations, including various ways of including/excluding that data, as well as examples of state policies. This is useful in that it provides examples of what other states are doing – not only in terms of including student achievement data, but also training evaluators, supporting teachers, and recognizing effective teachers.
http://www.ecs.org/clearinghouse/86/21/8621.pdf

Teacher evaluation systems: The Window for opportunity and reform – Little (NEA) (2009)
This paper reviewed five evaluation systems (TAP, FFT, ProComp, PAR, and BEST), including the relation to student outcomes and how each was 'received' by teachers and administrators. The following characteristics were noted of successful systems: credibility and meaningful (involve multiple stakeholders; valid; use multiple measures); and linked and integrated (include embedded PD; opportunities for career advancement [includes a discussion of performance pay, that it be linked to valid measurements and that performance pay include additional opportunities for additional compensation such as additional responsibilities and working in hard-to-staff schools]). The author reinforces that multiple measures be used, that teachers be involved in deciding how to incorporate student performance data, that value-added be used cautiously and

not beyond its capabilities as a signaling tool, and that multiple outcomes beyond student learning be incorporated into any system.

Teachers, schools, and academic achievement - Rivkin, Hanushek, and Kain (2005). This was a technical analysis of variation is student performance. The main conclusion was that teachers and schools had large effects on student achievement, "in a way that rules out the possibility that the observed differences are driven by family factors" (p. 449). Additional conclusions: no evidence that a master's degree improves teaching skills; gains in the first three years of teaching, with little evident beyond; class size had significant effect on math/reading growth, especially in earlier grades. The authors framed their conclusions in terms of whether or not certain investments were worthwhile, and supported the conclusion that supports are a greater investment than tightening standards.
http://www.economics.harvard.edu/faculty/staiger/files/HanushekRivkinKain%2BEcta%2B2005.pdf

Teacher value-added and credentials as tools for school improvement (Powerpoint) – Douglas Harris, UWisconsin (2009)
This is an analysis of research addressing credentials versus value-added as they pertain to student achievement.
http://www.tqsource.org/webcasts/evaluateEffectiveness/Harris_VAM_WC.pdf

Teaching matters: Strengthening teacher evaluation in Massachusetts – MassPartners (2004)
This work summarized research on evaluation systems and provided a list of recommendations, including new regulations and professional standards; using multiple data sources; providing adequate time; and training. There was also a description (more detailed than other works) on differentiating evaluation by career stage. In addition, results from a MA survey were reported, which set the local context. Appendix A also contains a useful graphic organizer for teachers needing improvement.
http://www.mespa.org/news/Teaching_Matters.pdf

Using competency-based evaluation to drive teacher excellence: Lessons from Singapore – Lucy Steiner, Public Impact (2010)
Nearly a decade ago, Singapore revised its evaluation system to be performance-based, and the country is reluctant to reveal many of the 'secrets' of their system. But the author determined that one clear element was the use of performance competencies. Singapore shifted from a system that relied on observable characteristics to emphasize these underlying competencies (patterns of thinking, feeling, acting, or speaking). The shift was based on a simple research process – identify high- and average-performers based on a universal measure, then conduct a structured Behavior Event Interview with both groups, and code the interviews for patterns. There are graphics and tables detailing Singapore's competencies, and descriptions of the evaluation process. In addition, the career ladder is described. This is a fascinating, fast read, and could be useful if the group is willing to consider models outside the common US models.
http://opportunityculture.org/images/stories/singapore_lessons_2010.pdf

Using data about classroom practice and student work to improve professional development for educators – NEA Foundation (2003)
This paper used a broader concept of 'data' to include videos of classroom performance. This is a novel approach to some of the questions we will be attempting to answer, as it could relate to observational protocols, an additional measure of performance, using evaluation to influence PD, and how to disseminate best practices. The work suggested a balance of data sources to inform PD through the 'inquiry cycle'.
http://www.neafoundation.org/downloads/NEA-Using_Date_Classroom_Practice.pdf

Using open innovation to reform teacher evaluation systems – Hope Street Group (2010)
Contains recommendations to improve teacher evaluation systems, including: objective measures, clearly-defined standards; supportive administrators; trained evaluators; and information that is comparable across schools and districts. Also contains a thorough overview of the problem and the results of an online collaboration tool that connected practitioners across the country.
http://www.hopestreetgroup.org/content/images/stories/documents/policy2.0policypaper.pdf

Using student performance data to identify effective classroom practices. John Tyler, Eric Taylor, Thomas Kane, and Amy Wooden (2009)
This study used data from Cincinnati schools, including student performance data and observations. Classroom management and instructional skills measures by TES predicted student growth. The authors claim that a core of 'Distinguished' teachers could close the achievement gap in 5-6 years relative to 'Proficient' teachers. This report is slightly technical in nature, but not long.
http://www.econ.brown.edu/econ/events/Tyler%20teachers.pdf

The Value of value-added data - Craig Jerald, The Education Trust (2009).
In this paper, the author described the ways in which access to value-added data could assist teachers in both assessing student performance and in strengthening their own performance. The author included examples from districts and states around the country where teachers are using value-added data, as well as specific ways in which having access to the data assisted them in their work.
http://www.edtrust.org/sites/edtrust.org/files/publications/files/ValueAdded_0.pdf

When the stakes are high, can we rely on value-added? Exploring the use of value-added models to inform teacher workforce decisions – Goldhaber, Center for American Progress (2010)
In this report, the controversy of using VAM in the context of low-, medium-, and high-stakes decisions is analyzed. The author contends that there are potential issues with using VAM on its own, but argues that similar risks of misclassification exist for any measure used in educator evaluation, and that it is not a reason to exclude such measures. VAM measures, according to Goldhaber, allow for differentiation of educator performance and there is a review of evidence that points to the utility of such measures, to be used as "honest brokers" in performance evaluation. Parts of the report are technical in nature, reviewing error rates and the predictive quality of VAMs, but the ongoing review of the literature makes the report readable and the results worth considering.

http://www.americanprogress.org/issues/2010/12/pdf/vam.pdf

The Widget effect: Our national failure to acknowledge and act on difference in teacher effectiveness – The New Teacher Project (2009)
This report has been cited numerous times by various groups in the past year and has been influential in bringing educator evaluation to the front of the reform agenda. Examining 12 districts in 4 states, the authors found that evaluation systems failed to provide feedback on teacher performance. Other notable findings: less than 1% teachers received unsatisfactory ratings; 73% teachers said most recent evaluation did not identify areas for development; there was no system to identify and promote the most effective teachers; and large percentages of teachers and administrators said they knew at least one tenured teacher who was performing poorly (higher numbers in high-poverty schools).
http://widgeteffect.org/downloads/TheWidgetEffect.pdf

Working with teachers to develop fair and reliable measures of effective teaching – Gates Foundation (2010)
This described the purpose of the project – to develop fair and reliable systems of observation that may be used for different purposes. There were useful and simple graphics that illustrated the current system of evaluation and the working theory that the project sought to test. The project proposed 5 measures: student gain scores, observations, pedagogical content knowledge, students' perceptions of the classroom, and teachers' perceptions of working conditions. Preliminary findings are due to be published in Fall, 2010.

Would accountability based on teacher value-added be smart policy? An Examination of the statistical properties and policy alternatives. Douglas Harris, TQR (2009)
This research compares teacher value-added to other policy alternatives (credentials, school value-added, and formative use of test data), and considers the validity, costs, and purposes of value-added policies. There is a significant section on the statistical validity of teacher value-added measures, including the assumptions behind such models. The author finds that the assumptions are violated, but that the violations are not so severe to preclude the use of value-added measures, because they correlate highly with other measures of effectiveness. This provides some cautions for using *only* value-added measures, and provides a research base. This report is a bit technical in nature, but could be considered complementary to the slideshow by Harris (above).
http://www.wceruw.org/news/events/VAM%20Conference%20Final%20Papers/VAMandPolicy_DHarris.pdf

Appendix E – Presenters to the Task Force

Dr. Elizabeth Arons, Senior Human Resources Policy Advisor, New York City Department of Education
Performance Evaluation: A Cornerstone of Human Capital Development
Presented September 7, 2010

Dr. Susan Moore Johnson, Pforzheimer Professor of Teaching and Learning
Harvard Graduate School of Education
Building School Capacity Through Teacher Evaluations
Presented September 7, 2010

Bob Lee & Matt Deninger, Department of Elementary & Secondary Education
Growth Model Presentation
September 22, 2010

Dr. Joseph Murphy, Frank W. Mayborn Chair of Education; Associate Dean for Special Projects
Peabody College, Vanderbilt University
Learning-Centered Leadership & *Principal Evaluations*
Presented October 4, 2010

Dr. Thomas J. Kane, Deputy Director and Professor of Education and Economics. Harvard University
Measures of Effective Teaching
Presented October 20, 2010

Beverly Miyares, Professional Development Specialist
Kathie Skinner, Director of Policy & Practice, Massachusetts Teacher Association
MTA Proposal for Educator Evaluation
November 5, 2010, Administrator Working Group

Dr. Jon Saphier, Founder of Research for Better Teaching, Inc. (RBT)
Perspectives on Evaluation
November 9, 2010, Teacher Working Group

Paul Toner, President, Massachusetts Teacher Association
MTA Proposal for Educator Evaluation
November 9, 2010, Teacher Working Group

Thomas Gosnell, President, American Federation of Teachers Massachusetts
AFT MA Work on Peer Assistance & Review
December 7, 2010

Seth Moeller, Director, Talent Management, Fidelity Investments
Deb Morsi, Nurse Manager, Baystate Health
Linda Noonan, Executive Director, MA Business Alliance for Education
Paula Squires, VP for Human Resources, Baystate Health
Private Sector Performance Review
January 4, 2011

Appendix F – Educator Evaluation Policy in Massachusetts

Well before the spur of Race to the Top the Massachusetts Board of Elementary and Secondary Education had identified the development of an effective, diverse, and culturally proficient educator workforce as a top strategic priority, and ESE had embarked on a path to develop a coherent and aligned set of state performance-based policy initiatives designed to achieve that goal.

Almost exactly a year before winning the Race to the Top award (September 2010), the Board of Elementary and Secondary Education (BESE) considered a series of core propositions to support educator effectiveness:

- Classroom teaching quality is the top school-based factor in raising student achievement.
- The quality of an education system cannot exceed the quality of its educational workforce.
- Massachusetts has many great teachers but *every* child deserves a great teacher *every* year.
- The proficiency gap and the teacher quality gap are related and must be addressed together.
- Effective instructional leadership is critical to developing conditions for effective teaching.[10]

In September 2009, BESE charged the Commissioner and ESE staff with creating an aligned, performance-based framework of educator policy at the state-level by:
- Developing better measures of educator effectiveness and using them to inform the key state policy levers of educator preparation program approval, licensure, induction, evaluation, professional development, and compensation reform;
- Promoting more effective and aligned human resource practices and policies at the district level; and,
- Providing more robust tools and resources to assist districts in attracting retaining, and supporting effective teachers and administrators.

Collectively, these initiatives have been designed to ensure that the Commonwealth's schools and classrooms are staffed with effective educators, its educator workforce is both diverse and culturally proficient, and that the Commonwealth's schools and districts are organized to support student achievement and success.

As a result, The state's Race to the Top initiatives build on a strong theory of action that includes the articulation of district standards and indicators, including eleven conditions for effective schools; targeted assistance and support to struggling schools and districts; and a comprehensive approach to educator development throughout the careers of teachers and leaders. RTTT includes a range of interconnected strategies at the state and district level to strengthen student learning, growth and achievement; develop improved systems and supports for teaching and learning; promote educator effectiveness; ensure career and college readiness; support struggling learners; and close proficiency gaps.

---

[10] September, 2009 Meeting of the BESE (http://www.doe.mass.edu/boe/minutes/09/0921spec_0922reg.pdf)

Appendix G – The Essential Conditions for School Effectiveness

http://www.doe.mass.edu/sda/review/school/process.html?section=essential

These 11 essential conditions are necessary conditions for schools to educate their students well; they guide the actions taken by both districts and the Department at all levels of the accountability and assistance system. While schools are responsible for developing the school level practices that ensure implementation of these essential conditions, schools need to be supported in these efforts by the policies and practices of their districts.

Districts are ultimately responsible for ensuring that these essential conditions are being implemented for all students in all schools. Districts at Level 3 of the system will be required to conduct a self-assessment following Department guidance to inform their improvement planning; this self-assessment will also be made available for use by districts at Levels 1 and 2. Districts at Levels 4 and 5 will be required to implement all of these conditions in their Level 4 or 5 schools or provide a compelling rationale for alternative approaches designed to achieve comparable or superior results. The commissioner will determine whether the rationale is sufficiently compelling to warrant an exception to any of the specific requirements of these essential conditions.

1. **Effective district systems for school support and intervention:** The district has systems and processes for anticipating and addressing school staffing, instructional, and operational needs in timely, efficient, and effective ways, especially for its lowest performing schools.
2. **Effective school leadership:** The district and school take action to attract, develop, and retain an effective school leadership team that obtains staff commitment to improving student learning and implements a clearly defined mission and set of goals.
3. **Aligned curriculum:** The school's taught curricula are aligned to state curriculum frameworks and the MCAS performance level descriptions, and are also aligned vertically between grades and horizontally across classrooms at the same grade level and across sections of the same course.
4. **Effective instruction:** Instructional practices are based on evidence from a body of high quality research and on high expectations for all students and include use of appropriate research-based reading and mathematics programs; the school staff has a common understanding of high-quality evidence-based instruction and a system for monitoring instructional practice.
5. **Student assessment:** The school uses a balanced system of formative and benchmark assessments.
6. **Principal's staffing authority:** The principal has the authority to make staffing decisions based on the School Improvement Plan and student needs, subject to district personnel policies, budgetary restrictions and the approval of the superintendent.
7. **Professional development and structures for collaboration:** Professional development for school staff includes both individually pursued activities and school-based, job-embedded approaches, such as instructional coaching. It also includes content-oriented learning. The school has structures for regular, frequent collaboration to improve

implementation of the curriculum and instructional practice. Professional development and structures for collaboration are evaluated for their effect on raising student achievement.

8. **Tiered instruction and adequate learning time:** The school schedule is designed to provide adequate learning time for all students in core subjects. For students not yet on track to proficiency in English language arts or mathematics, the school provides additional time and support for individualized instruction through tiered instruction, a data-driven approach to prevention, early detection, and support for students who experience learning or behavioral challenges, including but not limited to students with disabilities and English language learners.

9. **Students' social, emotional, and health needs:** The school creates a safe school environment and makes effective use of a system for addressing the social, emotional, and health needs of its students that reflects the behavioral health and public schools framework.

10. **Family-school engagement:** The school develops strong working relationships with families and appropriate community partners and providers in order to support students' academic progress and social and emotional well-being.

11. **Strategic use of resources and adequate budget authority:** The principal makes effective and strategic use of district and school resources and has sufficient budget authority to do so.

Appendix H – Glossary of Terms

Artifacts – items that demonstrate the completion of specific educator practices; products of an educator's work.

Educator Evaluation System – a complete approach to the evaluation of educators, including its purpose, the rules and regulations that apply, the target group to be evaluated, the domains to be covered, the procedures and methods to be employed, the instruments to be used, the persons to be involved, and the types of reports and feedback to be provided. Used to describe district-wide organization of evaluation.

Evaluator – a person who assembles data and information collected about an educator, analyzes them, makes judgments as to whether that educator's performance level meets the pre-specified standards, prepares a summary report, writes recommendations, and may provide feedback to the educator, directly or through another person.

Exemplary – consistently and significantly above proficiency on a standard or overall.

Feedback – the information and recommendations provided to an educator about his/her performance based on the results of that educator's evaluation and designed to help the educator improve his/her performance and make decisions concerning professional development and improvement.

Formative Assessment – the formal and informal processes an evaluator uses to gather evidence and provide the educator with feedback on how to improve practice. Data used to inform evaluation decision.

Framework – the State's regulatory language that set the parameters for educator evaluation. Districts use the framework in developing their local educator evaluation systems.

Goal – a statement of intent or an end that a person or a group strives to attain. In the proposed framework, goals include elements related to educator practice against standards and to improvement in student learning and growth outcomes.

Measurable – that which can be classified or estimated, in relation to a scale, rubric or standard.

Model System – an educator evaluation system that ESE is in the process of developing, for use in Level 4 and other school districts.

Needs Improvement – demonstrating lack of proficiency on a standard or overall.

Observation System – a category of measurement that includes: notes and judgments made during a series of observations, as well as artifacts of practice that support the judgments made relative to a series of observations

Peer Assistance and Review – a process by which exemplary teachers work with other teachers in order to increase performance; usually a joint venture of a district and a teachers union local.

Plan – the formal arrangement, discussed between an educator and evaluator, that sets professional goals and addresses areas of growth or improvement through the use of targeted resource allocation

Professional Teaching Status (PTS) – in its current format, an educator who has been granted a license by ESE and has served the public schools of a school district for three previous consecutive years, or less than three if granted PTS status by the Superintendent of that district.

Proficient – demonstrating the expected performance on a standard, or overall.

Rating – a judgment of the attainment of some attribute of teaching using a numerical or descriptive continuum.

Reflection – the process an educator undertakes in order to make preliminary judgments about their practice, relative to performance standards.

Rubric – a matrix that provides descriptions of attainment of multiple areas of knowledge or skills.

Self-assessment – the process of judging one's own teaching performance and outcomes for the purpose of self-improvement. An educator may use such techniques as self-viewing on videotape, observing and modeling exemplary educators, reflections, and analysis of student learning and growth outcomes.

Summative Evaluation – evaluation used to arrive at a rating overall and to make personnel decisions. In the proposed framework, summative evaluations would include evaluator judgments of educator performance against standards and/or progress made toward completion of a performance plan.

Unsatisfactory – demonstrating lack of competence on a standard or overall.

Appendix I – Draft Standards and Indicators for Effective Leadership

Student learning and growth is a central goal of effective administration. The Administrator Working Group of the Task Force recommends amending the language that the Board adopted in each Standard description to include "promoting the learning and growth of all students…" and also recommends that appropriate multiple measures of student learning and growth are included as sources of evidence in meeting each Standard. [11]

	I. Curriculum, Instruction, and Assessment (C.I.& A)	II. Management and Operations	III. Family and Community Engagement	IV. Professional Culture
Description of Standard	*The education leader promotes the learning and growth of all students and the success of all staff by cultivating a shared vision that makes powerful teaching and learning the central focus of schooling*	*The education leader promotes the learning and growth of all students and the success of all staff by ensuring a safe, efficient and effective learning environment*	*The education leader promotes the learning and growth of all students and the success of all staff through partnerships with families, community members, and other stakeholders that support the mission of the school and district*	*The education leader promotes success for all students by nurturing and sustaining a school culture of professional growth, high expectations, and continuous learning for staff*
	Core Indicators[12]			
Core Indicator	a. **Reflective Practice.** Involves staff as participants in continual inquiry, using meetings with teams and work groups to gather information, analyze data, examine issues, and develop new approaches in order to improve teaching and learning.			

---

[11] ESE proposes that the following language appear in this document: *The evaluation of administrators will adhere to the procedures outlined in the Framework for Educator Evaluation. The process will begin with discussion between the administrator and his/her evaluator to review the administrator's written self-reflection, which must address the Administrator's practice across all four Standards, and contain at least two goals. One goal must be focused on the professional practice of the administrator and the other focused on student learning, growth and achievement. These goals must be approved by the evaluator. This discussion and the status of the administrator will determine the administrator's professional growth plan for current year, unless already established the year before, as one of the following: 1) Developing Educator Plan, 2) Self-Directed Growth Plan, 3) Directed Growth Plan or 4) Improvement Plan. Progress made towards meeting the goals, along with the evaluation based on the four Standards, will inform the plan on which the administrator is placed the following year.*

[12] The Task Force recommends that these "Core" indicators, along with the four Standards, be adopted by the Board of Elementary and Secondary Education. ESE staff recommends the term "Core" rather than "Power" indicators, because we believe it will be more readily understood by the public.

	I. Curriculum, Instruction, and Assessment (C.I.& A)	II. Management and Operations	III. Family and Community Engagement	IV. Professional Culture
Core Indicator	**b. Curriculum**. Ensures that teachers design effective and rigorous standards-based units of instruction consisting of well-structured lessons with measurable outcomes	**b. Environment.** Develops and executes effective plans, procedures, and operational systems to address a full range of safety, health, emotional, and social needs of students Establishes routines that give staff and students a sense of order, discipline, and predictability within a caring environment	**b. Family and Community Connections.** Welcomes and encourages every family to become active participants in the classroom and school community	**b. Commitment to high Standards.** Fosters a shared commitment to high standards of teaching and learning with high expectations for achievement for all
Core Indicator	**c. Instruction.** Ensures that instructional practices reflect high expectations, engage all students, and are personalized to accommodate diverse learning styles, needs, interests, and levels of readiness	**c. Human Resources Management and Development.** Implements a cohesive approach to recruitment, hiring and induction that promotes high quality and effective staff	**c. Sharing Responsibility.** Continuously collaborates with families to support student learning and development both at home and at school	**c. Open Communications.** Addresses concerns and problems in a way that invites dialogue with those impacted by the issue Demonstrates strong interpersonal, written, and verbal communication skills, facilitates groups effectively, including: accepting feedback from supervisor, staff and stakeholders to improve performance to foster clear communication

	I. Curriculum, Instruction, and Assessment (C.I.& A)	II. Management and Operations	III. Family and Community Engagement	IV. Professional Culture
Core Indicator	**d. Assessment**. Ensures that teachers use a variety of formal and informal methods and assessments to measure student learning, growth and understanding, and that teachers makes necessary adjustments to their practice when students are not learning.	**D Scheduling and Management Information Systems.** Utilizes systems to insure that time is optimized for teaching, learning and collaboration	**d. Communication.** Engages in regular, two-way meaningful communication with families about student learning and performance	**d. Continuous Learning.** Develops and nurtures a culture where the staff seeks out and applies current research, best practices and theory **and/or** Understands the adult learning needs of staff and creates a culture of inquiry and collaboration, and supports a comprehensive professional development program for all staff that is ongoing, job-embedded
Core Indicator	**e. Evaluation.** Provides effective and timely supervision and evaluation in alignment with state regulations and contract provisions.	**e. Laws, Ethics and Policies.** Complies with state and federal laws/mandates, local school committee policies, and collective bargaining agreements and negotiations.	**e. Family Concerns.** Addresses family concerns in an equitable, effective, and efficient manner.	**e. Shared Vision.** Engages all stakeholders successfully in a shared educational vision in which every student is prepared to succeed in postsecondary education, and become responsible citizens And community contributors

	I. Curriculum, Instruction, and Assessment (C.I.& A)	II. Management and Operations	III. Family and Community Engagement	IV. Professional Culture
Core Indicator	**f. Data-Informed District Decision-Making.** Uses multiple sources of evidence related to student learning, including state, district and school assessment results and growth data, to inform school and district goals and improve organizational performance, educator effectiveness and student learning.	**f. Fiscal Systems.** Develops for the School Committee a budget that supports the district's vision, mission and goals, and allocates, manages and audits fiscal expenditures consistent with district/school level goals and available resources.		**f. Managing Conflict.** Employs strategies for responding to disagreement and dissent, constructively addressing conflict,, and building consensus throughout a district/school community

Appendix J – Draft Standards and Indicators for Effective Teaching

ESE proposes that the following language appear in this document: The evaluation of teachers will adhere to the procedures outlined in the Framework for Educator Evaluation. The process will begin with a discussion between the teacher and his/her evaluator to review the teacher's written self-assessment, which must reflect on the teacher's practice across all four Standards and contain at least two goals: one goal must be focused on the professional practice of the teacher and to the other focused on student learning, growth, and achievement. These goals must then be approved by the evaluator. This discussion and the status of the teacher will determine the teacher's Plan for current year, unless already established the year before, as one of the following: 1) Developing Educator Plan, 2) Self-Directed Growth Plan, 3) Directed Growth Plan or 4) Improvement Plan. Progress made towards meeting the goals, along with the evaluation based on the four Standards, will inform the plan on which the teacher is placed for the following year.

	I. Curriculum, Instruction, and Assessment	II. Teaching All Students	III. Professional Culture	IV. Family and Community Engagement
Description of Standard	*The teacher promotes the learning and growth of all students through planning, instructional, and assessment activities that support a cycle of creating lessons focused on clear learning objectives, designing authentic and meaningful student assessments, analyzing student performance and growth, and continuously refining learning objectives.*	*The teacher promotes the learning and growth of all students through instructional practices that establish high expectations, create a safe and effective classroom environment, and demonstrate cultural proficiency.*	*The teacher promotes the learning and growth of all students through ethical, culturally proficient, skilled, and collaborative practice.*	*The teacher promotes the learning and growth of all students through effective partnerships with families, caregivers, community members, and organizations.*
	Core Indicators[13]			
Core Indicator	a. Demonstrates the capacity to reflect on and improve her/his own practice, using meetings with teams and work groups to gather information, analyze data, examine issues, and develop new approaches in order to improve teaching and learning.			

---

[13] The Task Force recommends that these "Core" indicators, along with the four Standards, be adopted by the Board of Elementary and Secondary Education. ESE staff recommends the term "Core" rather than "Power" indicators, because we believe it will be more readily understood by the public.

	I. Curriculum, Instruction, and Assessment	II. Teaching All Students	III. Professional Culture	IV. Family and Community Engagement
	Core Indicators			
Core Indicator	b. **Assessment.** Uses a variety of informal and formal methods of assessment to measure student learning, growth, and understanding, develop differentiated and enhanced learning experiences, and improve future instruction.	b. **Learning Environment.** Creates and maintains a safe and collaborative learning environment that values diversity and motivates students to take academic risks, challenge themselves, and claim ownership of their learning	b. **Collaboration.** Develops respectful, appropriate, and collaborative partnerships with administrators, teachers, students, families, and the community to build a positive school culture and improve instruction, assessment, and student performance OR Collaborates effectively with colleagues in teams on a wide range of tasks	b. **Engagement.** Welcomes and encourages every family to become active participants in the classroom and school community.
Core Indicator	c. **Curriculum.** Designs effective and rigorous standards-based units of instruction consisting of well-structure lessons with measurable outcomes.	c. **Cultural Competence.** Frames instruction to honor the fact that each student is a member of many groups, with numerous identities, challenges, and strengths.	c. **Decision-Making.** Becomes involved in school-wide decision making, and takes an active role in school improvement planning	c. **Collaboration.** Collaborates with families in creating and implementing strategies for supporting student learning and development both at home and at school.
Core Indicator	d. **Instruction.** Uses a range of instructional techniques to meet the learning and growth needs of all students.	d. **Expectations.** Plans and implements lessons that set high expectations and make knowledge accessible for all students	d. **Shared Responsibility.** Shares responsibility for the performance of all students within the school.	d. **Communication.** Engages in regular, two-way, meaningful and culturally proficient communication with families and caregivers about student learning and performance.