INTERIM REPORT

NEW JERSEY EDUCATOR EFFECTIVENESS TASK FORCE

March 1, 2011

# TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Task Force report presents recommendations for improving student achievement in New Jersey by revamping our educator evaluation system. Our recommended system is based on the knowledge that educator effectiveness is the most important in-school factor for improving student achievement. New Jersey, like the vast majority of other states, does not have an evaluation system that accurately differentiates the effectiveness of educators. High-quality evaluation systems for our teachers and principals will enable districts and the state to vastly improve personnel decisions, such as the awarding of tenure and the setting of compensation levels, and drive significant improvements in student learning.

The report consists of four sections: teacher evaluations, principal evaluations, conditions for success, and next steps.

Teacher Evaluations
In the first section, the Task Force recommends the development of a new teacher evaluation system that is based entirely on student learning; that is, all measures used to assess effectiveness should be linked to achievement. Initially, it would comprise equal parts teacher practice (inputs) and direct measures of student achievement (outputs). Over time, however, the Task Force encourages the state to increase the percentage of the evaluation contributed by measures of student achievement.
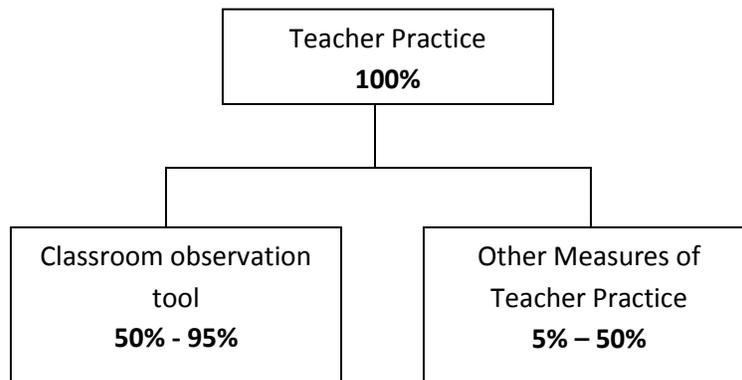
Recommended Framework for the New Teacher Evaluation System
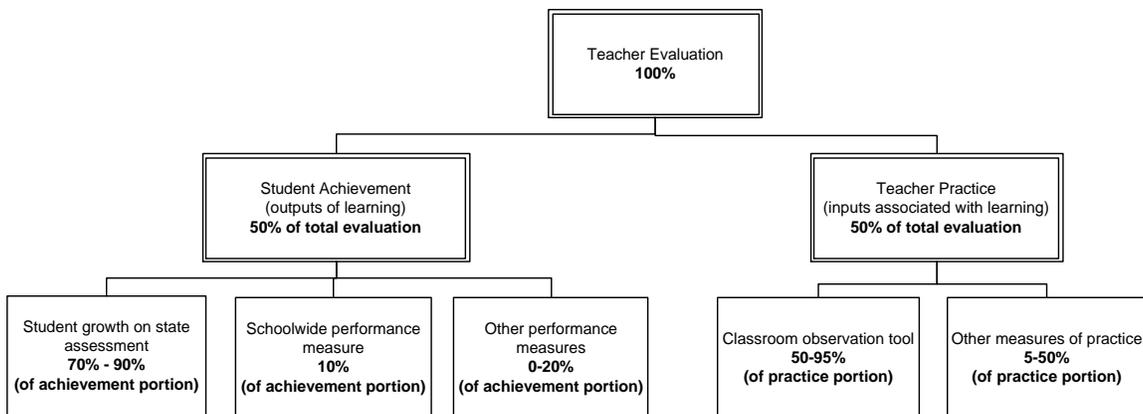
Measures of Teacher Practice

The measures of teacher practice should be based on clear performance standards that define effective teaching. The Task Force recommends that New Jersey use the new national core standards, reviewed and adapted as needed, as the basis for teacher evaluations.

Once clear standards have been established, measurement tools are needed to collect and review evidence to determine if teachers are meeting the standards. The Task Force recommends that all districts use one high-quality state-approved observation protocol and at least one additional state-approved tool to assess teacher practice.

```
                    ┌─────────────────────┐
                    │  Teacher Practice   │
                    │       100%          │
                    └─────────────────────┘
                              │
              ┌───────────────┴───────────────┐
    ┌─────────────────────┐         ┌─────────────────────┐
    │ Classroom observation│         │  Other Measures of  │
    │        tool          │         │  Teacher Practice   │
    │      50% - 95%       │         │     5% – 50%        │
    └─────────────────────┘         └─────────────────────┘
```

Because observation can be such a comprehensive tool for gathering information, the Task Force recommends that it alone comprise at least half of the weight within the teacher practice section, accounting for 50%-95% of this component. We recommend that every district use at least one additional measurement tool, and that each of these tools comprise at least 5% of the teacher practice component, but not more than 50% in combination.
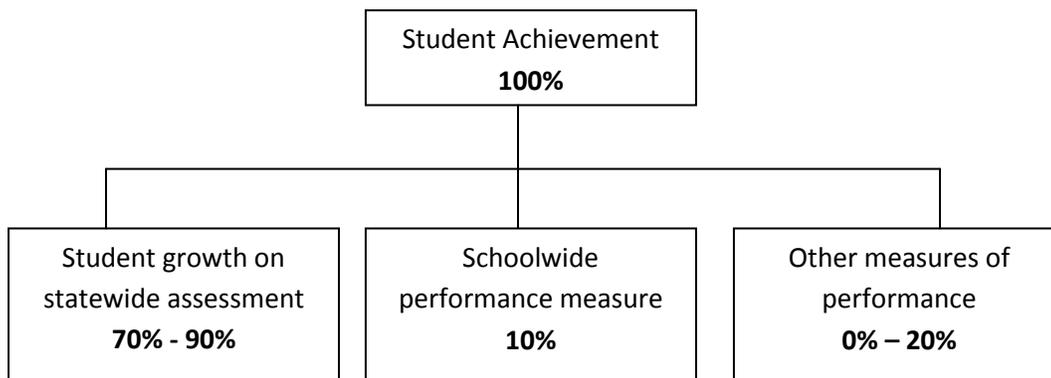
The New Jersey Commissioner of Education should develop a list of approved observation protocols and measurement tools from which districts may choose. The state may also consider developing a waiver process so districts have the opportunity to submit for approval a measurement tool that has not yet been accepted by the state.

The state's review and approval of measurement tools and their protocols will assure that they are sufficiently rigorous, valid, and reliable for measuring teacher effectiveness, and that all teachers are held to the same high standards. Providing districts some flexibility to create their own measurement tools will encourage innovation and experimentation in this area.

Measures of Student Achievement

Fifty percent of a teacher's evaluation should be based on direct measures of student achievement as demonstrated by assessments and other evaluations of student work. The Task Force recommends that the student achievement portion of the evaluation comprise two required components and one optional component. The largest required component (70% - 90%) would be an individual teacher's contribution to his/her students' progress on a statewide assessment. The other required component would be a state-approved schoolwide

performance measure (10%). A third, non-required component, would be another measure of performance (0% - 20), also State-approved.

```
                        ┌──────────────────────┐
                        │  Student Achievement  │
                        │         100%          │
                        └──────────────────────┘
                                   │
        ┌──────────────────────────┼──────────────────────────┐
┌───────────────────┐   ┌───────────────────┐   ┌───────────────────┐
│ Student growth on │   │    Schoolwide     │   │  Other measures of │
│statewide assessment│   │performance measure│   │    performance    │
│    70% - 90%      │   │        10%        │   │     0% – 20%      │
└───────────────────┘   └───────────────────┘   └───────────────────┘
```

Measures of student growth
Growth measures are preferable to attainment measures because they account for a student's academic starting point and give credit for progress made during the school year. The state will be able to generate growth scores in fall 2011. By fall 2012, the State will be able to tie growth scores to teachers.

However, because not all subjects and grades have statewide assessments, growth scores can be computed for a limited number of teachers. The Task Force recommends that the state develop assessments capable of generating growth scores in as many additional subjects and grades as appropriate and financially feasible so growth scores can be calculated for more teachers. This work can be done in partnership with districts, teachers, subject matter experts, and others.

Schoolwide performance measure
The Task Force recommends that a total school performance measure comprise 10% of the student achievement portion. This measure could be a schoolwide aggregation of all students' growth on state assessments. Alternatively, teachers could share credit for meeting a school-specific goal.   A school-specific goal would reflect an area of need identified by the school or district and approved for use by both the Commissioner and district superintendent.

Other measures of student performance
The Task Force recommends that districts be permitted to choose one or more additional measures of student achievement from a list of state-approved measures. Such measures might include student performance on nationally normed assessments or State-mandated end-of-course tests. These measures could comprise up to 20% of the achievement portion of the evaluation.

Leader Evaluations
The Task Force recommends that the principal evaluation comprise the following components and weights:

- Measures of effective practice: **40%**
- Differential retention of effective teachers (hiring and retaining effective teachers and exiting poor performers): **10%**
- Measures of student achievement: **50%**

Components of Principal Evaluations

Measures of effective practice, 40%

Measures of student achievement, 50%

Retention of effective teachers, 10%
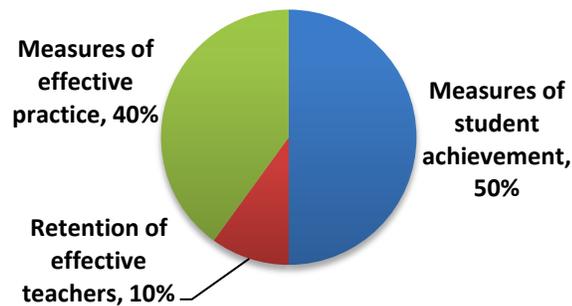
Measures of Leadership Practice

The Task Force recommends that New Jersey adopt the updated Educational Leadership Policy Standards: ISLLC 2008.[1] The ISLLC standards have been adopted by most states, are widely accepted by the profession, and serve as a credible and useful foundation for principal evaluations.

The Task Force recommends that the Commissioner develop or adopt statewide performance indicators to establish clear and consistent expectations for all principals. Districts should be able to choose the data sources and tools they wish to use from a list of state-approved rubrics, templates, and tools. The Commissioner may also develop a waiver process for districts to submit locally developed tools to the state for approval.

Retention of Effective Teachers

The principal's success in building and maintaining a high-quality faculty is critical to school success. Differential retention of effective teachers means hiring and retaining effective teachers and exiting poor performers. The Task Force recommends that differential retention of effective teachers contribute 10% of the principal evaluation.

The following indices should be used to measure differential retention:
- Principal's effectiveness in improving teacher effectiveness (i.e., growth of teachers' ratings)
- Principal's effectiveness in recruiting and retaining effective teachers
- Principal's effectiveness in exiting ineffective teachers

---

[1] New Jersey uses an older version of the ISLLC standards, adopted in 2003 and based on 1996 ISLLC standards, to accredit leadership preparation programs, license school leaders, and approve professional development activities.

It is critical to note that principals can only be judged against this measure if they are given a clear role in teacher hiring, organizing professional development, dismissing ineffective teachers, and more.

Measures of Student Achievement

The Task Force recommends that a principal's evaluation be based substantially on empirical measures of student learning. We have identified two different measures of achievement that should be included in the principal's evaluation: aggregated student growth on standardized assessments and "school-specific goals."

Components of Student Achievement
Portion of Principal Evaluations

- Student achievement: Aggregated performance on assessments
- Student achievement: School-specific goals
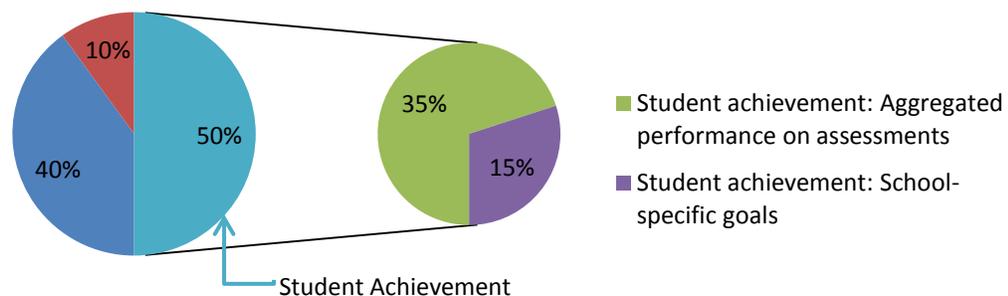
The Task Force recommends that principals be evaluated on the aggregated growth of all students on statewide assessments for all subjects and grades. This measure should comprise 35% of the total evaluation. The Task Force recommends that every principal also be measured on at least one school-specific goal, such as high school graduation rate increase. A school-specific goal would reflect an area of need identified by the school or district and should be approved by the Commissioner of Education. This measure or combination of measures would comprise 15% of the total evaluation.

Conditions for Success

The Task Force believes that in order to maximize the positive influence of these new evaluation frameworks, the State should simultaneously pursue a number of related policies and activities. These "Conditions for Success," will lay the foundation and build the support structure for this new system. This list of issues to consider include the following: training for those conducting observations, informing educators of the new system's components and implications, ensuring high-quality data systems, continuously monitoring the system's effects after implementation, and more.

Next Steps

The Task Force has identified a number of additional activities to be pursued over the next several months. This includes soliciting feedback from the State Board of Education and other education experts and stakeholders; further study of appropriate performance measures for teachers of special populations and non-tested subjects and grades; and developing recommendations for implementing the new evaluation system, including the possible use of pilots.

INTRODUCTION

In recent years, policymakers and other education stakeholders have pursued a wide array of strategies designed to improve academic outcomes, especially for our most disadvantaged children. Of these, efforts to improve educator effectiveness have been among the most prominent, popular, and important.

For decades we have known that a number of out-of-school factors, most notably poverty, can substantially depress student learning. But research has shown conclusively that teachers and principals have the ability to overcome these obstacles and help all students achieve at high levels.

The impact of our most effective teachers is remarkable. Studies have shown that if we were to give at-risk students access to our highest-performing teachers, we could close the achievement gap, helping deliver on our nation's promise to provide equal opportunity to all. But the data also show that if a child is placed in the classrooms of a series of ineffective teachers, he/she will struggle mightily to recover academically and may never catch up.

The cornerstone of any broad initiative to improve educator effectiveness is an evaluation system that accurately measures our educators' influence on student learning. Evaluations that fail to account for differences in effectiveness are unfair to families and their children.

But they are also unfair to the adults working in our schools. These professionals will never receive the respect they deserve if we continue to treat teachers and administrators like machines on an assembly-line instead of the highly skilled professionals that they are.

The purpose of this report is to help New Jersey create a new system for evaluating teachers and principals that leads to substantial and lasting improvements in public education. Such a system will provide actionable information to schools, parents, taxpayers, and policymakers. As a consequence, the state will be better positioned to help educators improve, rethink compensation plans, reform tenure, and much more.

Guiding Principles
The recommendations of this Task Force flow from three guiding principles. The first is that the needs of students are paramount.

Public education cannot function without adults, and changes to the system inevitably affect their day-to-day work and long-term careers. We must keep this in mind and be sensitive to its implications.

But public education exists for the benefit of children. It is society's means of ensuring that all children have the chance to reach their full potential and lead healthy, productive, and satisfying lives. We believe that the reforms recommended here are good for both children and

adults. But we understand that some elements of this report may generate opposition from adult-oriented interest groups. We believe that when the interests of adults and the interests of children don't align it is our duty to side with the latter.

The second principle relates to our belief that all children can achieve at the highest levels. Some contend that a child's neighborhood, race, and family income amount to destiny—that we can only expect so much from public schools because external forces are determinative. This would suggest that an educator evaluation system based on student achievement is unfair because teachers and principals would be held to account for something over which they have no control.

We believe that the purpose of public education is to lead all students to high levels of achievement no matter where they begin.

Our third principle is our belief in the efficacy of educators. We believe that educators, equipped with the right skills, knowledge, and dispositions and given the proper supports, have the power to inspire, engage, and broaden the life opportunities of students.

The evaluation system recommended in this report reflects these convictions.

Finally, we would not argue that our plans are perfect, only that they will substantially improve the status quo. Similarly, we do not argue that this report should be the final word, but the beginning of a long-avoided conversation.

Process
Governor Christie established the Education Effectiveness Task Force through a September 28, 2010 Executive Order. Nine members, with experience in and knowledge of education policy, administration, and teaching were selected (members are listed in the Appendix) on October 28, 2010.

The Task Force was charged with recommending an educator evaluation system based on measures of effectiveness. According to the Executive Order, its recommendations must include measures of student achievement (representing at least 50% of the evaluation); demonstrated practices of effective teachers and leaders; and weights for the various components.

An initial report was mandated by March 1, 2011. After the submission of the report, the Task Force is to receive comments from the public, stakeholders, and the State Board of Education and to review and revise its recommendations.

To complete its work, the Task Force, with the support of staff from the Department of Education, reviewed the latest research on educator evaluations, examined systems in use both in-state and nationally, and studied a range of issues related to the development of high-quality

evaluation systems, such as observation protocols, growth measures, and special education considerations. The Task Force met 12 times between November 16, 2010 and March 1, 2011.

A full list of the resources utilized by the Task Force, including presenters and written materials, is included in the Appendix.

Report Outline
The report is composed of four sections. The first offers recommendations for a new teacher evaluation system. It includes two subsections, one for measures of teacher practice; the other for measures of student achievement.
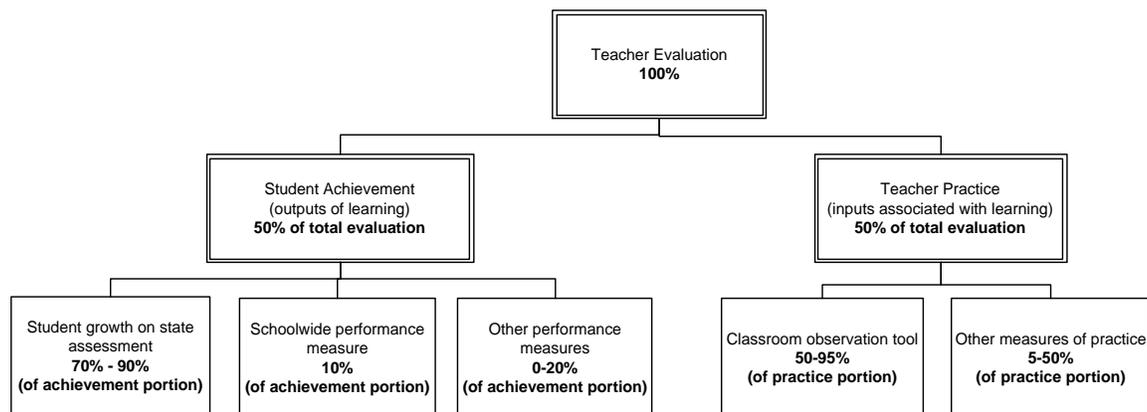
The second section offers recommendations for a new principal evaluation system. It has three subsections dedicated to measures of practice, retention of effective educators, and student achievement, respectively.

The third section includes a set of recommendations regarding additional considerations. Through our work, the Task Force developed a great appreciation for the broad infrastructure needed to build high-quality evaluation systems. We highlight a number of issues, such as the need for additional assessments and expanded administrator training, that the state might consider alongside our other recommendations.

The final section is on next steps. The Task Force has identified a number of activities to pursue in the months to come to help further advance the cause of improved educator evaluations.

SECTION I:
SECTION I:
TEACHER EVALUATIONS

Recommended Framework for the New Teacher Evaluation System

```
                          ┌─────────────────────┐
                          │ Teacher Evaluation  │
                          │       100%          │
                          └─────────────────────┘
```

Teacher Evaluation 100%	
Student Achievement (outputs of learning) **50% of total evaluation**	**Teacher Practice** (inputs associated with learning) **50% of total evaluation**

Student growth on state assessment **70% - 90%** (of achievement portion)	Schoolwide performance measure **10%** (of achievement portion)	Other performance measures **0-20%** (of achievement portion)	Classroom observation tool **50-95%** (of practice portion)	Other measures of practice **5-50%** (of practice portion)

Purpose of an Educator Evaluation System

Teachers have a powerful influence on student learning. No in-school factor has a greater bearing on achievement than the effectiveness of the adult in front of a classroom. Though out-of-school factors certainly exert a significant influence, for years we have known that teachers can help even the most disadvantaged students excel.

A high-quality evaluation system has the power to accurately assess the effectiveness of teachers and differentiate between those excelling and those struggling. In this way, an evaluation system can be the foundation for a wide range of critical personnel decisions. If we have reliable information on effectiveness, districts and the state can make highly informed decisions related to hiring, tenure, compensation, dismissal, and more.

And when used properly, a strong evaluation system will also help educators become more effective.[2]

It will help clarify expectations. Teachers will know what behaviors, practices, and results are expected and by what metrics they will be evaluated.

---

[2] For more on this subject, see the discussion in DC IMPACT:
http://dc.gov/DCPS/Learn+About+Schools/School+Leadership/IMPACT+(Performance+Assessment)

12

12

It will provide meaningful feedback. Results from observations, test scores, and more will clearly delineate strengths and weaknesses and provide a path for improvement.

It will facilitate collaboration. By providing a common evaluation framework and language, the system will enable educators to work together, within and across schools, to improve their collective work.

It will improve and target professional development. A strong evaluation system will indicate areas for improvement, enabling schools, districts and the state to develop improved professional development opportunities and ensure that each teacher receives training that matches her needs.

In these ways, an effective evaluation system will help earn the trust and support of teachers. They will know that the system isn't in place merely to declare winners and losers; it exists to help teachers improve their capacity to help students succeed.

The Task Force recommends that as it develops a new teacher evaluation system, the State ensures that it succeeds on both fronts: assessment and development.

Essential Features
Through our research, we have noted that the most compelling evaluation systems share a number of key characteristics. These features contribute to the fairness and transparency of evaluations and, most importantly, help ensure that they are highly correlated with and, therefore, help drive gains in student achievement.

The Task Force recommends that a new teacher evaluation system adhere as closely as possible to the follow principles:

- The system should be based on clear standards that describe the characteristics of effective and ineffective teaching.
- The standards and evaluative criteria should reflect a high level of rigor, meaning the system has the highest expectations for all teachers and students.
- To the greatest extent possible, the system should have a uniform design so measures are consistent across districts and within schools.
- The system should allow for differences in teaching positions (performing arts, career tech, special education, for example, do not lend themselves to the same types of assessments as math and science).
- The system should make use of multiple measures or data sources so an array of evidence is utilized when assessing a teacher's effectiveness.
- Care should be given to ensuring that the measures assess educator effectiveness with reasonable accuracy (validity) and generate consistent results across different raters and contexts (reliability).

- Those implementing the evaluation system must faithfully adhere to the system's measurement process, including the collection of data and the observation of teachers.

Summative Rating Categories
The Task Force recommends that the new system have four summative categories: Highly Effective, Effective, Partially Effective, and Ineffective. The number of rating categories should be large enough to give teachers a clear picture of their performance, but small enough to allow for clear, consistent distinctions between each level and meaningful differentiation of teacher performance[3].

---

[3] "Teacher Evaluation 2.0," p. 7, The New Teacher Project, 2010.

MEASURES OF TEACHER PRACTICE

Definition of effective teaching
Most evaluations and personnel decisions have not adequately distinguished teachers of varying levels of effectiveness. In a robust evaluation system, effective teaching is defined by practices that contribute to student learning and empirical measures of student achievement.

The Task Force recommends that measures of effective teacher practice represent 50% of a teacher's evaluation.

Teaching Standards
Teaching standards serve as the foundation for teacher evaluations by outlining the professional responsibilities, behaviors, and expectations of teachers. New Jersey's current standards for teachers were adopted by the State Board of Education in 2003.

According to New Jersey regulations, the standards are used in the accreditation of teacher preparation programs, the recommendation of candidates for certification, and the approval of professional development programs. However, they have not been a required part of teacher evaluations.

The Task Force recommends that these standards serve as the basis for teacher evaluations in the state.

However, new draft core teaching standards have been developed by the Interstate New Teacher Assessment and Support Consortium (INTASC). Unlike the original 1992 INTASC standards that were designed for "beginning" teachers, these are intended as professional practice standards for use at different developmental stages of the teacher's career. They differ from the previous standards in several other ways: there is greater emphasis on the learner, greater knowledge and skill is expected around the use of assessment data to improve instruction and support learner success, and technology is infused throughout all the standards.

The Task Force recommends that the new national standards, when finalized, be carefully reviewed by the state and considered for adoption. If New Jersey is to have a robust, trusted, and transparent evaluation system, it must be grounded in a widely acknowledged and respected set of standards.

Summary of the Draft Model Core Teaching Standards

	Standard	Description
The Learner and Learning	1. Learner Development	The teacher understands how children learn and develop, recognizing that patterns of learning and development vary individually within and across the cognitive, linguistic, social, emotional, and physical areas, and designs and implements developmentally appropriate and challenging learning experiences.
	2. Learning Differences	The teacher uses understanding of individual differences and diverse communities to ensure inclusive learning environments that allow each

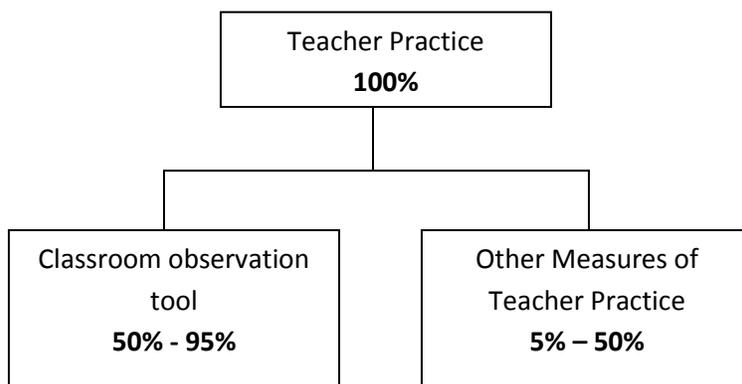|		learner to reach his/her full potential.
	3. Learning Environments	The teacher works with learners to create environments that support individual and collaborative learning, encouraging positive social interaction, active engagement in learning, and self motivation.
Content	4. Content Knowledge	The teacher understands the central concepts, tools of inquiry, and structures of the discipline(s) he or she teaches and creates learning experiences that make these aspects of the discipline accessible and meaningful for learners.
	5. Innovative Applications of Content	The teacher understands how to connect concepts and use differing perspectives to engage learners in critical/creative thinking and collaborative problem solving related to authentic local and global issues
Instructional Practice	6. Assessment	The teacher understands and uses multiple methods of assessment to engage learners in their own growth, to document learner progress, and to inform the teacher's ongoing planning and instruction.
	7. Planning for Instruction	The teacher draws upon knowledge of content areas, crossdisciplinary skills, learners, the community, and pedagogy to plan instruction that supports every student in meeting rigorous learning goals.
	8. Instructional Strategies	The teacher understands and uses a variety of instructional strategies to encourage learners to develop deep understanding of content areas and their connections, and to build skills to access and appropriately apply information
Professional Responsibility	9. Reflection and Continuous Growth	The teacher is a reflective practitioner who uses evidence to continually evaluate his/her practice, particularly the effects of his/her choices and actions on others (students, families, and other professionals in the learning community), and adapts practice to meet the needs of each learner.
	10. Collaboration	The teacher collaborates with students, families, colleagues, other professionals, and community members to share responsibility for student growth and development, learning, and well-being.

Measurement Tools

Once clear standards have been defined for an evaluation system, measurement tools are needed to collect and review evidence to determine whether teachers are meeting the standards. These measurement tools must be valid (the capacity to measure what they are intended to measure) and reliable (the capacity to measure accurately and consistently).

The Task Force recommends the use of a high-quality observation protocol and at least one additional measurement tool to assess teacher practice. The Commissioner should develop a list of approved measurement tools and protocols from which districts can choose. In addition, the Commissioner should develop a waiver and review process through which districts could submit alternative tools for approval.

The state review and approval of measurement tools and their protocols will assure that they are sufficiently rigorous, valid, and reliable while also providing districts flexibility to innovate and develop their own tools.[4]

Because observation can be such a comprehensive tool (it is able to cover most teaching standards), the *Task Force recommends that it alone comprise at least half of the weight within teacher practice, accounting for 50%-95% of this portion. We further recommend that every district use at least one additional measurement tool and that each of these tools comprise at least 5% of the teacher practice score, but not more than 50%.*

```
                    ┌─────────────────────┐
                    │   Teacher Practice   │
                    │        100%          │
                    └─────────────────────┘
                              │
              ┌───────────────┴───────────────┐
    ┌─────────────────────┐         ┌─────────────────────┐
    │ Classroom observation│         │   Other Measures of  │
    │        tool          │         │   Teacher Practice   │
    │     50% - 95%        │         │      5% – 50%        │
    └─────────────────────┘         └─────────────────────┘
```

Classroom Observations

Observation protocols are the most common tool for measuring teacher practice, but how thoroughly and frequently they are conducted and what they evaluate vary widely. Observations are required in New Jersey, and they are used in all the model systems we reviewed.

Some of the model systems have created their own observation protocols (e.g., DC IMPACT and Harrison, Colorado) and some have adopted existing observation protocols (e.g., Delaware uses Danielson's Framework for Learning). Essential elements of successful observation practices include well-trained observers, a high-quality rating rubric, and the faithful administration of the selected protocol.

The Task Force recommends a minimum of four observations a year, as well one annual summative evaluation for all teachers. Successful districts often conduct frequent observations and provide feedback to the teachers on a regular basis. In Washington, DC, every teacher has five formal observations per year, and in Harrison, Colorado, every teacher has at least four spot observations (between 10-15 minutes each); probationary teachers have eight.

---

[4] The Bill and Melinda Gates Foundation in collaboration with many prominent research organizations are in the process of testing a wide array of measurement tools in the Measuring Effective Teaching project: http://metproject.org/

There are numerous observation protocols in use and many are well grounded in research. Among the most well-known is Charlotte Danielson's "Framework for Teaching," which is currently used by more than 30% of New Jersey districts.

<u>Additional Tools</u>
We recommend that the Commissioner develop a list of approved additional tools from which districts can choose. Potential options include the following.

- Documentation logs/portfolios*:* Logs or portfolios can provide evaluators with information about student learning that might not be uncovered by assessments or standard in-class observations. Teachers can collect artifacts showing how well their practices adhere to performance standards (e.g., planning and preparation, lesson plans, student assignments).   If these tools are utilized, the state and districts should take care to ensure that the material collected is truly representative of the teacher's work.

- Student surveys: Students have a unique and valuable perspective on classroom environment and their teachers' effectiveness. Studies have found that the results of student surveys can be tightly correlated with student achievement results. Persuasive evidence can be found in the Gates MET study, which uses a survey instrument called Tripod.[5] It asks students if they agree or disagree with statements about their classroom's instructional environment, such as:

  > "My teacher knows when the class understands and when we do not."
  > "My teacher has several good ways to explain each topic that we cover in class."
  > "My teacher gives me useful feedback that helps me improve."

- Assessments of teachers' pedagogical content knowledge: The MET study is also testing the use of assessments developed by ETS to measure a teacher's ability to recognize and diagnose students' misunderstandings of lessons. The assessments measure teachers' general, specialized, and pedagogical content knowledge. If these assessments, or others, are found to be valid measures of teacher effectiveness, the Department should consider including them as an approved tool.

Reviewers
Any evaluation system that emphasizes the value of teacher practice will inevitably increase the demands on principals and other administrators; observations and other reviews of teacher work require significant investments of time. The Commissioner might consider addressing this issue through the use of the following models, which have the potential to both reduce the burdens placed on administrators and generate stakeholder support.

---

[5] *Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project*, Bill and Melinda Gates Foundation, 2009

Peer Assistance and Review (PAR)

PAR was created to be a collaborative assessment process, with peer teachers and a greater emphasis on professional development. The program identifies underperforming teachers and provides them with a supportive yet consequential professional improvement plan. Teachers that participate in a PAR program relinquish their tenure rights. The following components of a PAR system are recommended:

- *A PAR Panel:* An oversight panel comprising both teachers and school leaders that provides assistance and makes decisions on dismissal. The panel members should be outside the bargaining unit so as to eliminate any possible conflicts of interest.

- *Consulting Teachers*: Educators also outside the bargaining unit that provide instructional support to teachers under review and collect data through observations. They report monthly on the progress of the teachers to the PAR Panel.

Based on information gathered through the review program, the PAR Panel makes recommendations to the principal and superintendent for both provisional and tenured teachers regarding contract renewal, recommendation for a second year in PAR, or contract termination.

Master Teachers

Several evaluation systems studied by the Task Force use "master teachers" (in addition to the principal) to conduct teacher reviews; DC IMPACT and the system developed by Colorado's Harrison District Two are notable examples. The use of master teachers can be valuable because they can confirm the accuracy of a principal's evaluation and offer teachers an additional set of suggestions for improvement.[6] In both the referenced systems, the master teachers are from the district—not the teacher's school.

---

[6] In DC, master educators are expert practitioners who work at the district level. They conduct observations without prior knowledge of the scores given by principals. Over the course of one year, the principal conducts three formal observations and a master educator will conduct two. In Harrison school district, district-level observers review only those teachers at the very low and high ends of the rating scale. Principals conduct spot observations of instruction eight times each semester for probationary teachers and four times each semester for non-probationary teachers.

MEASURES OF STUDENT ACHIEVEMENT

The Task Force recommends that measures of student achievement initially comprise 50% of a teacher's evaluation in the new system. Over time, as the system is improved and gains support, we recommend that measures of student achievement grow to a larger portion of the evaluation.

Principles Guiding Recommendations

Use Multiple Measures

No single empirical measure can fully summarize a teacher's performance, so evaluation systems should use a number of measures to determine whether a teacher is effective.

A brief explanation of growth models

Growth models measure student progress. Such systems assess student performance at two points in time and generally control for factors such as previous performance or demographic characteristics.

Growth scores can be tied to teachers: in simple terms, if the students in a teacher's class make greater gains than similar students elsewhere, that teacher is credited with effectively raising student achievement.

Some say growth scores should not be used in evaluations. But based on our research, we believe that they provide important, if not perfect, information. When used in conjunction with other measures, growth can tell us a great deal. Despite limitations, these scores tell us *something*; that is, evaluations are better off using them than disregarding them altogether.

We recommend that the new system use growth alongside other measures and that the State work with testing experts to continually improve their validity.

Use Growth Models

Measuring attainment, for example whether a student reaches proficiency on a state assessment, doesn't take into consideration academic growth. Failing to account for progress is particularly unfair in the case of students who start a school year academically behind their peers.

Growth scores are a fairer and more accurate means of measuring student performance and teachers' contributions to student learning. In fact, over half of the states surveyed by the Council of Chief State School Officers (CCSSO)— 24 out of 43—reported that they either already do or plan to use student growth in analyzing teacher effectiveness.[7]

The state will be using a growth model to measure student achievement on state assessments with data from 2009-2010. These scores will be released in fall 2011.

Use the Best Assessments Possible

The state does not have a single, comprehensive system of assessments covering all subjects and all grades.[8] The new evaluation system should use the best assessments available to generate

---

[7] *State Growth Models for School Accountability: Progress on Development and Reporting Measures of Student Growth*, 2010, by the Council of Chief State School Officers.
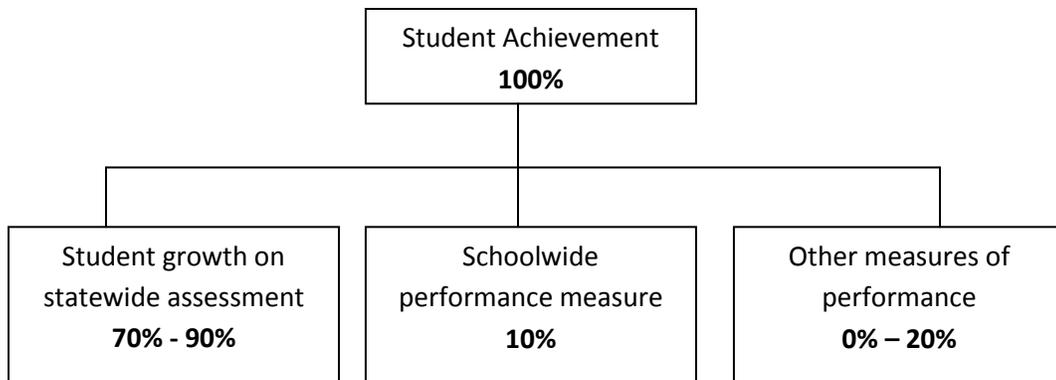
[8] New Jersey conforms to the federal NCLB requirements that students be tested in math and language arts in grades 3-8, and once again in high school. Some science assessments are also required by the state.

empirical measures of student performance. Where possible, teachers should be evaluated using state standardized tests. For currently non-tested subjects and grades, the assessments used should be rigorous and comparable across classrooms and should measure learning growth.

Measures of Student Achievement
The Task Force recommends that the student achievement portion of the evaluation comprise two required components and one optional component. The largest required component would be an individual teacher's contribution to her students' progress on a statewide assessment. The other required component would be a schoolwide performance measure. A third, non-required component would be another measure of performance.

The schoolwide and non-required performance measures that districts could choose would be approved by the Commissioner to assure goals are appropriate and sufficiently challenging yet attainable. The Commissioner should also consider creating a waiver and review process by which districts could submit for approval some other performance measure to be used in the evaluation.

```
                    ┌─────────────────────────┐
                    │  Student Achievement    │
                    │         100%            │
                    └─────────────────────────┘
                                 │
        ┌────────────────────────┼────────────────────────┐
┌───────────────────┐  ┌───────────────────┐  ┌───────────────────┐
│ Student growth on │  │    Schoolwide     │  │  Other measures   │
│ statewide         │  │ performance       │  │  of performance   │
│ assessment        │  │ measure           │  │                   │
│    70% - 90%      │  │       10%         │  │     0% – 20%      │
└───────────────────┘  └───────────────────┘  └───────────────────┘
```
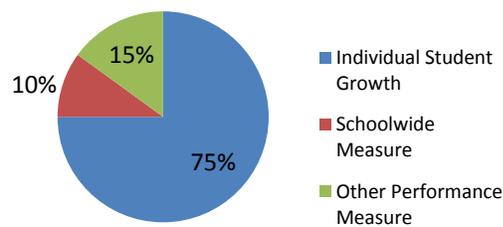
The Task Force recommends that each district be allowed to choose whether to use two or three components and have discretion over how to weight these components within the bands recommended here.

For example, District A may choose to use only the two required components. In that case, growth on the statewide assessment would comprise 90% and the schoolwide measure 10%. District B, however, may choose to use all three components, deciding to weight individual growth at 75%, the schoolwide measure at 10%, and another measure at 15%.

District A

10%

90%

- Individual Student Growth
- Schoolwide Measure

District B

15%

10%

75%

- Individual Student Growth
- Schoolwide Measure
- Other Performance Measure
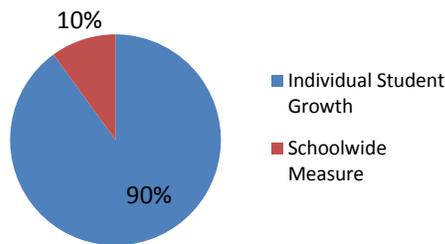
Individual Student Growth

The Task Force recommends that a teacher's student growth score make up the core of the student achievement section of her evaluation: 70% - 90% of the student achievement portion (or 35%-45% of the total evaluation).

Because not all subjects and grades have statewide assessments, currently growth scores can be computed for a limited number of teachers. For math and language arts/literacy in grades 4 – 8, these scores will be available in the fall of 2012.[9]

The Task Force recommends that the State work to develop standardized assessments in as many additional subjects and grades as appropriate so growth scores can be calculated for a growing number of teachers.

The Task Force recommends that the Commissioner approve the types of assessments that are acceptable for use in these areas in advance of the development of standardized assessments.

In some subjects, standardized year-end assessments may never be suitable (e.g., art, music, physical education, or career-tech fields). *In these cases, the Task Force recommends the use of other rigorous performance-based evaluations of student work. The use of re- and post- tests would be ideal so student growth, not merely attainment, can be gauged.*

A general rule embraced by the Task Force is that, within a district, different categories of teachers may be evaluated differently (e.g., gym teachers vs. 4th grade math teachers), but all teachers within a category should be evaluated using the same measures and weights.

Several states, such as Delaware, have assembled subject-specific groups of teachers and subject-matter experts to develop recommendations for addressing assessments in untested grades and subjects. New Jersey should consider convening similar groups. The groups of experts could provide guidance on how to develop new standardized assessments, how to measure growth before such assessments are available, and how to measure growth in subjects

---

[9] The link between students' growth scores and individual teacher will be completed in Fall 2012.

where standardized assessments are inappropriate. This work can be done in partnership with districts, teachers, subject matter experts, and others.

Schoolwide Performance Measure ("Shared Attribution")
The Task Force recommends that a total-school measure comprise 10% of the student achievement portion (or 5% of the total evaluation).

This measure could be a schoolwide aggregation of all students' growth on state assessments. Alternatively, teachers could share credit for meeting a school-specific goal. A school-specific goal would reflect an area of need identified by the school or district and approved for use by both the Commissioner and district superintendent.[10] The list of state-approved measures might include:

- High school graduation rate increase
- Promotion rates from 9th to 10th grade
- College matriculation rate increase
- Proficiency level increases for an underserved subgroup
- Advanced-level increases for the school or subgroups
- Student attainment level or proficiency increase on nationally normed or supplemental assessments (e.g., Iowa Test of Basic Skills, Stanford 9, International Baccalaureate, APA, SAT, ACT, early childhood)

The Task Force believes that the use of such shared attribution scores would focus all teachers on a school or district priority, thereby facilitating collaboration among educators and increasing the likelihood of accomplishing a major task.

Other Measures of Performance
The Task Force recommends that the Commissioner promulgate a list of state-approved student achievement measures. Interested districts would be permitted to choose a measure or measures from this list to comprise up to 20% of the student achievement portion of the evaluation (or up to 10% of the total evaluation).

Possibilities might include:
- Growth or attainment on a nationally normed tests (e.g., Iowa Test of Basic Skills)
- Growth or attainment on supplemental assessments (e.g., Stanford 9)
- State-mandated end-of-course tests (e.g., biology)
- Student achievement goals, also called "student learning objectives" (e.g., DC's IMPACT system, Harrison, CO)[11]

---

[10] This concept came from IMPACT, DC's Effectiveness Assessment System for School Leaders, 2010-2011.

[11] Teachers set goals for student growth, subject to certain parameters, with their principal's approval. Teacher evaluation is based on students' progress on the established goals, as determined by an end-of-the-year principal review using pre-determined assessments. While not comparable across classrooms, student learning objectives (SLOs) have been shown to be effective measures of student achievement growth.

- Grade- and subject-specific student outcomes (e.g., graduation/college acceptance rates)

Scoring

There are many different ways to combine the scores of the components of educator evaluations (e.g., the index system or panel approach). For example, since four summative rating categories are required, a district might choose to rate a teacher's performance on each component on a 1 – 4 scale, weight the components, and then sum the results.

So a teacher in District B who was found to be effective (a score of 3) on her students' growth scores (75% weight), partially effective (2) on her school's other performance measure (15%), and highly effective (4) on the schoolwide measure (10%) would earn for the achievement section:

$$(3 * .75) + (2 * .15) + (4 * .1) = 2.95 = \text{Effective}$$

This is just one of many ways to combine the component parts. *The Task Force recommends that the Commissioner develop guidelines and model scoring systems for districts to follow.*

SECTION II:
SCHOOL LEADER EVALUATIONS

School leaders play a crucial role in raising student achievement. According to research, principal and teacher quality account for nearly 60% of a school's total impact on student achievement, with principals alone accounting for 25%. The influence of school leaders is so significant because of their enormous contributions to schoolwide success conditions. Key among these contributions are activities related to teacher effectiveness, such as hiring, professional development, evaluation, retention, and dismissal.

Furthermore, even though a single teacher can have a profound impact on student learning over the course of a year, that effect generally fades unless a student's subsequent teachers are equally effective.[12] In order for a student to have high-quality learning gains year after year, the entire school must have a culture that supports learning and that school must be populated by the most effective teachers. These conditions are only brought about by high-performing school leaders.

In New Jersey, school leaders include principals, assistant principals and supervisors. Each of these positions has unique responsibilities, and therefore each should be evaluated based on their performance of those responsibilities. The Task Force recommends that *all* school leaders be evaluated, but has developed specific evaluation recommendations only for principals in this report.

Purpose of Principal Evaluation
As is the case with teacher evaluations, the Task Force believes that the purposes of principal evaluations are two-fold: assessment and development. In order for a principal evaluation system to be truly successful it must accurately assess the current performance of a principal and provide feedback on where and how to improve.

Definition of Effective Leadership
A large body of research has identified the leadership practices that produce successful schools. Principal evaluation systems have used this information to varying degrees.[13] But very few principal evaluation systems have held principals accountable for the academic outcomes of their students. It is only recently that reform-minded policy experts and education researchers have concluded that principal evaluations must include measures of both practice and student performance.

---

[12] "Principal Effectiveness: A New Principalship to Drive Student Achievement, Teacher Effectiveness, and School Turnarounds with Key Insights from the UEFTM" by New Leaders for New Schools, 2009

[13] Source: The Evaluation of Principals: What and How do States and Districts Assess Leadership?; Ellen Goldring, Andrew Porter, et. al., 2008

A principal's work has direct and indirect influences on school success.[14] Through the direct actions of hiring and retaining high-quality teachers, supporting their work, fostering a culture of student achievement, and more, the principal indirectly influences student achievement. Thus, for the principal, "…achieving results through others is the essence of leadership."[15] Schools with high at-risk populations that exceed expectations share a common element: a strong leader committed to education.[16]

Summative Categories
T*he Task Force recommends that the new principal evaluation system have the same four summative categories as the teacher evaluation system: Highly Effective, Effective, Partially Effective, and Ineffective.*

The Components of Principal Evaluations
The Task Force recommends that the new principal evaluation comprise the following components with the following weights:
- Measures of effective practice: **40%**
- Differential retention of effective teachers (hiring and retaining effective teachers and exiting poor performers): **10%**
- Measures of student achievement: **50%**

Components of Principal Evaluations

Measures of effective practice, 40%

Retention of effective teachers, 10%

Measures of student achievement, 50%

---
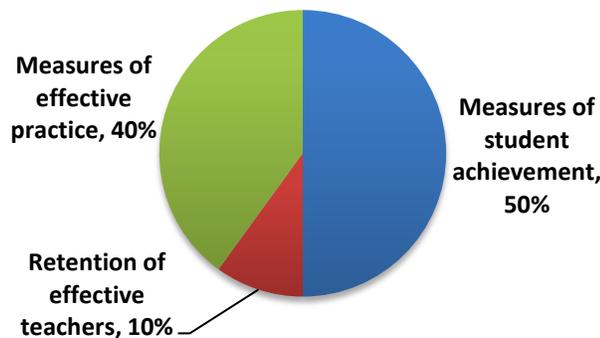
[14] One study conducted with the Dallas, Texas, Public Schools found "…that the quickest way to change the effectiveness of a school, for better or worse, is to change the principal" (Mendro, R.L. (1998). *Student achievement and school and teacher accountability.* Journal of Personnel Evaluation in Education, 12, pp. 263- 264.

[15] Mendro, R.L., p. 39.

[16] Cawelti, G. (1999). *Portraits of six benchmark schools: Diverse approach to improving student achievement.* Education Research Service.

MEASURES OF PRACTICE

Performance Standards
Before we can recommend how to evaluate principal effectiveness we must define the essential skills and responsibilities of an effective principal. This is the purpose of performance standards*.*

The Task Force recommends that New Jersey adopt the updated and revised Educational Leadership Policy Standards: ISLLC 2008.[17]   The Interstate School Leaders Licensure Consortium (ISLLC) standards have been adopted by most states, are widely accepted by the profession, and serve as a credible and useful foundation for principal evaluations.

New Jersey uses an older version of the ISLLC standards, adopted in 2003 and based on the 1996 ISLLC standards, to accredit leadership preparation programs, license school leaders, and approve professional development activities. However, they are *not* currently required by code for use in principal evaluation. Using the same standards across the continuum from preparation through practice will promote consistency and help drive systemic change.

The 2008 ISLLC standards provide high-level guidance and insight about the traits, functions of work, and responsibilities expected of school and district leaders, and are organized into six domains:

PERFORMANCE EXPECTATION 1: Vision, Mission, and Goals
Education leaders ensure the achievement of all students by guiding the development and implementation of a shared vision of learning, strong organizational mission, and high expectations for every student.

PERFORMANCE EXPECTATION 2: Teaching and Learning
Education leaders ensure achievement and success of all students by monitoring and continuously improving teaching and learning.

PERFORMANCE EXPECTATION 3: Managing Organizational Systems and Safety
Education leaders ensure the success of all students by managing organizational systems and resources for a safe, high-performing learning environment.

PERFORMANCE EXPECTATION 4: Collaborating with Families and Stakeholders
Education leaders ensure the success of all students by collaborating with families and stakeholders who represent diverse community interests and needs and mobilizing community resources that improve teaching and learning.

PERFORMANCE EXPECTATION 5: Ethics and Integrity
Education leaders ensure the success of all students by being ethical and acting with integrity.

---

[17] Revised ISLLC standards were adopted through the Council of Chief State School Officers (CCSSO) as model standards in 2008. ,

PERFORMANCE EXPECTATION 6: The Education System
Education leaders ensure the success of all students by influencing interrelated systems of political, social, economic, legal, and cultural contexts affecting education to advocate for their teachers' and students' needs.

The Task Force recommends that the domains within the standards be weighted equally by all districts throughout the state. When research identifies which domains are most highly correlated with school success, this issue should be reconsidered.

Performance Indicators
Performance indicators provide descriptions of observable or demonstrable behaviors for each standard. That is, the performance indicators describe the types of performance that will occur if a standard is being met successfully.

New Jersey has not adopted a set of performance indicators for each standard, leaving to individual interpretation what specific actions and results are expected from an effective principal. New Jersey is not alone in this. A flurry of activity is now underway across the nation as states work to develop principal evaluation systems aligned to clear standards and performance indicators.

A handful of principal evaluation systems, complete with evaluation instruments and tools, already exist (e.g., McREL's Principal Evaluation System, New Leaders for New Schools Leadership Rubric, the New Jersey Principals and Supervisors Association's *Teacher and School Leader Evaluation Standards and Data Sources*). Some states and district have created systems of their own (e.g., DC IMPACT, Harrison, Colorado, and Rhode Island). The Task Force has reviewed many of these systems.

The Task Force recommends that the Commissioner develop a set of performance indicators or adopt existing performance indicators for the state. The state should also establish a waiver process by which districts could develop rigorous, comparable performance indicators that meet guidelines established by the Commissioner.

Evidence of Performance
No single data source can adequately capture the complexities of a school leader's work. A holistic view of professional practice and performance is needed. *The Task Force recommends that the principal evaluation include multiple data sources for gathering evidence of performance.*

We further recommend that the evaluation include the following data sources:
- Observations of instructional meetings, faculty meetings, professional learning communities, and other activities in which principals should be deeply engaged; such observations should be conducted by the superintendent or a designee and occur twice per year, at minimum.

- Annual surveys of teachers and families to assess school culture, learning climate, community engagement and other key elements.
- Document logs or portfolios (prepared by the principal) that provide evidence of success associated with the standards; interviews to review portfolios should occur twice per year.[18]
- Evidence of the principal's progress toward meeting district goals; assessment should be conducted twice per year.

The Commissioner may consider approving other data sources that may be used by districts, for example 360 degree survey tools (e.g., VAL-Ed).[19]

Evaluation Tools
The Task Force recommends that the Commissioner develop a list of approved rubrics, templates and tools that have been validated for use in leader evaluation, and develop a review process for districts to submit their own locally developed tools for review and approval.

Requiring each district to use state-approved measurement procedures and data collection protocols will enhance clarity, increase fairness, and ease inter-district comparisons. Should the NJDOE select only one set of tools for use across the state, it would provide a common language for evaluation and provide the opportunity for realizing economies of scale, especially for professional development.

Evaluators and Frequency of Evaluations
The Task Force recommends that principal evaluations be performed by superintendents or their appropriately trained designees. Thorough training should be provided to the evaluators so that the review process is implemented in a rigorous and consistent manner.

The Task Force recommends that reviews of leadership practice occur at least twice per year. This will enable principals and their evaluators to engage in constructive conversations that provide the opportunity for principals to make needed adjustments.[20] *In addition, an annual summative evaluation should occur at the end of the year.*

---

[18] In current code, a professional growth plan is required for all principals. This plan should be based on the professional growth goals established as a result of the evaluation.
[19] VAL-ED:
http://www.wallacefoundation.org/KnowledgeCenter/KnowledgeTopics/CurrentAreasofFocus/EducationLeadership/Documents/VAL-ED-Technical-Manual.pdf

[20] In DC's IMPACT evaluation system, instructional superintendents evaluate principals twice each year, which guarantees regular formative feedback. However they are expected to be in their principals' schools *at least* once every two weeks.

RETENTION OF EFFECTIVE TEACHERS

The principal's success in building and maintaining a high-quality teaching staff is critical to the success of the school. Differential retention of effective teachers means hiring and retaining effective teachers and exiting poor performers. *The Task Force recommends that differential retention of effective teachers contribute 10% of the principal evaluation.*

The following indices should be used to measure differential retention of effective teachers:

- Principal's effectiveness in improving teacher effectiveness (i.e., growth of teachers' ratings)
- Principal's effectiveness in recruiting and retaining effective teachers
- Principal's effectiveness in exiting ineffective teachers

The Task Force recommends that principals be empowered with the role of human capital manager. It is critical to note that principals can only be judged against this measure if they are given a clear role in teacher hiring, organizing professional development, dismissing ineffective teachers, and more. Current New Jersey law states that superintendents are responsible for most of these personnel decisions. To make the individual school accountable for its student achievement outcomes, the school principal must be given more control over the inputs. The Commissioner should develop policies to ensure principals and superintendents have responsibility for personnel decisions.

As previously discussed, teacher effectiveness is the most important in-school factor related to student achievement, and principals influence teacher effectiveness by providing instructional leadership and through their personnel decisions.
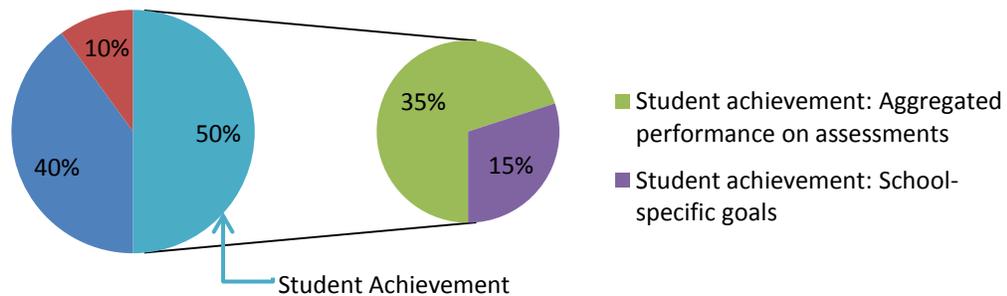
MEASURES OF STUDENT ACHIEVEMENT

A principal's primary indicator of success is the improvement of student achievement throughout her school. Accordingly, a principal's evaluation should be based substantially on empirical measures of student learning.

As discussed above, a principle guiding our recommendations is the use of multiple measures overall and within the student achievement category.   The use of multiple measures will provide a district a number of angles by which to inspect principal performance, and it will broaden the list of performance indicators on which the principal, and therefore her faculty, will focus.

The Task Force has identified two different measures of achievement that should be included: aggregated student growth on standardized assessments and "school specific goals."

**Components of Student Achievement
Portion of Principal Evaluations**

The Task Force recommends that principals be evaluated on the aggregated growth of all students on statewide assessments (all subjects and grades). This measure should comprise 35% of the total evaluation (or 70% of the achievement portion of the evaluation). The state's development of end-of-year assessments across a broader swath of subjects and grades will facilitate the availability of a larger number of growth scores, providing a fuller measure of the school's overall performance.

The Task Force recommends that every principal also be measured on at least one school-specific goal.[21] A school-specific goal would reflect an area of need identified by the school or district and should be approved for use by both the Commissioner and district superintendent. This measure or combination of measures would comprise 15% of the total evaluation, or 30% of the student/school performance portion of the evaluation.

---

[21] This concept came from IMPACT, DC's Effectiveness Assessment System for School Leaders, 2010-2011.

The list of state-approved measures might include:
- High school graduation rate increase
- Promotion rates from 9th to 10 grade
- College matriculation rate increase
- Proficiency level increases for an underserved subgroup
- Advanced level increases for the school or subgroups
- Student attainment level or proficiency increase on nationally normed or supplemental assessments [e.g., Iowa Test of Basic Skills (ITBS), Stanford 9, International Baccalaureate, APA, SAT, ACT, early childhood]
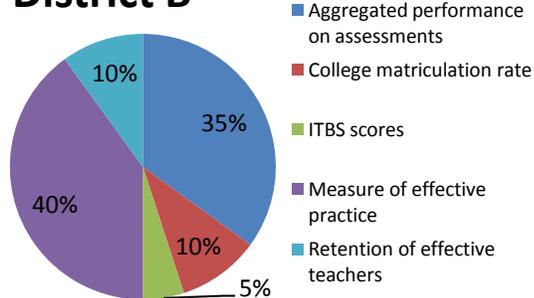
Implementation

District A could choose to select only one school-specific metric—in this instance, high school graduation rates. This would account for 15% of the principal's evaluation. District B, however, could choose two school-specific measures—here, college matriculation rates and ITBS scores. These two measures would combine to total 15% of the evaluation.
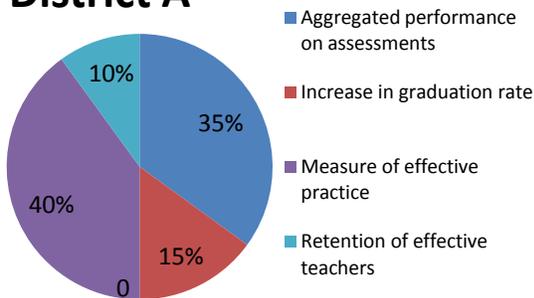
Scoring

As in the teacher section, a district might decide to use a 1 – 4 scale for each of the principal's components, weight the components, and then sum the results.

So a principal in District A found to be highly effective (a score of 4) on aggregated growth scores (35%), effective (3) in raising graduation rates (15%), highly effective (4) in measures of her practice (40%), and partially effective (2) in her retention of effective teachers (10%) would earn:

$$(4 * .35) + (3 * .15) + (4 * .4) + (2 * .1) = 3.65 = \text{Highly Effective}$$

This is just one of many ways to combine the component parts. *We recommend that the Commissioner develop guidelines and model scoring systems for districts to follow.*

SECTION III:
CONDITIONS FOR SUCCESS

As the Task Force studied the complex field of educator evaluations, it became clear that evaluation systems cannot be considered in isolation. In order for an evaluation system to have a meaningful and lasting impact, many other supportive policies and practices must be in place. That is, the success of the evaluation systems recommended here will depend largely on the environment into which they are introduced. Though the identification of these conditions for success was not required by the governing Executive Order, the Task Force believed that the cause of improving educator effectiveness would be well served by raising these interrelated issues.

What follows is an overview of the key issues the Governor and his administration might consider as they build and implement an improved evaluation system.

Evaluator Capacity and Training
The evaluation system recommended in this report calls for a substantial portion of a teacher's evaluation to be based on observations of teacher practice. The responsibility of conducting classroom observations rests on principals, other administrators, and possibly seasoned, skilled teachers. In order for these observations to be fair to teachers, to elicit a high level of trust and confidence in the system, and ultimately drive improvements in student learning, high-quality evaluator training is essential.

As the instructional leaders for their schools, principals will need adequate training on the observation protocol and other measurement tools used to evaluate teachers. If evaluation results are to be tied to a wide array of personnel decisions, the importance of proper training cannot be overstated. We strongly recommend that New Jersey's Commissioner of Education prioritize such training and work with districts to ensure that those conducting teacher observations--and therefore exerting an enormous influence on teachers' professional standing--be qualified to do so. The system depends on a high level of reliability and accuracy in the evaluations.

The Task Force recommends that the Commissioner consider the development of regional training centers, so that training will be consistent and high-quality. These training centers could be modeled on those that existed under previous administrations or are operating successfully in other states.

The same recommendations apply to those evaluating school leaders. Superintendents must be adequately trained in the observation protocol and other measurement tools used to evaluate principals.

Engaging and Educating Teachers and Principals

Most teachers and principals are accustomed to the current, longstanding systems of evaluation. The recommendations in this report would, if implemented, represent a major change to their professions; without fully explaining the new system and its implications, the state would risk confusing, and possibly alienating, its educators.

We recommend that the Commissioner develop plans for ensuring that educators are made aware of the contours and consequences of the new system, given the opportunity to learn why and how it will work, and engage in its implementation. This could include developing statewide professional development programs, working with existing programs, or partnering with districts, membership organizations, or other nonprofits to develop tools, practices, or policies to successfully implement the system. One possible model to emulate can be found in Delaware, which has formed groups of teachers and subject matter experts to develop measures of student achievement in non-tested subjects and grades.

Observation Frequency and Teacher Feedback

A commonality among the strong systems we studied was the increased frequency of observations. In many schools, classroom visits by administrators are rare or perfunctory. This means a teacher is given few opportunities to demonstrate her skills and knowledge, and little opportunity to receive constructive feedback. We believe a cornerstone of a robust evaluation system is a commitment to frequent observations coupled with an ongoing dialogue between teacher and observer that offers the opportunity for continuous improvement. The Commissioner should set guidelines around the minimum number of observations teachers should receive.

Reconsidering Priorities

The demands of implementing a quality educator evaluation system present a number of challenges to schools and districts. One of the greatest is the need to reconsider how educators spend their time. For principals, conducting observations, writing evaluations, and then conferring with teachers require a significant commitment of time. With so many other responsibilities and regulatory requirements, administrators will be hard-pressed under current conditions to find such time. As the state and its districts develop a comprehensive strategy for improving educator effectiveness, finding ways to enable administrators to adequately do this important work should be a priority.

The Commissioner should conduct a thorough code review to eliminate redundancies and unnecessary mandates that pose a burden on the school leaders' time. Another possible solution is to shift some non-instruction functions to other administrators or the central office.

Teachers may need to spend more time learning how to reach the most disadvantaged students, use data in the classroom, and align instruction with clear performance goals. This suggests the possible need for changes in teacher preparation programs, different or expanded professional development opportunities and more opportunities to engage in professional learning communities.

Developing High-Quality Assessments

Currently, fewer than half of educators teach in tested grades and subjects, so student growth scores can only be generated for a portion of the state's teaching corps. Growth scores, however, are absolutely essential for the system recommended here; they provide a measure of how far students have progressed in the span of a school year, thereby taking into account each student's starting point.

The state should determine how best to develop valid and reliable empirical measures of student performance in all subjects and grades. Whether traditional standardized assessments or others tools that accurately assess learning, these measures should be tightly aligned with clear standards and, to the fullest extent possible, measure growth in addition to attainment.

The state could engage teachers and other subject matter experts in an initiative to develop these assessments. This would go far toward ensuring these assessments measure what matters most and generating support among practitioners.

Though this will be a challenging and time-consuming task, the state should not delay taking it on. Developing empirical measures of student learning in all subjects and grades will send a powerful message about the importance of standards, assessments, and student achievement. Moreover, the new evaluation system will not be complete and internally consistent until every teacher's evaluation has some empirical measure of her students' learning.

Developing High-quality Data Systems

The success of our recommended evaluation system will depend largely on the quality of the data systems that undergird it. We must have systems that not only calculate student growth scores and tie these results back to teachers, but also process this information swiftly so it can be used by the state, districts, and schools in a timely fashion. Moreover, this information must make its way to teachers if they are to have a true opportunity to learn from the data. Similarly, if districts are to use interim assessments for evaluation or formative purposes, data systems must be prepared for this additional responsibility.

The state should give particular attention to several issues related to growth scores. The NJDOE will be able to link student achievement scores to individual teachers by fall 2012. However, the state must find ways to verify student rosters for all teachers in a timely fashion, and apportion responsibility for student performance in team teaching settings and when students change teachers during the year.

Additional Observers

Although principals and assistant principals are typically the primary observers, it may be the case that they lack the specific content knowledge to effectively evaluate all teachers, especially those in higher grades and specialized subjects. Developing a cadre of "master teachers" with content expertise who are empowered to observe would help address this issue.

It would also decrease the total number of observations a time-strapped administrator must conduct and give a school another view of a teacher's practice.

For example, Washington, D.C.'s IMPACT system uses district-level "Master Educators" in addition to administrators to conduct classroom observations. A Master Educator is defined as "an expert practitioner in a particular content area who will serve as an impartial observer" of teacher practice. Master Educators give confidence to teachers that their evaluations will be less vulnerable to the subjectivity of a single person and that the observer is knowledgeable in their content area.

Access to Resources
The type of robust evaluation system recommended in this report places new responsibilities on schools and districts. To help those on the ground implement this new system the state should consider developing a range of supports. The list could include tools that ease data collection and facilitate the tracking of students or programs that help teachers with interim assessments and data analysis.

Continuous Improvement
A common refrain from those with the most impressive evaluations systems is that the work of building a great system is never done. A number of our presenters noted that no evaluation system is perfect and that each year they must strive to make it fairer, more accurate, and more transparent. Despite our faith in its attributes, we know that the system recommended here will not be perfect from the start. Much will be learned about its strengths and weaknesses during its implementation. We strongly encourage policymakers and practitioners to continuously study this new system and make modifications over time to ensure that it is both improving educator effectiveness and driving student learning. This could be accomplished through a variety of means, such as empirical studies of changes in student performance over time or regular surveys of teachers and principals. This kind of feedback loop will also help build support for the new system, as those in schools will see that it is responsive to changing conditions and new information and tightly aligned to explicit results.

Increased Principal Autonomy
Our recommended system would make a principal highly accountable for the gains of her school's students and the effectiveness of her teaching faculty. This increased accountability should be coupled with increased authority at the school level; that is, more responsibility for outputs requires greater control over inputs. Principals, in collaboration with superintendents, should have the power to select and develop their teachers and dismiss those not succeeding at their craft. They should also have greater authority over their budgets and other resources.

Teachers of Special Populations
During the course of our research, we came to the conclusion that in some cases the framework recommended here may not apply fully. Teachers of special populations, including ELL and special education students, may need to be evaluated using different measures. We

recommend that the Commissioner convene work groups to determine how best to evaluate teachers who work in these areas.

Superintendent Evaluation

The Task Force recommends that superintendents and their professional staffs be evaluated in part based on the quality of their principals' teacher evaluations and their records of development and differential retention. Each level of the education system must be held accountable for student achievement and each must be aligned along the same goals.

Evaluations for All

The executive order charged the Task Force with recommending evaluation measures for teachers and school leaders. However, for schools to be most effective, all staff should be evaluated, including librarians, nurses, school social workers, secretaries and custodians. In this way, each person will be treated as an important member of the school community responsible for contributing to student achievement.

Implementation

Given that the long-term sustainability of this new system will be influenced by its early effectiveness, we recommend that the administration carefully plan an implementation process. Several issues mentioned in this report, such as the availability of growth scores, the development of additional assessments, the timely delivery of data to districts, the need for data collection and other types of support, and high-quality training for reviewers need to be addressed.

The state might also consider piloting the system in a limited number of districts before taking it statewide. This would afford policymakers and practitioners the opportunity to build support and resolve initial challenges before attaching high stakes to the results. A gradual roll out would also give the state time to align other policies and practices, such as reforms to tenure and compensation, with the new evaluations.

One possible implementation plan would look as follows:

> Fall 2011: Pilots
> Measures of student achievement and the link to individual teachers are needed:
> - Student growth scores using 2009/10 and 2010/11 will be available in Oct. 2011.
> - Participating pilot districts would need to provide student-teacher roster data for the 2010/2011 school year (roster data will be collected statewide for 2011/2012).
> - Other measures of student achievement will need to be developed for teachers of non-tested subjects and grades.
>
> Measures of teacher practice must be identified and evaluators need to be trained:
> - Districts will need to identify an observation protocol and at least one additional measure of teacher practice.
> - Training for evaluators will be needed.

- Expectations for teachers and how they will be evaluated must be clearly communicated.

Fall 2012: Statewide rollout without "high stakes"
The link of student achievement data to individual teachers will be available statewide:
- The first growth scores attributable to teachers of language arts and math in grades 4, 5, 6, 7 and 8 will be available in fall 2012 using 2010/2011 and 2011/2012 data.

Fall 2013: Full implementation statewide; impact on personnel decisions:
After two years of testing the evaluation system and making adjustments, it should be ready to be used in making personnel decisions.

SECTION IV:
NEXT STEPS

The completion of this report represents the first step in developing improved educator evaluation systems. The Task Force has identified the following next steps that it might pursue in an effort to continue advancing this important work:

- Solicit feedback on the report's recommendations from the State Board of Education and other stakeholder groups in order to make revisions and refinements.
- Convene sub-groups to develop recommendations for student achievement measures for teachers of special populations and non-tested subjects and grades.
- Develop detailed recommendations for piloting the evaluation system in selected districts.

APPENDIX

Task Force Members

- Brian Zychowski, Task Force Chair: Superintendent, North Brunswick schools

- Derrell Bradford, Executive Director, Excellent Education for Everyone (E3)

- Donna Chiera, Executive of the American Federation of Teachers, NJ, and Special Education Resource Teacher (Perth Amboy)

- Jane Cosco, retired teacher (Paramus) and Director of Operation Goody Bag

- Ross Danis, former Associate Dean of Education, Drew University; current Executive Director, Newark Education Trust

- Rafael Fajardo, former President of the Elizabeth Board of Education

- Peggy Sue Juliano, Executive Board Member of the Lacy Township High School PTA (invited)

- Rev. Edwin Leahy, Headmaster of St. Benedict's Prep in Newark

- Jesse Rector, Principal, North Star Academy Charter School; Clinton Hill (Newark) Campus

Presenters

- Evaluation experts:

    o Laura Goe (Teacher Quality Source; ETS): Using evaluation to improve teacher effectiveness

    o Drew Gitomer, Distinguished Researcher and Director of the Understanding Teaching Quality Center, ETS

    o Margaret Terry Orr, Bank Street College of Education: Evaluating Principals-- Considerations and Recommendations from Research and Practice

    o Charlotte Danielson: Operationalizing performance standards for all teachers

- Model evaluations systems:

    o Mike Miles, Superintendent of Harrison, CO: Using performance data in an evaluation system

    o Paul Bambrick, North Star Academy Charter School, Newark: Uncommon's evaluation system

    o Jason Kamras: DCPS: Washington DC's IMPACT system

    o Tony Davis, Principal Consultant: McRel principal and teacher evaluation systems

    o Ted Herschberg, Professor, Public Policy and History; Director, Operation Public Education, University of Pennsylvania

- o Ulcca Joshi Hansen, Associate Director for Educator Effectiveness, Colorado Legacy Foundation: Colorado's Framework for Evaluating Educators
- New Jersey districts
  - o Nathan Parker, Summit Superintendent: Summit's system and approach
  - o Brian Osborn, Maplewood-South Orange Superintendent: Principal evaluation
- Stakeholder groups
  - o EQuATE: Creating a More Perfect System: A Draft Report on Improving Educator Effectiveness from Concerned Practitioners and Policymakers: Earl Kim, Superintendent of Schools, Montgomery Township
  - o NJDOE Professional Development Advisory Committee: Brian Cory, Co-Chair; Vice Principal at Tenafly HS
  - o State Special Education Advisory Council: Howard Lerner, Chair; Kathy Roberson, Vice Chair
  - o New Jersey Association of School Administrators: Richard Bozza, Executive Director
  - o New Jersey Principals and Supervisors Association: JoAnn Bartoletti, Executive Director, Debra Bradley, Director of Government Relations, Jay Doolan, Director of FEA
- NJDOE Staff:
  - o Elaine Davis, Director of Leadership Development: Principal evaluation
  - o Jeffrey Hauger, Director of State Assessment: Assessment issues
  - o Bari Erlichson, Director, Office of Education Data: NJSMART and Growth Models

Works Cited and Works Reviewed

"Area of Focus - Area One: Effective Teaching and Leadership." www.cms.k12.nc.us . Charlotte-Mecklenburg Schools. 10 September 2010.

"Area of Focus – Area Two: Performance Management." www.cms.k12.nc.us. Charlotte-Mecklenburg Schools. 10 September 2010.

"Assessing the Effectiveness of School Leaders: New Directions and New Processes." *The Wallace Foundation.* March 2009.

Babo, Gerard, et al. *A Proposed Statewide System of Evaluation for New Jersey Public School Principals*. February 2011.

Baker, Eva L., et al. "Problems of the Use of Student Test Scores to Evaluate Teachers." *Economic Policy Institute*. #278. 29 August 2010.

Bell, Courtney A. and Gitomer, Drew H. "Evaluating Teaching and Teachers." *Educational Testing Service.* DRAFT

Betebenner, Damian W. "A Primer on Student Growth Percentiles." *National Center for the Improvement of Educational Assessment (NCIEA)*. 8 February 2008.

Blank, Rolf K. "State Growth Models for School Accountability: Progress on Developing and Reporting Measures of Student Growth." *Council of Chief State School Officers.* June 2010.

Carson, Becca, et al. "The Evaluation of Principals: What and How Do States and Districts Assess Leadership?" *Presented at the Annual Meeting of American Educational Research Association.* March 2008.

Center for Education and Policy Research at Harvard University. *Teacher Employment Patters and Student Results in Charlotte-Meckleburg Schools*. 23 February 2010. Powerpoint Presentation.

Colorado State Council for Educator Effectiveness. "New Jersey Educator Effectiveness Task Force." 4 February 2011. Powerpoint Presentation.

Danielson, Charlotte and McGreal, Thomas L. *Teacher Evaluation: To Enhance Professional Practice*. Princeton, NJ: Educational Testing Service, 2000.

Danielson, Charlotte. "Assessing Teacher Effectiveness." February 2011. Powerpoint Presentation.

Darling-Hammond, Linda. "Evaluating Teacher Effectiveness: How Teacher Performance Assessments Can Measure and Improve Teaching." *Center for American Progress.* October 2010.

"Effectively Linking Teachers and Students." *Data Quality Campaign*. July 2010.
www.dataqualitycampaign.org

EQuATE. "Creating a More Perfect System: A Draft Report on Improving Educator Effectiveness from Concerned Practitioners and Policymakers." *Submitted to the Governor's Task Force on Educator Evaluation*. 4 February 2011.

"Evaluating Principals: Balancing Accountability with Professional Growth." *New Leaders for New Schools*. 2010.

Glazerman, Steven, et al. "Evaluating Teachers: The Important Role of Value-Added." *Brown Center on Education Policy at Brookings.* 17 November 2010.

Goe, Laura, et al. "A Practical Guide to Evaluating Teacher Effectiveness." *National Comprehensive Center for Teacher Equality*. April 2009.

Goe, Laura. "Using Student Growth as Part of a Multiple Measures Teacher Evaluation System: Teachers in Non-Tested Subjects and Grades." *2010 NSAII/NEA Teaching and Learning Conference*. 9 November 2010. Powerpoint Presentation.

Goe, Laura. "Teacher Evaluation in Transition: Using Evaluation to Improve Teacher Effectiveness." *New Jersey Educator Effectiveness Task Force*. November 2010. Powerpoint Presentation.

Goldstein, Jennifer and Noguera, Pedro A. "A Thoughtful Approach to Teacher Evaluation." *Education Leadership.* March 2006: 31-37.

Hamilton, Laura S., et al. "Incorporating Student Performance Measures into Teacher Evaluation Systems." *Center for American Progress.* Rand Corporation: 2010.

Hershberg, Ted and Robertson-Kraft, Claire. "A Grand Bargaining for Education Reform: New Rewards and Supports for New Accountability." *Task Force on Teacher Effectiveness - NJ Department of Education*. 20 January 2011. Powerpoint Presentation.

Hinchey, Patricia H. "Getting Teacher Assessment Right." *National Education Policy Center*. December 2010. http://nepc.colorado.edu/publication/getting-teacher-assessment-right.

"Impact: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel – Group 1 and Group 2 Teachers." *District of Columbia Public Schools*. 2010.

"Impact: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel – School Leaders." *District of Columbia Public Schools*. 2010.

Keeling, David, et al. "The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness." *The New Teacher Project.* 2nd Ed. www.widgeteffect.org.

Massachusetts Department of Elementary and Secondary Education. *MCAS Student Growth Percentiles: Interpretive Guide*. October 2009.

"McREL's Principal Evaluation System." *Mid-continent Research for Education and Learning.* 2010.

"McREL's Teacher Evaluation System." *Mid-continent Research for Education and Learning.* 2009.

MET Project. "Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project." *Bill and Melinda Gates Foundation*.

MET Project. "Working with Teachers to Develop Fair and Reliable Measures of Effective Teaching." *Bill and Melinda Gates Foundation.* June 2010.

Milanowski, Tony. "Principal Evaluation." *Center for Educator Compensation Reform*. Powerpoint Presentation.

New Jersey Association of School Administrators. *Statement to the New Jersey Educator Effectiveness Task Force.* 4 February 2011.

New Jersey Principals and Supervisors Association. *Recommendations on Educator Evaluation Systems to the Education Effectiveness Taskforce.* 8 February 2011.

North Star Academy. *Teacher Evaluation Rubric*. Revised. 25 October 2010.

Orr, Margaret Terry. "Evaluating Principals: Considerations and Recommendations for Research and Practice." *Bank Street College of Education*. *New Jersey Educator Effectiveness Task Force*. 4 February 2011. Powerpoint Presentation.

 "Principal Effectiveness: A New Principalship to Drive Student Achievement, Teacher Effectiveness, and School Turnarounds." *New Leaders for New Schools*. 2009.

Professional Development Advisory Committee for School Leaders. *Presentation to the New Jersey Educator Effectiveness Task Force. 26 January 2011*

Rothstein, Jesse. "Review of *Learning about Teaching*." *National Education Policy Center*. January 2011.

School District Two – Harrison. *Professional Educator Evaluation System: Rules and Procedures*. Adopted 5 January 2010.

State of Rhode Island and Providence Plantations Department of Elementary and Secondary Education. *The Rhode Island Model Educator Evaluation System*. Working Draft. January 2011. www.ride.ri.gov.

Stronge, James H. "Teacher and School Leader Evaluation Standards and Data Sources." *New Jersey Principals and Supervisors Association.* 1 February 2011.

Stronge, James H. "The Design and Implementation of Teacher and School Leader Evaluation Instruments and Systems – Technical Report." *New Jersey Principals and Supervisors Association.* 1 February 2011.

Teacher Effectiveness Task Force – Final Report. *Los Angeles Unified School District.* 13 April 2010. http://etf.lausd.net

The New Teacher Project. *Teacher Evaluation 2.0*. www.tntp.org.