State of New Jersey
DEPARTMENT OF EDUCATION
PO BOX 500
TRENTON, NJ 08625-0500

CHRIS CHRISTIE
Governor

KIM GUADAGNO
Lt. Governor

CHRISTOPHER D. CERF
Commissioner

February 5, 2013

TO: Chief School Administrators
            Charter School Lead Persons

FROM: Peter Shulman, Assistant Commissioner/Chief Talent Officer *PS*
            Division of Teacher and Leader Effectiveness

SUBJECT: **Educator Evaluation Research, Reports, and Policy Development**

As you know, the effort to improve educator evaluation in New Jersey has been a top priority for the State since the Educator Effectiveness Task Force (EETF) was convened in 2010. Woven into the fabric of this initiative is the need to strike a balance between two important drivers: a sense of urgency to make changes that will benefit students and educators, and a strong desire to proceed thoughtfully and to incorporate feedback from New Jersey educators along the way. I am writing today to share some results from the first year of our teacher evaluation pilot and to explain how these results – along with several other sources of information – continue to inform development of the statewide evaluation framework.

We published on our website today three important reports that address the first year of the teacher evaluation pilot (2011-12) and the national impact of new teacher evaluation systems: the Evaluation Pilot Advisory Committee (EPAC) Interim Report, the Rutgers University Graduate School of Education (RUGSE) Year 1 Report, and the Culminating Findings from the Measures of Effective Teaching (MET) Project's Three-Year Study. Even though these reports have only recently been compiled, we have been analyzing both qualitative and quantitative data from pilot districts and national experiences to improve our evaluation activities for well over a year.

While I encourage you to read the reports in full, Appendix A includes a brief summary of each, along with some of the unique findings. Highlighted within these pages are common lessons learned, Department actions, and practical recommendations that will help all districts better prepare for new evaluation systems.

I. Background, Timeline, and Progress to Date

At the New Jersey Department of Education ("the Department"), we know that teachers and leaders matter. Educators are the heart of our school system – and research supports our belief in their impact. Principal and teacher quality account for nearly 60 percent of a school's total impact on student achievement.[1] The effect of increases in teacher quality swamps the impact of any other educational investment, such as reductions in class size.[2] Replacing one poor teacher with an average one increases a classroom's lifetime earnings by several hundred-thousand dollars.[3]

Most likely, you've heard these statistics before – but we cannot emphasize enough the value we place on the people who help students succeed. And while we all agree that fostering educator quality is a top priority, we recognize that it is a complicated endeavor. With evaluations, we are changing entrenched systems that have failed to distinguish among educators and to provide meaningful feedback for many years. Fortunately, our state is home to committed educators who are willing to roll up their sleeves and work with the Department to effect difficult but necessary change.

After *The Widget Effect*[4] report showed that the vast majority of all teachers are rated satisfactory or better – despite troubling achievement gaps and drop-out rates across the country – the Obama Administration linked state commitments to reforming evaluation systems to competitive grant funds. In true New Jersey fashion, we are paving the way for other states and districts by responding to this national call. In fact, improving educator effectiveness was a critical element in both our state's $38 million Race to the Top III award and our approved Elementary and Secondary Education Act (ESEA) waiver.

And so we have embarked upon a multi-year program of research and development to construct an evaluation system that will accurately assess educator performance. From the beginning, we designed a process that would evolve based on lessons learned along the way. For this reason, it is important to consider the findings from the EPAC and RUGSE reports – which cover one phase of the initiative – in context of evaluation reform as a whole. Given that the first year of the pilot ended last summer, and we are well into the second year, we have already incorporated many of the recommendations outlined in each.

The timeline included as Appendix B depicts various aspects of the multi-year process to improve evaluations.

---

[1] Marzano et al., 2005
[2] Goldhaber, 2009
[3] Chetty et al., 2011
[4] 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.

II. Common Themes and Department Action

As depicted below, several sources of guidance are informing New Jersey's evaluation initiative.
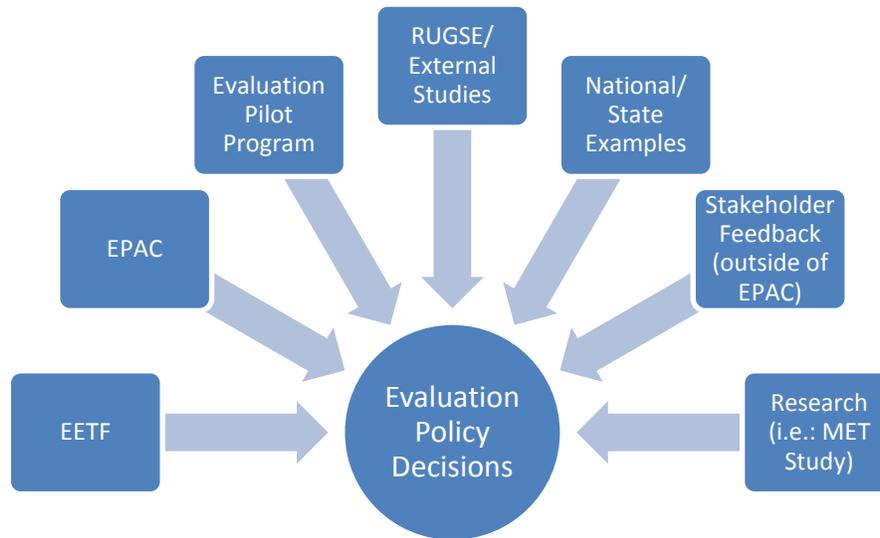
We are pleased that recommendations from the two reports released today on our first year pilot echo much of what we're learning from other sources, including the MET Study. While the three reports' authors are different, striking commonalities emerge in each. As expected, the findings they offer include successes and challenges. While we are gratified by areas of success, we are driven by the challenges and committed to employing the lessons from early implementation.

Perhaps most importantly, the MET study, which used randomized, controlled study groups of over 3,000 teachers across seven districts nationally, found that effective teaching can be measured through high-quality evaluations that utilize multiple measures of teacher practice and student outcomes. The study also provides valuable insight on conducting teacher observations, specifically spotlighting the duration, frequency, and types of observations that may be employed to produce the most accurate results.

Using multiple sources of data, we have identified some essential ingredients for success. Through collaborative approaches, quality training, committed leadership, and clear communication, districts that prioritize high-quality evaluation processes have achieved significant success.

As part of our measured approach, lessons from the first year of pilot implementation have informed the second year, which is informing proposed regulations. Several of these lessons are outlined below with corresponding actions taken by the Department to address the challenges and make improvements for the statewide system.

1. **Culture Change**

Adoption of new evaluation systems entails a culture shift in many schools – a shift away from compliance-based, low-impact, and mostly perfunctory evaluations to a focus on educators as career professionals who receive meaningful feedback and opportunities for growth.

Common Findings:
- Perhaps the most important outcome of the pilot is that in many schools, the new evaluation system has led to a collective refocus on the elements of effective teaching.
- The EPAC and RUGSE reports include first-hand examples of teachers and supervisors – sometimes for the first time – using common language and definitions of good teaching, and engaging in a collaborative process to help all educators continuously improve their practice. However, that positive collaboration requires a commitment from both district/school leaders and teachers to engage meaningfully in the new evaluation system.
- The MET Study affirms our core principle for improving evaluations: effective teaching can be measured – and effective educators impact student achievement. The culture shift required to successfully implement new systems is dependent upon a shared belief in this principle.

Department Takeaways and Actions:
- We are striving to provide common terminology at the state level, and encourage districts to build out these definitions relative to selected evaluation instruments.
- By sharing lessons learned and relevant guidance resulting from the pilot and national examples in regular communications, we are providing examples of effective strategies for all districts to consider.

2. **Timing of Initial Implementation and Building Infrastructure**

As with any new system, districts needed to engage in extensive and time-consuming capacity-building at the beginning of implementation.

Common Findings:
- Analyzing pilot data is difficult without first recognizing that pilot districts adopted new systems on a very short timeline. The ten districts selected to participate and receive funding were not aware of their grant status until late August of 2011 – the beginning of the school year in which they were expected to begin using a new system. Once selected, districts went on to form their local advisory body, train their administrators and teachers, and ensure that new protocols were in place to execute the evaluations.
- The EPAC and RUGSE reports point to this delayed timeline when explaining why in most cases districts did not execute the number of observations that were expected. For example, the RUGSE Report notes that "The observation data…represents a good faith effort by the districts to comply with [state] requirements in spite of a late start and the need for everyone to learn how to conduct observations in the new system."
- The MET Study was designed to occur over three years; this demonstrates that the activities required to improve evaluation systems build upon one another over time.

Department Takeaways and Actions:
- Perhaps the most significant early EPAC recommendation, which the Department acted upon and was codified in the TEACHNJ Act, was to delay statewide implementation of new evaluations (originally scheduled for 2012-13) and provide another year of study and capacity-building. We recognize the time commitment required for appropriate training and implementation of new evaluation instruments, and sought to provide additional time while continuing to move swiftly toward improvements for all educators and students.
- The EPAC also encouraged the Department to create a list of state-approved educator practice instruments from which all districts could choose to help expedite selection. We began that process in 2012 and have published the list on our website. This has allowed districts flexibility to work with a variety of instruments while receiving assurance that the instrument meets a series of criteria outlined in the law.
- Both the additional year of piloting and state approval process for instruments represent a markedly different course than a number of other states, which are piloting only one year if at all, and mandating one instrument for all districts.
- Pilot districts built essential infrastructure by choosing evaluation instruments and data management systems which they will use going forward; now all New Jersey districts are in the process of doing the same. The TEACHNJ Act and proposed regulations established a required timeframe for selecting instruments and conducting training during this capacity-building year to help prepare districts to implement new systems with greater fidelity and to avoid the "time crunch" experienced in the pilots.
- In the second year of the teacher evaluation pilot, we provided a toolkit to districts with various guidance documents, templates, and communications materials. We are receiving feedback on this toolkit and will provide enhanced guidance for teachers and principals for statewide use in the coming months.

3. Fidelity and Accuracy

A major challenge when using a new instrument is for observers to internalize and correctly implement the many domains and components of the new system.

Common Findings
- When looking at survey data and aggregate observation scores from the EPAC and RUGSE reports, it is clear that there was more score differentiation in some districts than others.
  - The EPAC Report identifies several factors that may explain these mixed results, from variations in length and quality of training provided by instrument vendors to district approaches to training to interpersonal dynamics within schools and districts.
  - The RUGSE Report notes that most teachers were assigned scores that were relatively strong but not at the top of the scale; smaller but substantial numbers of teachers received at least some ratings that were below the midpoint of the respective observation scales. Greater differentiation among teachers is necessary to paint a more accurate picture of educator quality; research studies often indicate that strong implementation of new systems leads to such differentiation.
- The MET Study finds that the use of additional observations and multiple observers helps increase reliability with ratings; shorter observations (at least 15 minutes) show promise for

including additional observers without excessive time burdens. Thorough training and certification of observers is also critical.

Department Takeaways and Actions:
- The second year of the teacher evaluation pilot includes changes to training timelines and observation requirements. Cohort 2 districts are conducting some double-scored observations, using external observers, and taking steps to assure observers possess thorough knowledge of the instruments in use. We are analyzing initial results from these procedures in the current pilot year to make decisions for the future.

4. Collaboration and Communication

Educators must be involved with all stages of evaluation improvements, and clear and open channels of communication are paramount for ensuring educator engagement and trust.

Common Findings
- The EPAC and RUGSE reports point to the collaborative and measured process adopted by the Department, inclusive of stakeholder feedback and recommendations.
- Many EPAC members commented on the usefulness of time to collaborate during regular meetings; representatives from pilot districts benefited from sharing experiences and the group grounded its recommendations in facts from the field.
- RUGSE Report survey data indicates that administrators view the new system more positively than teachers do, but many teachers still see the value in new evaluations. These responses may reflect different levels of training and exposure that administrators and teachers experienced in the first year, suggesting that the more time they spend with the new system, the more likely they are to believe that it is fair and accurate.

Department Takeaways and Actions:
- While districts encountered challenges in implementing new evaluation systems, some districts conducted this work effectively.
- A transparent and collaborative approach emerged as a contributing element to districts' success. The TEACHNJ Act and proposed regulations required the formation of a District Evaluation Advisory Committee (DEAC) and School Improvement Panel (ScIP) to ensure feedback loops at the district and school levels. We will provide additional guidance for the activities of these groups in forthcoming regulations and guidance.
- In the second pilot year, we have utilized monthly EPAC meetings to better provide information to and gather feedback and recommendations from the state advisory group. By facilitating open dialogue and sharing strategies to address common concerns, we are becoming better informed about strengths and weaknesses with the system. By engaging educators directly, we aim to set an example for all districts to follow as they engage their DEACs and ScIPs.

III. Next Steps

As outlined in the January Educator Evaluation Update, the Department intends to propose regulations in March that will clarify specific requirements for evaluations in 2013-14. These proposed regulations will reflect the many sources of feedback and input described in these pages. We will conduct the first wave of a major outreach initiative from March through May, including regional presentations and trainings.

The first year of statewide implementation, which will begin this fall, is the next stage in the process of improving evaluations. By no means will our work be done in September, however. Instead, we will continue to gather feedback, analyze research, and engage practitioners in a cycle of continuous improvement. Lessons learned from all districts across the state will inform future plans, including new or modified proposed regulations as needed.

While we recognize that this is difficult work, we are confident that it is work worth doing – and we are encouraged by evidence that shows we are moving in the right direction. We commend our pilot districts for leading the way and demonstrating that committed leaders and educators can accomplish great things for our students.

PS/TM/JP/E:\Communications\Memos\020513 Evaluation Research And Policy Memo-Final.Docx

c: Members, State Board of Education
     Christopher Cerf, Commissioner
     Senior Staff
     Diane Shoener
     Marie Barry
     Karen Campbell
     Mamie Doyle
     Jeff Hauger
     Robert Higgins
     Mary Jane Kurabinski
     Timothy Matheney
     Peggy McDonald
     Cathy Pine
     Megan Snow
     Ellen Wolock
     Amy Ruck
     Nancy Besant
     William Firestone
     Todd Kent
     Joel Zarrow
     CCCS Staff
     Executive County Superintendents
     Executive Directors of Regional Achievement Centers
     Executive County School Business Administrators
     Garden State Coalition of Schools
     NJ LEE Group

Appendix A: Report Summaries and Unique Findings

In addition to the common themes and Department actions detailed in this document, some of the unique findings in each report are detailed in this Appendix, along with corresponding Department responses.

[The Evaluation Pilot Advisory Committee (EPAC) Interim Report](#)

Description: This report explores activities and preliminary results from the first year of the teacher evaluation pilot (2011-12), as well as the formation, charge, and recommendations of the state advisory group during that period. The final report from EPAC will be released in late 2013.

Data Sources:
- EPAC subcommittee reports
- Pilot district reports
- Surveys and interviews of EPAC members

Unique Findings and Department Response:

Finding	NJDOE Action
New evaluation systems are helping to create a consistent definition of good teaching and providing teachers with more opportunities to engage in meaningful conversations about practice.	• Providing common terminology at state level • Encouraging districts to build out additional definitions and shared values
Time constraints, a heavy training load, and reprioritizing district activities provided significant challenges for most pilot districts.	• Delayed statewide implementation • Established capacity-building requirements for all districts
Pilot districts found value in a collaborative and transparent approach to adopting a new evaluation system, especially in the use of teacher leaders as facilitators and turnkey trainers.	• Providing additional time for collaboration in EPAC meetings • Requiring DEAC and ScIP advisory groups statewide
The growing size of the committee and meeting time constraints created frustrations among EPAC members; survey responses showed a range of opinions about the structure and usefulness of meetings. While some members felt engaged and able to share concerns, others questioned whether their participation was making a difference.	• Improving EPAC meeting structure and processes for feedback and recommendations in second pilot year
Within EPAC subcommittees, 75% of participants felt they were provided ample opportunities to give input, and teachers felt they discussed policy with administrators and Department officials as equals.	• Continuing to engage EPAC members and other educators in sharing input
At the end of SY11-12, there were several topics that had not been adequately addressed, such as student achievement in non-tested grades and subjects and the calculation of summative ratings.	• Researching these topics in expanded pilot • Discussing issues in EPAC meetings

Description: As part of the pilot program, the Department contracted with RUGSE – specifically, a research team led by Dr. Bill Firestone – to conduct an independent study of teacher evaluation activities. The RUGSE Year 1 Report covers 2011-12 pilot work and focuses on the observation practices in pilot districts. RUGSE set out to explore the status of teacher observations, teacher and administrator perceptions of observations and evaluations, and barriers and facilitators affecting initial implementation.

Data Sources:
- Teacher and administrator surveys
- Six site visits, which included interviews and focus groups
- Teacher observation data

Unique Findings and Department Response:

Finding	NJDOE Action
Data management tools are essential for implementing teacher observation systems; learning to use them was a major challenge.	• Provided capacity-building year for additional district time
More administrators than teachers agreed that the training they received helped them to understand the observation process; the extensive time administrators spent in training may have contributed to their greater appreciation.	• Requiring training activities on a specified timeline • Providing additional tools and guidance for statewide implementation
A large proportion of eligible teachers were observed at least once, which was lower than the recommended amount; participation rates varied substantially across districts.	• Required training for all districts as part of capacity-building • Establishing universal observation requirements in forthcoming proposed regulations
The process for ensuring accuracy and fidelity of implementation with observations needs improvement; educator practice instruments should meet standards of rigor and accuracy in order to gain credibility with educators and the public.	• Maintaining state-approved list of instruments for district selection • Establishing observation requirements in forthcoming proposed regulations
Districts did very little to prepare to use teacher observation data to make personnel decisions or to plan collective professional development – which was anticipated in the first year of the pilot.	• Using pilot information to create informed plans for linking evaluation results to personnel actions and professional development
Frequent and accurate communication is needed within districts and schools to ensure that all educators are informed about evaluation procedures and have the opportunity to ask questions and provide feedback.	• Requiring DEAC and ScIP advisory groups statewide • Planning additional guidance for these groups' activities

Culminating Findings from the Measures of Effective Teaching (MET) Project Three-Year Study

Description: Funded by the Bill & Melinda Gates Foundation, the MET Study was a large scale, multi-year effort to provide credible scientific evidence for the best way to measure the effectiveness of a teacher. While we recognize the limitations of any single piece of research, we are encouraged that several findings support New Jersey's plans for improved evaluations.

Data Sources:
- Randomized, controlled study groups of over 3,000 teachers across 7 districts: Charlotte-Mecklenburg (NC), Dallas (TX), Denver (CO), Hillsborough County (FL), Memphis (TN), New York City (NY), Pittsburgh (PA)
- Classroom observation instruments
- Student perception surveys
- Student achievement gains

Unique Findings and Department Response:

Finding	NJDOE Action
Effective teaching can be measured; by controlling for other factors, findings demonstrated that effective teachers continued to be labeled as such, even with randomly assigned students – and that their impact on student learning reached beyond state tests into "more cognitively challenging" assessments.	• Establishing this core principle - as linked to the importance of educator quality - as the basis of efforts to improve evaluations
Using a balanced approach with multiple measures such as student test data, classroom observations, and student surveys provides a more accurate and stable picture of a teacher's practice.	• Including multiple measures from the pilot stage • Ensuring no educator is ever rated based on one test score or observation alone
A weighting of between 33%-50% for student achievement in a teacher's overall evaluation is recommended, including multiple years of data.	• Considering this range in finalizing proposed student achievement weighting for SY13-14, to be explained in forthcoming proposed regulations
There is a large increase in reliability of teacher observation ratings when two or more observations are used, especially when they are done by different administrators.	• Changed second year pilot observation requirements to include double-scoring, additional observers, and assurances of observer knowledge in instrument
Well-designed and carefully administered student surveys are one of the most reliable indicators of teacher effectiveness; such surveys show promise and can provide meaningful feedback to teachers.	• Worked with multiple SY12-13 pilot districts to employ student surveys • Continuing to learn about their impact locally and nationally
The use of videos to train and provide feedback for teachers shows promise and can result in cost-savings, accessibility, and improved quality.	• Continuing to study video effectiveness • Considering ways to support districts in their use

Appendix B: Timeline of Major Evaluation Activities, 2010-13

	2010		2011				2012				2013			
	Fall	Winter	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter

TEACHNJ Act passed

NJDOE Announces SY11-12 Teacher Evaluation Pilot

NJDOE Announces SY12-13 Evaluation Pilots & Requirements

Districts form DEACs

EPAC Interim Report Released

EETF Created By Exec. Order

EETF Releases Recs.

NJ receives $38m Race to the Top Award

NJDOE Proposes First Round Of Eval. Regs.

RUGSE Year 1 Report Released

All Districts Launch Improved Evaluations

Sept — Mar — Jun — Sept — Dec — Feb — Jul — Aug — Sept — Oct — Dec — Feb — Mar — *Continued Capacity-Building* — Sept — *Ongoing Improvements*

11 Districts Selected for SY11-12 Pilot

NJDOE Announces Pilot Study Partnership w/ RUGSE

NJDOE Releases Pilot Teacher Eval. Toolkit

NJDOE Proposes Second Round of Eval. Regs.

Eval. Pilot Advisory Comm.(EPAC) Selected

24 Districts Selected for SY12-13 Pilots

Districts Select Evaluation Instruments

NJDOE Launches Outreach Initiative

RUGSE – Rutgers University Graduate School of Education
EETF – Educator Effectiveness Task Force
NJDOE Activity – Regulation and Policy Development, Communications

Key

EETF	EPAC	NJDOE Activity	Pilot Activity	External Pilot Study