

2011-12

Evaluation Pilot Advisory Committee
(EPAC)

Interim
Report



State of New Jersey
DEPARTMENT OF EDUCATION

TABLE OF CONTENTS

Executive Summary	1
Pilot District Experiences	1
EPAC Activities	2
Conclusions	2
Introduction.....	4
Part One: Educator Evaluation Reform in New Jersey:	
Background and Context	4
Rationale for Evaluation Reform.....	4
Educator Effectiveness Task Force.....	5
New Jersey Department of Education’s Educator Evaluation Initiative: Goals and First Steps	6
Part Two: Teacher Evaluation Pilot – Cohort One	7
Introduction	7
Evaluation Instrument Implementation.....	7
Professional Development	15
Teacher Support	15
District Evaluation Pilot Advisory Committees (DEPACs)	16
Measures of Student Achievement	17
Cohort One Takeaways	18
Part Three: The Work of the Evaluation Pilot Advisory Committee (EPAC) in SY11-12	20
Formation and Charge	21
EPAC Meetings	21
EPAC Successes	21
EPAC Challenges	23
EPAC Subcommittees	24
Subcommittee Recommendations	24
Outcomes of Subcommittee Recommendations	27
Unanswered Questions and Next Steps	27
Conclusion	28
Appendices	29
Appendix A: Teacher Evaluation Pilot Participants, 2011-2012	29
Appendix B: Evaluation Instruments and Data Management Systems by District.....	30
Appendix C: EPAC Members, 2011-2012	31
Appendix D: EPAC Meeting Background Reading, 2011-2012.....	32
Appendix E: EPAC Presentations	33
Appendix F: EPAC Subcommittee Chairs, 2011-2012.....	34
Appendix G: Subcommittee Reports, 2011-2012	35

Executive Summary

New Jersey is preparing to implement a statewide educator evaluation system in the 2013-14 school year. As part of this process, an Evaluation Pilot Advisory Committee (EPAC) was convened in September 2011. The EPAC consisted of 22 appointed members from various stakeholder groups. Its charge was to make recommendations to the New Jersey Department of Education (“the Department”) using current evaluation research and lessons shared by districts who were piloting new evaluation instruments. Monthly meetings organized by the Department were attended by the 22 appointees plus representatives from pilot districts. This interim report summarizes the key lessons learned from pilot districts as well as recommendations from the EPAC’s activities during the 2011-12 school year.

Pilot District Experiences

Pilot districts provided a window into the challenges of implementing a new evaluation system. A more in-depth look at their experiences reveals several areas that should be the focus of districts and the Department before and during statewide implementation in September 2013. Among these is that effective training of observers in the teacher practice instrument is crucial. Districts must provide good training and accuracy checks that will ensure consistent and valid observations of teachers. Furthermore, when districts develop a process for evaluators to demonstrate that they can accurately use the instrument, their teachers will be reassured that observations will be performed competently and objectively. Finally, when districts provide good training, it will allow observers to perform more efficient observations earlier in the year, thereby reducing some of the time pressure to which all districts are sensitive. The Department should support districts in their efforts to provide excellent training to teachers and administrators and require instrument providers to offer and districts to use rater accuracy checks.

Despite facing a variety of unique challenges, teachers and school leaders are seeing the benefits of adopting a high quality evaluation instrument.

In addition to providing effective evaluation instrument training, districts should leverage their District Evaluation Advisory Committees (DEACs) to create a transparent process during planning and implementation. Furthermore, clear and consistent communication will be essential to success, as will recognizing teachers as partners and facilitators in this work.

Districts were faced with a number of challenges during the first-year pilot. Time constraints, amplified by late notice of their grant awards, inhibited pilot districts’ efforts to complete the required number of observations. Intense focus on training delayed work on other important areas of the evaluation such as developing measures of student achievement in non-tested grades and subjects and providing effective professional support for teachers across the performance continuum. Going forward, and as soon as possible, the Department should provide clear guidance to districts in these areas.

EPAC Activities

EPAC members were exposed to a large amount of information from both national experts and pilot districts and asked to provide feedback and recommendations for statewide policy development. Most notable of the EPAC's recommendations was that the Department delay full implementation for a year, moving the start date from September 2012 to September 2013. This recommendation to delay was acted on by the Department and codified in the [TEACHNJ Act](#). It led to an expanded pilot program and a capacity-building year for all other districts in New Jersey.

In addition to this overarching recommendation, EPAC's subcommittee work was important in providing specific ideas to the Department regarding teacher and principal evaluation. Subcommittees considered several sources of information including presentations and research papers, feedback provided by pilot districts, and the individual professional experiences of the committee members.

Recommendations from subcommittees for teachers in the areas of Special Education, English Language Learners, and early childhood centered on ensuring appropriate evaluation instruments were developed for these groups. These subcommittees noted that the specific needs and developmental levels of students make the use of traditional test data challenging. In addition, they recommended that teacher evaluation must take into account the specific methodologies used and challenges faced while teaching diverse students.

The principal evaluation subcommittee made several practical recommendations on implementation of evaluation instruments. Some of these recommendations were incorporated into a principal pilot program that was launched in September 2012. Others continue to inform the Department's work on principal evaluation.

Several key components written into the second year of the teacher evaluation pilot came from the teacher practice subcommittee. These included allowing districts to continue to choose from a list of approved evaluation instruments, requiring comprehensive observer and teacher training, and using informal observations to provide feedback for growth and support.

Practical recommendations also came from the summative rating subcommittee, which suggested that vendors be asked to provide clear guidance on summative rating calculations and that districts be required to develop assessments in non-tested grades and subjects. The subcommittee for professional development and school culture strongly supported the development of collaborative structures and appropriate administrative support to enhance teacher learning. In addition, the subcommittee suggested that districts develop teacher leadership roles for those receiving "highly effective" ratings.

Conclusions

There is still a great deal of work to be done. Clarity is especially needed in the following areas: measuring achievement of students in non-tested grades and subjects; calculation of summative ratings; and processes that will ensure observers use evaluation instruments accurately and appropriately. However, just because the work is difficult does not mean it should not be done. Initial reports from pilot districts provide reason to hope that we are moving in the right direction. Despite facing a variety of unique challenges, teachers and school leaders are seeing the benefits of adopting a high quality evaluation instrument. They are witnessing a transformation in the type and quality of conversations surrounding teacher practice and student learning. Some districts are more effectively differentiating between the

¹ *Teacher Effectiveness and Accountability for the Children of New Jersey Act; New Jersey's "Tenure Reform Law" enacted on August 6, 2012*

performance of teachers, even in these early stages of implementation, and are developing new systems of recognition for teachers with excellent practice and professional support for everyone. These districts are the vanguard for others in the state who are just beginning this work. Lessons learned from the pilots have been invaluable and continue to inform the Department's work, allowing it to make wiser and more practical recommendations, such as allowing districts flexibility in the type of evaluation instruments they use.

Additionally, the first year of the EPAC's activities have demonstrated that dozens of strong educators, educational leaders, and officials from the state can come together, learn from one another, and strive to make a very difficult task possible. Even when tensions were high and frustrations many, dedicated professionals continued to make deliberate progress. The Department is approaching evaluation activities carefully and thoughtfully with the continued guidance of the EPAC in 2012-13.

It is this type of perseverance and continued collaboration between the Department and educators throughout New Jersey that will be crucial in creating an environment conducive to the growth and success of such an ambitious program of reform. It is in this spirit of collaboration that difficult work must be done if we are to make educator evaluation reform more than just a passing fad, but a lasting legacy that will benefit all of New Jersey's children for years to come.

Introduction

This report is the first of two that will represent the work of New Jersey's Evaluation Pilot Advisory Committee (EPAC). The final report from EPAC will be released later in 2013; this interim report summarizes the key lessons learned from the 2011-12 teacher evaluation pilot and provides recommendations for statewide roll-out of a more effective educator evaluation system. The writing and recommendations within it are the result of a collaborative effort between representatives of the EPAC and the Department.

The report includes three sections. Part One outlines the background and context for evaluation reform in New Jersey. Part Two explores the activities and preliminary results from Cohort One of the teacher evaluation pilot (2011-12). Part Three describes the formation, charge, activities, and recommendations of the EPAC.

Many sources were used to inform the lessons learned and recommendations of this report including the EPAC's subcommittee reports, pilot district reports, and surveys and interviews of EPAC members.²

While this report attempts to capture EPAC's work as thoroughly and fairly as possible, any particular recommendation or viewpoint that it contains does not necessarily represent the opinion of all members of the committee.

Part One:

Educator Evaluation Reform in New Jersey: Background and Context

Rationale for Evaluation Reform

New Jersey, like the majority of states across the country, is undergoing comprehensive educator evaluation reform. This work stems from a growing body of research and national education priorities that emphasize the importance of teacher quality on student achievement – and the inadequacy of old evaluation systems. The findings of *The Widget Effect*,³ a 2009 study of evaluation policies and practices in 12 school districts across four states, found most teachers were rated good or great in their evaluations, despite the fact that significant student achievement gaps and poor graduation rates persist. Evaluation reforms seek to provide schools with effective systems of evaluation that encourage all teachers to engage in a cycle of continuous improvement.

² All quotes taken from these sources are purposefully kept anonymous to protect the confidentiality of the districts and educators involved.

³ 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.

At the national level, the Obama administration's education reform agenda identifies improving educator effectiveness as a key priority. Both Race to the Top and the Elementary and Secondary Education Act (ESEA) flexibility process, for example, have required state commitments to reforming evaluation systems.

Like all states undertaking this work, New Jersey's current evaluation system does not clearly differentiate performance among educators, provide adequate feedback and targeted opportunities for professional development, or produce quality data to inform staffing decisions. Therefore, improving educator evaluation was a critical element in both the state's \$38 million Race to the Top III award and its approved ESEA waiver.

Educator Effectiveness Task Force

As the first key step in New Jersey's evaluation reform effort, Governor Chris Christie created the Educator Effectiveness Task Force (EETF) by Executive Order in September of 2010. The EETF was composed of nine members with education experience. It was charged with developing recommendations to guide the creation of an evaluation system that utilizes both student achievement and educator practice.

In March 2011, the EETF presented its [report](#)⁴ its report with the following high-level recommendations:

- Teacher Evaluation
 - o All measures used to assess teacher effectiveness should be linked to achievement.
 - o In the initial roll-out, half of the teacher's evaluation should be based on measures of student achievement and half on measures of teacher practice.
 - o Over time, the state should increase the percentage of the evaluation contributed by measures of student achievement.
 - o The evaluation system should include four rating categories: Highly effective, effective, partially effective, and ineffective.
- Principal Evaluation
 - o Half of the principal's evaluation should be based upon measures of student achievement and half upon measures of professional practice.
 - o 10 percent of the evaluation should measure principal practice in retaining effective teachers.
 - o The evaluation system should include four rating categories: Highly effective, effective, partially effective, and ineffective.
- Next Steps
 - o The Department should solicit feedback from the State Board of Education and other stakeholder groups.
 - o The Department should convene sub-groups to develop recommendations for student achievement measures for teachers of special populations and non-tested subjects and grades.
 - o The Department should develop detailed recommendations for piloting the evaluation system in selected districts.

New Jersey Department of Education's Educator Evaluation Initiative: Goals and First Steps

Informed by the guidance of the EETF, the Department's goals for a reformed educator evaluation system are to:

1. Increase student achievement;
2. Accurately assess the effectiveness of teachers and differentiate between those who are excelling and those who are struggling;
3. Improve the effectiveness of educators (as defined by professional practice and student outcomes) through a system that:
 - a. Clarifies expectations for teacher practices and the metrics that will be used in evaluation, and
 - b. Provides meaningful feedback to identify strengths and weaknesses that will result in a relevant growth plan for teachers; and
4. Facilitate school- and system-wide collaborative cultures focused on continuous improvement by:
 - a. Providing a common vocabulary and understanding of what teachers need to know and be able to do to be effective;
 - b. Promoting the use of student and teacher data to improve teacher practice and student learning; and
 - c. Fostering a culture of openness and sharing where educators work together to improve their collective work.

As the first concrete step toward accomplishing these goals, the Department launched a teacher evaluation pilot program in 2011. At the onset of this initiative, Commissioner Chris Cerf convened the Evaluation Pilot Advisory Committee (EPAC) to help inform the development of an improved statewide educator evaluation system.

Evaluation reforms seek to provide schools with effective systems of evaluation that encourage all teachers to engage in a cycle of continuous improvement.

Part Two:

Teacher Evaluation Pilot – Cohort One

Introduction

Based on the recommendations offered in the March 2011 [Educator Effectiveness Task Force Report](#), the Department launched a teacher evaluation pilot program in the fall of 2011. This project was designed to enable the experiences of pilot districts to inform the development and implementation of the evaluation system to be launched statewide.

Having successfully applied for grants through a competitive Notice of Grant Opportunity (NGO), 10 districts were selected to participate in the pilot program, splitting \$1.1 million in funds. Newark Public Schools also participated in the pilot using funding provided by another grant. In addition, 19 schools receiving federal School Improvement Grant (SIG) funds were required to participate. A full list of pilot participants is provided in [Appendix A](#). This section of the report uses information gathered only from the 10 districts awarded grants through the NGO.

All pilot district participants were required to implement the following elements of a teacher evaluation system during the 2012-13 school year:

- Thorough training of evaluators and teachers in effective teaching practices based on professional standards;
- Annual teacher evaluations that include multiple observations and result in clear, actionable feedback for improvement;
- Multiple measures of teacher practice and student performance, proven to be valid and reliable, with student academic progress or growth as a key measure;
- A summative rating that combines the scores of all the measures of teaching practice and student achievement;
- Four summative rating categories that clearly differentiate levels of performance; and
- A link from the evaluation to professional development opportunities that meet the needs of educators at all levels of practice.

This section outlines key components of the evaluation pilot and lessons learned about each. Important takeaways for all districts to consider are provided at the end.

Evaluation Instrument Implementation

Districts chose to use a teaching practice evaluation instrument – including an evaluation instrument – that met requirements specified by the Department. In some cases, the instrument was similar to that which the district had been using previously. Each of the instruments allowed for a four-point rating of a teacher's practice. The instruments that each pilot used are listed in [Appendix B](#).

Training Administrators

Districts used a variety of approaches for training observers, including in-person training by a vendor, whole group activities, video training, and in-district instructional leaders. All but two districts included teacher leaders in their administrators' training sessions. These included academic coaches, team leaders, and other educators tapped to be train the rest of the teachers, i.e. turnkey trainers

Most districts reported that training was very successful for administrators and that vendors provided excellent training. Project directors noted that "training was very comprehensive," and that there was "excellent support from our Evaluation Tool Provider." Districts highlighted the consistency, transparency, and commitment with which they implemented training for observers. All pilot districts claimed that they had fully trained more than 80 percent of their observers.

However, despite reports of thorough training from all districts, only two-thirds of them reported that more than 80 percent of their observers were able to show proof of mastery.⁵ In fact, one district confessed to having only 40-60 percent of its evaluators demonstrate mastery in the new tool and two districts did not report on this aspect of the training. This indicates that in certain cases, a disconnect exists between what districts consider to be full training and the actual outcomes of that training. This finding is reflected in the opinions of some teachers. One commented that, in his district "administrators are not sufficiently trained." Another, from a different district said, "Observers need to be more consistent."

Teacher leaders were "essential in increasing the understanding and credibility of the evaluation process."

Training Teachers

"The district created a comprehensive book of the definitions of every element and standard at each performance level and school level along with a glossary. This assisted in bringing everyone into the conversation with the same understandings. It took a tremendous amount of work but was more than worth the value of that work."

- Project Director

As with their administrators, districts spent significant time on training teachers – 12-18 hours in most cases. They also used a variety of methods, including bringing in external trainers, using video and online instruction, leveraging professional learning communities, having mini-sessions within school hours, and calling extensively on turnkey trainers. One district developed a "Definitions Book" that was of great use in getting everyone on the same page. (see box)

Districts that used turnkey training generally agreed that this was a very successful approach to building training capacity. One project director noted that the teacher leaders were "essential in increasing the understanding and credibility of the process," and another said that they became, "a tremendous resource at the school level."

All districts used whole group instruction at some point in training. Of note is one district that closely monitored its training process by collecting feedback from teachers using a variety of forums and methods highlighted below:

Feedback Forums

Teacher-leader panel
Department meetings
Grade level meetings
Faculty meetings
Professional development days

Feedback Methods

Open-ended questionnaires
Jig saw activities
Gallery walks
Discussion

⁵ "Mastery" in the case of the evaluation instrument most widely used by the pilot districts involved passing a test to demonstrate that the evaluator could apply the instrument with accuracy, i.e. certification tests for Danielson's Framework for Teaching.

The focus on training seems to have paid dividends in most districts. Most stated that teachers in their schools felt that they had a very good grasp of the evaluation instrument. One project director reported that “Teachers felt much more comfortable with the model once they were trained.” Even though this is seemingly obvious, it underscores the importance of getting training right as soon as possible.

Understandably, getting it right did not happen overnight and teachers highlighted some of the challenges. In one case a teacher noted, “It took a full year to get everyone on the same page.” Another claimed that even though the turnkey trainers were knowledgeable, the 30-minute sessions conducted after school were not long enough and not willingly attended by staff. Some external trainers were well received, with a project director stating that “teachers wanted to learn from her,” but this was not always the case. One pilot teacher noted that even though the situation improved, initially, consultants were “coming in and out and telling us different things.” Other external trainers were poorly received by teachers and had “very little credibility among our staff.” One teacher went so far as to use the words “arrogant” and “bullies” when referring to trainers from a particular company.

By the end of the pilot year, many initial obstacles had been overcome and districts reported that their teachers were well-trained. However, many educators echoed that the same level of training needs to be applied to all staff, including new hires. One teacher cautioned that “new hires have to be continuously trained.”

Time Constraints

Woven throughout comments from project directors and teachers alike was the challenge of time constraints. Only one district claimed that time was not an issue and was able to train all staff by November 2011, a feat that it attributed to “proactive and systematic actions.”

However, having only been notified in September that they would be receiving grants, most districts struggled with timely implementation.⁶ One district stated “the late timing of notice from the Department regarding our own pilot status eliminated many dates available for trainings.” Another district felt that the quality of the training suffered because of the tight timeframe. “Had we more time, we would have done more extensive training and included a trial period on the evaluation so that evaluators felt more confident in the initial phases.” Yet another pilot identified the emphasis on anti-bullying training that year as a hindrance to effective training in evaluation.

Not only was time in short supply, the training drained other resources including administrative staff. One director noted that “it was difficult for administrators to be out of the building for all of the trainings.” Having teacher leaders in training also placed a strain on buildings and incurred costs for substitutes.

Unforeseen Benefits

Pilot project directors found quality training for evaluators and teachers to be a crucial component of success in their first year. In addition to providing an instrument for better evaluations, training on the evaluation instrument created a school culture in which a shared educational language and set of expectations developed. In some instances, staff found they gained more than just knowledge of an evaluation instrument but “a deeper knowledge of the teaching and learning process.” The conversations that took place during and after training became “the greatest successes,” in

One district saw a great benefit in bringing together many separate school initiatives, seeing them through the lens of teacher evaluation.

⁶ Due to state grant procedures and the competitive number and nature of grant applications, the Department was not able to notify districts of awards as early as originally intended.

the words of one project director. Another administrator highlighted the teachers' pride in being a part of the pilot and that this "influenced their cooperation and enthusiasm for success." One district saw a great benefit in bringing together many separate school initiatives, seeing them through the lens of teacher evaluation.



Key Lessons Learned

While there are obvious challenges in training in the new instruments, inadequate training has significant ramifications for the quality of teacher observations. The Department must provide clear guidance to ensure that before observing teachers, observers must demonstrate mastery of the instrument through an established process. The Department should consider providing models of best practice that demonstrate how districts may ensure that training is carried out in a timely and effective fashion. Districts must reprioritize scheduled meetings and professional development days to ensure that adequate time is provided for effective training.

The Observation Process and Professional Dialogue

Improved conversations between staff and administrators were generated by the implementation of a new evaluation instrument. One administrator stated that, "Teachers became much more aware of their practice and collaborated with observers on reflection of such." The new instrument also seemed to provide a useful tool to bring clarity to many aspects of teaching. A director noted that it allowed deeper insight into teaching and learning and, "its interrelationship with the district's curriculum, instructional planning, assessment, and instructional practices." Another expressed that one of the values of evaluation reform is, "in finding ways for teachers to focus on feedback and identified areas of growth generated from the observation process rather than the ratings." The consistent framework that a new instrument provided led one administrator to observe that, "Teachers and evaluators felt that the observation process was very objective."

Even though some teachers agreed with the above statements, and had received positive observations and pre- and post-conferences, others noted their concerns. One teacher felt that the "average teacher was not very comfortable with the observations." A second noted that "some had conflicts with some administrators," a third that "certain administrators make people feel uncomfortable," and a fourth that they have a principal who "won't buy in" and "waits months for a conference." Finally, some teachers remarked that even though they had been observed, they had not had a post-conference.

Evaluation Instruments and Rating Teacher Performance

The 10 pilot districts receiving Department grants used one of four evaluation instruments and chose a data system that allowed the recording and management of observation data. These are shown in Figure 1. The combination of evaluation instruments and data management systems used by each district can be found in [Appendix B](#). Each evaluation instrument was used to provide a rating of one to four for teaching practice based on classroom observations and other components of a teacher's work, including planning and professional contributions to the educational community.

Evaluation Instrument			
Danielson	Marzano	McREL	James Stronge
6	1	2	1

Data Management System			
Observation	McREL	Oasys	Teachscape
3	2	1	4

Fig. 1: Evaluation instrument/data system type and number of districts using them.

One of the goals most fundamental to evaluation reform is to produce a system that will fairly identify and classify teacher performance. One of the results of such a system will be to reduce the so-called “widget effect” in which the vast majority of teachers are rated great or good and subsequently regarded as interchangeable parts in a school district. Modern teacher practice instruments, if implemented correctly, are designed to make such classification more possible. For instance, commonly heard during training in Danielson's Framework for Teaching is that the highest rating is something a teacher may achieve only sometimes, like “taking a trip to Hawaii,” in the words of one teacher.

Most pilot districts did make an effort to apply their teacher practice instruments appropriately resulting in fewer teachers than usual getting a top rating. This switch in expectation and concurrent decrease in ratings had consequences. One teacher noted that no one in her school was rated highly effective and that the teachers felt “the bar was too high.” Another thought that observation ratings became personal in that “some teachers think the lower rating is a punishment for something they did that the evaluator did not like.” One teacher explained that veteran teachers who are rated “basic” are angry because “they think their years of experience should make them ‘distinguished.’”

Despite sharing these comments, teachers in district pilot advisory committees seemed to understand the theory behind the four category system with higher standards. Echoing the statement above, one commented “Danielson says no one lives in the land of the distinguished.” Teachers in the field want a more granular rating than a one to four number. To provide this service, one teacher remarked that “some administrators use ranges within a level to give more specific feedback.”

Key Lessons Learned

Districts should provide guidance and communication tools to their principals and supervisors that allow them to provide context and set expectations for the new observations. In addition, they must put in place measures to ensure that observers conduct observations and conferences with the utmost professionalism and objectivity. The Department may support this work by sharing best practices and supporting professional development opportunities offered by stakeholder groups that help evaluators have “difficult” conversations with teachers regarding their observations.

⁷ 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.

Completion of Observations

While half of the first-year pilot districts claimed that all of their teachers had been observed at least twice with the newly adopted evaluation instrument, none of the districts who submitted data to the state met this mark for full observations. The data suggests that districts may have counted any type of observation including walk-through's and informal observations when reporting their numbers. Actual numbers of full observations along with the numbers of evaluators can be seen in Figure 2.

One of the reasons for these low averages of full observations may be explained by a project director who simply stated, "Training was not completed until March." The delay in training meant that the administrators only had two months to observe all teachers and provide end-of-year evaluations. Another project director summed up the stress of coping with increased requirements and a new system as having "an increased workload under a decreased timeline for observations." Another district referred to the pressure of "working non-stop around the clock to get all of our evaluations done," with the result that, "The wigit (sic) effect started to happen. It was more about getting everything done than it was about meaning."

District	Number of Teachers	Number of Observers	Observations Per Observer	Observations Per Teacher
1	180	10	29 (8-100)	1.6
2	51	4	13 (8-18)	1
3	1772	94	32 (1-98)	1.7
4	344	27	15 (1-36)	1.2
5	385	20	19 (5-37)	1
6	102	11	9 (1-25)	1

Fig. 2: Observations in pilot districts reporting information.



Key Lessons Learned

The NJDOE should consider challenges of time and capacity when deciding upon the appropriate number of observations it requires for each teacher. This number must provide an adequate sample of a teacher's practice while guarding against causing observers to cut corners in the process. In addition, the NJDOE should provide guidance for 2013-14 in the form of best practices taken from districts that were successful in striking this balance.

Ratings from Observations

One of the objectives of an improved evaluation system is to differentiate among educators who practice at various levels of effectiveness. Preliminary data from first-year pilot districts indicate that this may have been addressed but that there is substantial work to be done.

Even though more observations per teacher would provide a statistically more useful sample size, first-year data provides a useful snapshot. While some differentiation between teacher performance levels occurred in some districts, in the majority, there was a heavy weighting towards the effective and highly effective ratings. The observation data of three districts is depicted in Figures 2, 4, and 5. Districts A and B used the Danielson Framework for Teaching and District C used a different evaluation instrument (not specified for confidentiality purposes).

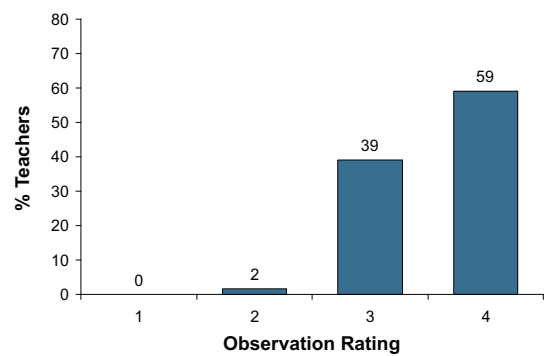


Fig. 3: District A observation data results (Danielson)

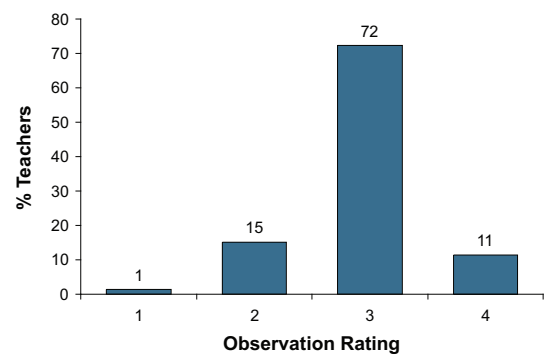


Fig. 4: District B observation data results (Danielson)

All of these figures show that a diminishingly small number of teachers were given an ineffective rating. Districts A and C had no teachers earn this rating (the zeroes are not rounded figures). In half of the districts for which there is data, no teachers were rated ineffective.

In all cases, teacher ratings skew heavily towards the upper two categories. District B provided the best example of a moderation of this effect but still rated 72 percent of its teachers effective and 11 percent of them highly effective. District B also had the highest percentage of teachers rated partially effective at 15 percent.

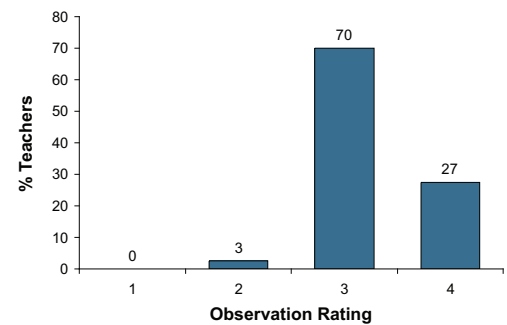


Fig. 5: District C observation data results

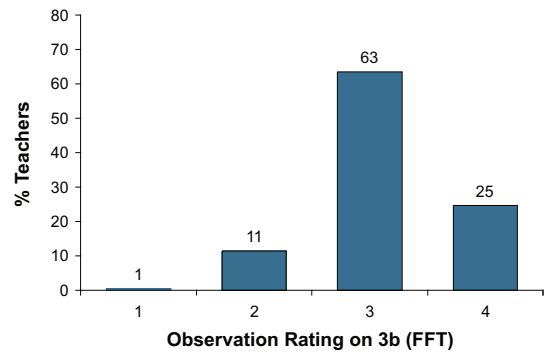


Fig. 6: Ratings for the “questioning and discussion” element in Danielson’s instrument across four pilot districts

Of the four districts reporting on the Danielson instrument, the most promising data in terms of distribution across the four required proficiency levels correlated with component 3b: Using Questioning and Discussion Techniques in Instruction. An average of these ratings across the four districts that used Danielson can be seen in Figure 6.

Factors Affecting Observation Ratings

Several factors, other than teacher performance, may have contributed to the distribution of ratings given to teachers in the pilot districts. Some of these are described below.

Accuracy of Observations

While districts found it challenging to guarantee rater accuracy in such a short period of time, seven out of 10 districts reported that they used at least one process to check for accuracy. All districts used video exemplars during training and some used additional processes including:

- Practice with a master scorer/vendor coaches
- LEA-developed definitions book plus discussion at administrative meetings
- Ongoing feedback on observation reports
- Double-scoring (same lesson scored by two observers)
- Superintendent review for consistency
- University of [Washington's 5D Assessment](#) rater accuracy score

Number of Ratings

Fewer than two of the observations per teacher required by the pilot grant were performed in each district and the number given in each district varied. This makes a district-to-district comparison more difficult. However, District 2 shown in Figure 4 – with the highest number of observers and observations per teacher – showed the most normally distributed data of all districts. The comfort level of an observer assigning any given rating in a new system is an important factor. This comfort level is likely to improve as observers complete more observations and become more familiar with the new evaluation instrument.

Relational Factors

Related to an observer's comfort with the system itself, is the suggestion that some administrators struggled to provide anything less than "good" ratings to teachers they know and have worked with closely for years. One pilot director noted this as an area of concern commenting that, "Personnel is personal. This is part of the larger transition of principal from manager to instructional leader. The seeds need to be cultivated for this change." Although the data doesn't tell us how widespread this issue is, it merits further study.

Training

As mentioned previously, evaluator training was completed by all districts, but in some instances, this may not necessarily have translated to accurate use of the instrument. In part, this may have been due variable training quality. One administrator noted that there were deficiencies in certain external trainers and "gaps that occurred between training and actual observations." This may have been compounded in some cases as trainees grappled with a whole new way of thinking about evaluation, labeled by one district as "a paradigm shift in the content of the evaluation model."



Key Lessons Learned

The Department must establish guidelines for how districts should check for rating accuracy in observers. Instrument providers and districts using "homegrown" tools must make certain that their instruments have formalized systems to check for reliability and accuracy of observations. Districts must continuously assess and improve training and implementation of evaluation instruments to ensure they are administered with fidelity.

Professional Development

Many pilot districts are optimistic that focusing on a high quality evaluation instrument will improve practice by generating “increased awareness of strengths and areas needing improvement.” However, for most districts in the first pilot year, lack of timely data from observations, lack of adequate time to train and do the observations, and the focus on learning the new instrument inhibited their ability initially to develop meaningful links between observations and professional development.

One district recognized that this should become easier in the future - “as the instrument becomes better understood, the linkages [to professional development] become clearer.” One district has begun this work and asked all of its teachers to reflect on their performance based on observation data. Another district has moved a little further; with information gathered from their teacher observation scores, administrators noted weakness for many teachers in using questioning techniques. Therefore, the district plans to require each teacher to embed higher level questioning in their 2012-13 Professional Improvement Plans. Another district has developed four new courses “that align with a different domain.” The district most advanced in this area is one that offers 30 daily workshops which are aligned to their evaluation rubric. These are optional for some teachers and required for others “based on the evaluation level of the teacher.”

The variability in how districts have moved towards effectively linking professional development to instruments used to observe teachers is mirrored in comments from teachers in these districts. Some teachers acknowledged there was some movement in a positive direction but according to the educators surveyed, only one of their districts is systemically doing this well. One teacher acknowledged that this is understandable and commented that it is “unrealistic to think that next year the districts can do everything.”

Key Lessons Learned

Identifying and providing the right professional development to teachers is a cornerstone of New Jersey’s evaluation initiative. School districts should proactively review observation data throughout the year in order to create focused professional development plans that will enhance teacher practice. School leaders should prioritize this work and customize professional development for both individual teachers, teams of teachers such as in professional learning communities, and the school as a whole.

Teacher Support

Teachers Whose Practice is Partially Effective or Ineffective

While two districts had nothing firm in place to address teachers rated partially effective, or ineffective, others had a variety of interventions planned for them. These included support from the building principal and peer groups, training mentor teachers to assist them, providing targeted professional development including out of district and online courses, and one-on-one coaching with content area and practice experts.

Teachers Whose Practice is Effective and Highly Effective

Four districts did not have clear plans regarding teachers rated effective or highly effective. However, other districts had a variety of ideas. These included creating teacher leader positions

with stipends and allowing these teachers to share their expertise with staff either at faculty meetings, professional development days, or through classroom videos. One district intended to let its teachers rated highly effective teach mini-courses.



Key Lessons Learned

Districts' newly formed district evaluation advisory committees (DEACs) consisting of representatives of each stakeholder group, can be a powerful tool for successful implementation of a new evaluation system. To be effective, they must be broad-based, transparent, and provide open lines of communication. The Department needs to identify best practices of pilot districts that used DEPACs most effectively and share this information with districts.

District Evaluation Pilot Advisory Committees (DEPACs)

As part of the evaluation pilot, districts were required to convene District Evaluation Pilot Advisory Committees (DEPACs) to help guide implementation. Most districts agreed that despite running into scheduling challenges, their DEPAC was key for keeping all stakeholders informed as their evaluation systems were being developed and implemented. The committees gave districts a way to show that work was not just being done behind closed doors by the administration. In one district, their DEPAC fostered “collaboration and investment in the process.” One district said the DEPAC was used to create “complete transparency” and included “teachers, parents, board of education members, and administrators.” This idea of open communication was echoed by another project director, who said that their 20-member DEPAC was sometimes unwieldy but that the size “created a more transparent process” and “stronger buy-in to the program.”

Part of the DEPAC role was to “share updates, provide data, present implementation challenges, and support open dialogue,” but in addition, some districts utilized the expertise of its various educational professionals. One district said that it was a good committee to “run ideas by before implementing them.” Another noted that frequent early meetings “allowed us to address issues that arose as a result of the new pilot.” Yet another credited the DEPAC with their overall success. “We were able to fully implement our plan. A major reason for this was proactive, systematic planning of the implementation process from the very start of the process.”



Key Lessons Learned

As shown by DEPACs in pilot districts, district evaluation advisory committees (DEACs) required by TEACHNJ, can be a powerful tool for successful implementation of a new evaluation system. To be effective, they must be broad-based, transparent, and provide open lines of communication. The Department needs to identify best practices of pilot districts that used their committees most effectively and share this information with districts.

⁸ DEPACs included teachers from each school level (e.g., elementary, middle, high school), central office administrators overseeing the teacher evaluation process, administrators conducting evaluations, a data coordinator, and local school board representation. Membership was extended to other groups at the superintendent's discretion. One member of the advisory committee was identified as the pilot program liaison with the Department.

Measures of Student Achievement

Non-Tested Grades and Subjects

"Internally, we have created benchmarks for every grade level and every content area. The benchmarks are given four times a year, as a tool to measure student knowledge and understanding and prepare them for standardized assessments. Teachers, in conjunction with content area supervisors and outside content area experts, construct the benchmark exams. The benchmarks have been updated and revised every year, so as to align to the expectations in the curriculum."

- Project Director

In addition to performing more observations and training in a new evaluation instrument, some districts also developed benchmark tests to allow them to monitor growth of students in non-tested grades and subjects. Even though some districts had these in place at the beginning of the pilot (see box), the majority did not. One district had no common benchmarks but developed assessments in every untested area in 2011-12. The same district noted that "focused more on the process than the quality of the final product," acknowledging that the quality would take some time.

An outgrowth of this work was the generation of positive "conversations around standards, student learning objectives, and SMART goals." Another district's high school used its Professional Learning Communities and guidance from other states' models to create quarterly benchmark assessments. Three districts had struggled to move forward with this aspect of evaluation and noted that lack of time and guidance had made this work challenging. One project director commented that "teachers are willing to do this [assessment creation] but they must be trained." Another suggested that the state extend the work day to allow extra time without students to accomplish this.

Teachers' comments reinforced those from the project directors. Some reported that there was movement in this area in their districts, while in another the process had "stalled because other priorities took over," or had yet to begin. Teachers generally agreed that there were too many unanswered questions regarding non-tested grades and subjects.



Key Lessons Learned

The Department must provide clear guidance in the form of a "how to" publication and support for rigorous professional development to help districts produce measures of student achievement in non-tested grades and subjects. Districts should work in earnest to develop assessments that can be used by teachers in their non-tested grades and subjects that are as equitable and fair as possible when compared to tested areas.

Student Growth Percentile Data

Collecting, analyzing, and providing a useful summary of student growth percentile (SGP) data¹⁰ in each pilot district was impractical for the Department within the timeframe of this interim report. Given this reality, combined with the difficulties faced by districts in the area of non-tested grades and subjects, it is impossible at this point to draw a correlation between teacher observation data and student outcomes. SGP data was made available to pilot districts in January 2013 for the 2011-12 school year; these districts are currently analyzing the data and will provide feedback to the Department in the coming months. This critical analysis will be presented in detail in the final EPAC report that will be released later in 2013.

¹⁰ Student growth percentiles will be used in New Jersey as measure of student achievement for teachers of ELA and math, grades 4-8.

Cohort One Takeaways

Summarized below are key lessons and observations from pilot districts that will be of value to any district working towards full implementation of a new teacher evaluation model.

Changes in School/District Culture

Project directors were asked about the impact the evaluation initiative had on their school or district culture. Quotes from each district are summarized below.

How the Evaluation Initiative Affected School Culture
Increased professional discussion. "Marginal teachers" have become more aware of what is needed to become effective.
Increased anger, bitterness, and resentment. Teachers are angry with the government but still trust the district.
Boosted school and community pride. Increased reflection on practice and focus on student achievement.
Teachers demonstrated commitment, cooperation, and support.
Increase in district pride. Increased fear of what will happen to teachers (especially with tenure). Gratitude from teachers for having clear expectations.
No change yet.
More collaborative and comprehensive.
Forced the synthesis of several initiatives into one focused project.
Improved positive feeling surrounding teacher instrument and use of student data.

Districts should expect positive and negative changes in culture as they implement new evaluation systems. While some districts noted a spike in negativity and fear, others noted real benefits to school culture in the first year of pilot implementation. Pride, commitment, collaboration, and gratitude are mentioned by more than one district.

Best Practices

District project directors were all asked to provide at least one example of a best practice based on work in their first pilot year. Summarized responses from each district are shown below:

Best Practices
Train observers and teacher facilitators together.
Involve the teachers; this must not be a top-down process. Define the standards and elements in the instrument clearly and consistently across the district.
Focus on one domain and one component for walk-through's; this sets a positive tone towards the process. Turn the administrative team into a professional learning community.
Provide comprehensive and ongoing training for administrators.
Develop partnerships with a wide-range of stakeholders. Commit to a language of learning for all involved.
Train the entire district well.
Re-calibrate administrators' work days to prioritize the observation process.
Train teachers in the creation, scoring, and interpretation of valid student assessments.
Be up front about the rules of engagement. Have teacher leaders on the districts' evaluation committee. Establish rater accuracy, agreement, and reliability.



Key Lessons Learned

Themes taken from pilot district takeaways are:

- 1) Make the evaluation initiative transparent.
- 2) Include teachers as partners.
- 3) Provide excellent training.

Part Three:

The Work of the Evaluation Pilot Advisory Committee

Formation and Charge

In the summer of 2011, the Department solicited nominations and selected Evaluation Pilot Advisory Committee (EPAC) members representing a diverse cross-section of the New Jersey education landscape. These members included teachers of various subjects and grade levels, principals, superintendents, other administrators, parents, school board members, and representatives from private, charter, and vocational schools and the higher education community ([Appendix C](#)). Dr. Brian Osborne, Superintendent of South-Orange-Maplewood Schools, served as chair of the group.

In addition to these appointees, each district participating in the teacher evaluation pilot ([Appendix A](#)) was asked to send two representatives to attend EPAC meetings. As part of pilot requirements, participants convened district-level advisory committees with various local stakeholders, known as District Evaluation Pilot Advisory Committees (DEPACs). The two additional district representatives to the EPAC were members of their DEPACs, and at least one was required to be a teacher in order to maximize educator feedback at the state level.

Representatives from the 19 New Jersey schools receiving federal School Improvement Grant (SIG) funds ([Appendix A](#)) also participated in the 2011-12 pilot and EPAC meetings. In total, roughly 80 members, 25 of whom were teachers, served on the EPAC.

The EPAC's primary charge was to provide recommendations on various aspects of a statewide evaluation system based on learning from national research, best practices, the experiences of large school districts and other states, and the state's evaluation pilot program. Specifically, it was asked to:

1. Identify challenges and make recommendations for pilot implementation and for statewide roll out of an evaluation system, and
2. Provide written recommendations in July 2012 for a statewide roll out of the evaluation system in 2012-13.

In the first EPAC meeting in September 2011, the Department made it clear that any recommendations coming from the committee must fall within the broad outline described by the Educator Effectiveness Task Force (EETF) report, and the Notice of Grant Opportunity parameters under which all pilot districts were working; specifically that 50 percent of a teacher's evaluation be based on measures of student achievement and 50 percent be based on measures of teacher practice. While some EPAC members expressed concern about these restrictions and the weighting placed on student achievement, they agreed to work within these parameters.

EPAC Meetings

The EPAC met monthly from September 2011 through June 2012. Each meeting generally consisted of presentations by national and state experts, an update from the Department, a report from pilot districts, and subcommittee work. Background reading was often assigned in advance of each meeting ([Appendix D](#)).

Monthly meetings were structured to meet the following goals:

1. Learn from national perspectives on best practices to inform statewide implementation
2. Receive Department updates on implementation plans;
3. Learn about pilot trends, successes, and challenges that may inform recommendations for statewide implementation; and
4. Work in subcommittees to develop recommendations for statewide implementation.

10 major presentations were given over the course of nine EPAC meetings, including the October 2011 Evaluation Summit ([Appendix E](#)). Topics presented included the Widget Effect Project by Dan Weisberg, the Measures of Effective Teaching (MET) study and the Framework For Teaching by Charlotte Danielson, and the use of mini-observations by Kim Marshall. 81 percent of the thirty three EPAC members who completed a survey at the end of the year, found these presentations to be useful in developing their recommendations to the Department.

Seven monthly reports from pilot participants were delivered by the Department's Implementation Manager. These presentations were well received by the audience with 81 percent of survey respondents finding value in them. Several participants noted that these reports, including a panel discussion of pilot representatives, were the most valuable part of EPAC meetings.

Brian Osborne, chair of the EPAC, also provided monthly updates based on the current thinking, concerns, and suggestions that he had gathered in conversation with EPAC members and from information provided at meetings by the Department.

EPAC Successes

In addition to providing recommendations for the Department, engaging educators on a regular basis to discuss matters of student achievement and the profession of education yielded significant benefits. Some of these successes are highlighted below.

Extension of the Pilot Program and Implementation Timeline

As part of their discussions, EPAC members provided ongoing input on pilot implementation and preparation for statewide rollout. Several concerns surfaced repeatedly in the course of the year. These included local districts' capacity for meeting heightened expectations for educator evaluation, broad stakeholder input into local evaluation decisions, and the timing of full implementation. In early spring 2012, the Department decided that the timeline for statewide implementation of the new evaluation system should be extended. In addition, the Department decided that a second year of pilot implementation would be valuable to ensure the success of the new evaluation model. Both of these decisions were codified in the TEACHNJ Law enacted in the summer of 2012.

The decision to extend (and expand) the pilot program in 2012-13 and push back full implementation to 2013-14 was met with universal acclaim by EPAC members.

According to Peter Shulman, Assistant Commissioner, Division of Teacher and Leader Effectiveness, this course correction was made in large part due to the feedback received via the EPAC from pilot districts and EPAC members. The decision to extend (and expand) the pilot program into 2012-13 and push back full implementation to 2013-14 was met with universal acclaim by EPAC members. One commented that she was “impressed when the EPAC pressed for the extended year and the Department took the recommendation.”

Educator Partnerships

It was notable that EPAC meetings blurred the distinctions between educators. This was a place where teachers discussed policy with administrators and Department officials as equals. There was little sense of the labor and management divide that often defines schools. One EPAC member noted, “I couldn’t tell what role any person had because it all blended, which is remarkable.” Teachers, principals, and superintendents were encouraged to give presentations as well as nationally known experts and for many, this signaled the idea that the Department truly wanted to engage New Jersey’s educators.

Subcommittee Work

The most meaningful aspect of this engagement for a large portion of participants was their work in subcommittees. 75 percent of surveyed participants felt that they were provided ample opportunities to give input into subcommittee work (see Figure 7). Moreover, when asked to identify the aspects of EPAC meetings that were most productive, 50 percent of participants chose subcommittee work. One educator found this work especially useful when, “we really got into the nitty gritty.” While several participants noted that not enough time was spent in subcommittees, others appreciated the Department’s use of a dedicated wiki and email dialogue to prepare for or follow up on conversations happening within these groups.

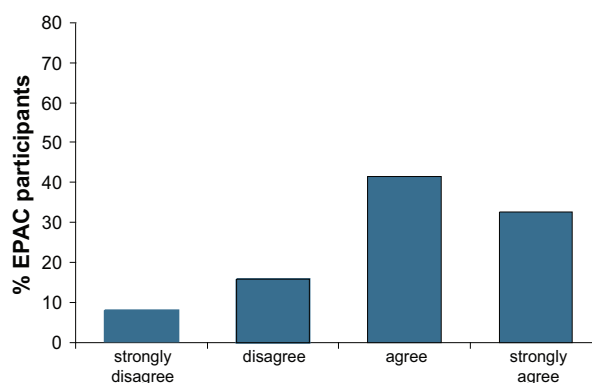


Fig. 7: Survey responses to question “EPAC members were given ample opportunity to provide input into subcommittee work.”

Professional Growth and Leadership Identification

Almost unanimously, individuals attending EPAC meetings considered the process a valuable experience professionally. Many enjoyed developing a view of the bigger picture in education, commenting that “the reading and experts have broadened my knowledge base,” and that it had helped develop “my awareness of other aspects of important educational issues.” The large amount of information and deep thinking occurring at EPAC meetings led one person to claim, “I feel like I have learned a whole new language.” Armed with this new information, one person noted, “I am better positioned to support districts,” and another that “I am able to talk about teacher evaluation in an educated manner making me a more credible advocate for change.”

Having educators engaged in the evaluation initiative has allowed schools, districts, and the state to identify individuals whose professionalism and commitment to the work have been useful in effectively moving the work forward. Teachers in particular have gained from this experience. Their work in DEPACs and as crucial turnkey trainers for district staff “has helped them to evolve into more reflective practitioners,” said one project director. Finally, one teacher noted a great deal of professional growth through participating in EPAC because, “I am rarely given the opportunity to contribute at the district and state level.”

EPAC Challenges

Despite the Department's willingness to engage educators in the process of evaluation design, it faced several challenges in making the most of this collaborative process. These were magnified by the short timeline and the complex nature of the work. Some of these challenges are described below.

Committee Size and Structure

When the EPAC was formed, it consisted of appointees who represented various stakeholder groups. The charge of this group was to listen, learn, and make recommendations for statewide rollout of improved educator evaluations. Subsequently, the group was expanded to include members of pilot districts, thereby providing a different perspective on this work and vastly increasing the size of the EPAC.

The large size of meetings, over 80 participants many months, added to management difficulties. One highly experienced educational leader said, "Large groups are challenging and hard to facilitate." However, recognizing the difficulty in reducing the meeting size, another committed EPAC attendee suggested that, "if we made the group small, then it would be a 'privileged group' and what we need is a balance between the two."

While the inclusion of members of districts' DEPACs was crucial for the learning of the group, there was some pressure to use the EPAC as a forum at which the Department provided guidance to take back to the field. Even though regular Department updates were provided, there were certain questions that some participants felt were not being answered with clarity or certainty, and this led to frustration.

In addition, the original core of appointees did not convene again apart from the larger group and some dissatisfied appointees self-selected off the committee.

Priorities and Presentations

Several participants noted other frustrations such as feeling that not enough time was spent on subcommittee work and that a reduction in presentation time by national experts should have been arranged to accomplish this. Some felt that information presented on different evaluation models was misplaced.

More than a few participants noted that these presentations were the least productive aspect of the meetings. Several participants claimed that they felt that vendors were there to sell products to them rather than "sharing global research." Even those who acknowledged the value of presentations, said they would have liked more time to process the information in either large or small groups.

Continuity and Voice

Even though the Department put a great deal of time and effort into designing EPAC meetings, some EPAC members noted that the lack of connections between the information presented created a challenge. "We heard presentations, but there was nothing to compare them with during the meeting," said one EPAC member. Another agreed, commenting, "Once we discuss something, we are not part of the next conversations that take place from the feedback." One more member said, "The information we get is random."

The challenges of committee structure and information continuity perhaps contributed to the feeling of some EPAC members that they were not heard. Although 86 percent of EPAC participants who completed the end of year survey said they felt that their voices were “respected and heard,” one EPAC member said, “I have not felt I’ve had that much voice throughout,” and that, “people want to be heard and have more time for discussion.”

EPAC Subcommittees

EPAC subcommittees were formed in November 2011 to address questions and provide recommendations. The topics of these subcommittees were chosen by the Department. Before the subcommittees convened, EPAC members were asked to state their top three group choices and were then assigned to one of these. The subcommittee topics were as follows:

- Early Childhood Teachers
- Teachers of English Language Learners
- School Leaders
- Professional Development and School Culture
- Special Education Teachers
- Summative Ratings
- Teacher Practice

The composition of each subcommittee was based on the areas of expertise and interests of participants. Each subcommittee was chaired by an EPAC member ([Appendix F](#)) and facilitated by a representative of the Department. The subcommittees generally met in the afternoon of each EPAC meeting and regularly reported out to the larger group. Sessions followed a similar format; the Department representative provided a background document outlining current thinking, along with one or two decision points. The subcommittees then discussed and made recommendations based on this document.

Subsequent meetings built on the work done in previous months. Background reading was often assigned by the Department to ensure that participants came ready to discuss the issues with the most up-to-date research available. Each subcommittee presented its recommendations to the Department in a report. The subcommittee recommendations are shown in the following summary tables and can be found in more detail in [Appendix G](#).

Subcommittee Recommendations¹¹

EARLY CHILDHOOD TEACHERS
Establish “Early Childhood” as a sub-group of non-tested grades and subjects and provide guidance for pre-K and K, and 1st through 3rd grades.
Provide guidance on collecting data in student portfolios that is based on appropriate early education performance criteria.
Provide guidance on adapting observation systems that take into consideration a developmentally appropriate curriculum.
Set a ceiling of 10 percent for the student growth percentile (SGP) as the Department continues to conduct research on the use of the SGP at the early childhood education level.

¹¹ Unless otherwise specified, each recommendation is suggested as action that should be taken by the Department.

TEACHERS OF ENGLISH LANGUAGE LEARNERS (ELL)

Provide observation protocols that effectively measure the specific methodologies used by ELL teachers.

Analyze and define the rate of growth on state assessments that can reasonably be expected of ELLs at each language proficiency level.

Since all teachers in a school are more or less responsible for language growth in all of its students, use language proficiency growth data in the evaluations of all teachers.

SCHOOL LEADERS

Ensure the Human Capital Management Responsibilities category in the principal evaluation pilot measures the effectiveness in the quality of and opportunities provided to improve teacher effectiveness and practice.

Provide a guidance document that explains the principal evaluation system. This may include items such as the selection of a system, how it may be aligned with a teacher evaluation system, and how surveys of various stakeholders may be used in the evaluation.

Expand the EPAC to include more experienced administrators and create subcommittees to address key areas of principal evaluation.

Support the creation of consortia to help with implementation and statewide professional development for administrators that will enhance instructional leadership capacity.

Design an evaluation system that takes into account principal experience and school context.

PROFESSIONAL DEVELOPMENT AND SCHOOL CULTURE

Require the establishment of structures for educator teams to have sustained collaboration focused on teaching and learning.

Require teachers and principals to collaborate in creating goals for teacher Professional Development Plans (PDPs) using multiple types of evidence.

Require teachers to engage in structured professional reflection as an ongoing process.

Hold district leaders and principals accountable for ensuring that teachers receive needed support.

Encourage districts to provide more latitude in professional development for “highly effective” teachers and create an infrastructure for those who elect to serve in leadership roles.

Require districts to use data to analyze trends in teaching practice and student achievement when developing PDPs.

SPECIAL EDUCATION TEACHERS

Allow districts to utilize multiple assessments for teachers who are in non-tested grades and subjects that may include statewide tests, progress towards meeting IEP goals and SLOs.

Develop a system to apply a weighted value to each of the multiple measures.

Provide guidance and professional development as districts select assessments and develop Individualized Education Plan (IEP) goals in order to meet the state's student learning objectives.

Create a specialized group to identify standards of effectiveness that characterize special education teachers. Districts should consult with evaluation instrument authors to ensure there is validity and reliability in the instrument based on these standards.

SUMMATIVE RATINGS

Establish a two-way matrix that demonstrates performance levels on the teacher practice side and the student achievement side that will be combined into one overall summative rating.

Provide definitions and guidance on how the observational data from the teacher practice side will correlate with the four rating levels either through a) the DOE approval process for instrument vendors or, by b) creating a common state summative rubric to crosswalk with vendors' rubrics.

Develop guidance on exactly how districts should reach a summative rating on the student achievement side of evaluation in areas of non-tested grades and subjects, tested grades and subjects, and school-wide measures.

Districts should develop assessments for non-tested grades and subject areas that are aligned with the standards, and include rubrics and/or SMART goals that link to teacher evaluation.

TEACHING PRACTICE

Using common vocabulary, review and refine criteria for the selection of teaching practice instruments.

Provide a list of teaching practice instruments that meet the criteria.

Require that comprehensive observer training be completed and required skills mastered, as demonstrated through an established process, before observations can be conducted. Guidance for training requirements should include specific criteria for skills and competencies, as well as standards for calibration to ensure inter-rater reliability.

Provide explicit guidance for training that ensures that before being observed, teachers are trained in all aspects of the process, including the additional practice measures.

Develop guidance on how novice teachers who are rated partially effective or ineffective will receive additional support.

Allow districts more flexibility in assigning the additional weights for teacher practice measures such as portfolios, self-reflection, and student surveys after doing additional research on how to evaluate them.

Promote observation practices that include formative processes such as walk-through's and peer observations that provide feedback for growth and support.

Outcomes of Subcommittee Recommendations

Subcommittee work provided rich feedback to the Department, and several of the recommendations have already been adopted. For example, the first two recommendations of the teaching practice subcommittee have been incorporated into proposed regulations. These include allowing districts to choose a teaching practice evaluation instrument that meets a set of specifications rather than having New Jersey adopt a statewide model. In addition, the Department has created a process for approving instruments, including those developed by districts, using a Request for Qualifications (RFQ) process and has posted [lists](#) of those already approved. Other recommendations are still under consideration and continue to inform the Department as the state moves towards full implementation in September 2013.

Subcommittee work provided rich feedback to the Department, and several of the recommendations have already been adopted.

Unanswered Questions and Next Steps

Participants sometimes felt that important questions were not being addressed adequately by the EPAC. Among these was the right approach for measuring achievement in non-tested grades and subjects. Even though a subcommittee was formed to tackle this difficult topic, it produced no recommendations and only met once. The lack of guidance from the state on this topic left many pilot participants feeling uneasy. Additionally, concerns over the weightings of student achievement set forth in the EETF report and the pilot requirements were frequently expressed but the Department indicated, as it had at the very first EPAC meeting, that this was a non-negotiable aspect of the evaluation process.

At the last 2011-12 meeting of the EPAC in June, participants signaled that they wanted more guidance and discussion on implementation questions in pilot districts. Specifically, EPAC members suggested that further conversation on the following topics would be critical:

- Lack of timely SGP data for making personnel decisions
- Use of student growth percentiles (SGPs) to measure teacher effectiveness
- Evaluation instrument implementation
- Data linkage and tags
- Alignment of professional development with observations
- Collecting data to improve process
- Increasing number/effectiveness of observations
- Effect of principal evaluation on teacher evaluation
- Calculation of summative ratings
- Evaluation rubric rollout plan

The Evaluation Pilot Advisory Committee has been expanded in 2012-13 to include representatives from a second cohort of teacher evaluation pilot districts and a cohort of principal evaluation pilot districts. Moving ahead into the next school year, and especially with the arrival of the [TEACHNJ Act](#) (new tenure law) in August 2012, it will be important for the Department to provide opportunities to discuss these unanswered questions and learn from the recommendations that the EPAC is able to make based on its collective knowledge.¹²

¹² At the time of publication of this report, many of the unanswered questions were being addressed by way of changes in the second round of NGOs and through discussions at EPAC.

Conclusion

Even when tensions were high, and frustrations many, dedicated professionals continued to make deliberate progress.

The State of New Jersey has embarked on an ambitious project to overhaul the evaluation of educators. The Department aims to develop an evaluation system that will more effectively recognize the true performance of classroom teachers and school leaders. Armed with this information, school districts will be able to make improved personnel decisions and provide appropriate professional support for educators across the spectrum of professional practice, elevating the quality of teaching and school leadership, so that all students may be better served by public schools.

While the theory behind the evaluation initiative is sound, it is clear from the deep work of the EPAC and the Cohort One pilot districts that the task of bringing these goals to fruition is difficult and complex. Even with an extended implementation timeline, an extra year of piloting, and additional districts engaged in trial runs of a new system, this work is just beginning and promises to present challenges for years to come.

Addressing capacity issues, developing measures of student achievement in areas of non-tested grades and subjects, and calculating summative ratings are just some of the challenges pilot districts continue to face. In addition, the delay in releasing SGP data from the state presents a challenge for making timely personnel decisions. Solutions need to be found for this.

However, just because the work is difficult does not mean it should not be done. Initial reports from pilot districts provide reason to hope that we are moving in the right direction. Despite facing a variety of unique challenges, teachers and school leaders are seeing the benefits of adopting a high quality evaluation instrument. They are witnessing a transformation in the type and quality of conversations surrounding teacher practice and student learning. Some districts are more effectively differentiating between the performance of teachers, even in these early stages of implementation, and are developing new systems of recognition for teachers with excellent practice and professional support for everyone. These districts are the vanguard for others in the state who are just beginning this work. Lessons learned from the pilots have been invaluable and continue to inform the Department's work, allowing it to make wiser and more practical recommendations, such as allowing districts flexibility in the type of evaluation instruments they use.

Additionally, the first year of the EPAC's activities have demonstrated that dozens of strong educators, educational leaders, and officials from the state can come together, learn from one another, and strive to make a very difficult task possible. Even when tensions were high and frustrations many, dedicated professionals continued to make deliberate progress. The Department is approaching evaluation activities carefully and thoughtfully with the continued guidance of the EPAC in 2012-13.

It is this type of perseverance and continued collaboration between the Department and educators throughout New Jersey that will be crucial in creating an environment conducive to the growth and success of such an ambitious program of reform. It is in this spirit of collaboration that difficult work must be done if we are to make educator evaluation reform more than just a passing fad, but a lasting legacy that will benefit all of New Jersey's children for years to come.

Appendices

Appendix A: Pilot Participants, 2011-2012

School Districts

School	County
Alexandria Township	Hunterdon
Bergenfield	Bergen
Elizabeth	Union
Monroe Township	Middlesex
Newark	Essex
Ocean City	Cape May
Pemberton Township	Burlington
Red Bank Borough	Monmouth
Secaucus	Hudson
West Deptford Township	Gloucester
Woodstown-Pilesgrove Regional	Salem

SIG Schools

School	District
Cramer College Preparatory Lab School	Camden City Board of Education
U.S. Wiggins College Preparatory Lab School	Camden City Board of Education
Camden High School	Camden City Board of Education
Essex Vocational West Caldwell	Essex Vocational Technical Schools
Cicely Tyson High School	East Orange
Lincoln High School	Jersey City Board of Education
Fred Martin School of Performing Arts	Jersey City Board of Education
Snyder High School	Jersey City Board of Education
Lakewood High School	Lakewood Board of Education
Newark Central High School	Newark Public Schools
Dayton Street School	Newark Public Schools
Newark Vocational High School	Newark Public Schools
Malcolm X. Shabazz High School	Newark Public Schools
Brick Avon Academy	Newark Public Schools
Barringer High School	Newark Public Schools
West Side High School	Newark Public Schools
Dr. Frank Napier School of Technology	Paterson School District
School Number 10	Paterson School District
Abraham Clark High School	Roselle Public Schools

Appendix B: Evaluation Instruments and Data Management Systems by District

Cohort One Pilot District	Instrument	Project Management System
Alexandria	James Stronge	Oasys
Bergenfield	Danielson	Teachscape
Elizabeth	Danielson	iObservation
Monroe Township	Marzano	iObservation
Ocean City	Danielson	iObservation
Pemberton	Danielson	Teachscape
Red Bank	Danielson	Teachscape
Secaucus	Danielson	Teachscape
West Deptford	McREL	McREL
Woodstown-Pilesgrove	McREL	McREL

Appendix C: EPAC Members, 2011-2012

- **Ms. Marie Bilik**, Executive Director, New Jersey School Boards Association
- **Mr. Carl Blanchard**, National Board Certified Teacher; Biology Teacher, Franklin High School
- **Ms. Marie Blistan**, Secretary/Treasurer, New Jersey Education Association
- **Ms. Jeanne DelColle**, Burlington County Teacher of the Year; History Teacher, Burlington County Institute of Technology
- **Ms. Patricia Donaghue**, Parent, Toms River, NJ
- **Ms. Carole Everett**, Executive Director, New Jersey Association of Independent Schools
- **Dr. Dorothy Feola**, Past President, New Jersey Association of Colleges for Teacher Education; Associate Dean, College of Education, William Paterson University
- **Ms. Darleen Gearhart**, Director, School Improvement Grants, Newark Public Schools
- **Mr. Timothy Matheney**, Principal, South Brunswick High School
- **Ms. Eileen Matus**, Retired Principal, Toms River Regional School District
- **Ms. Elizabeth Morgan**, National Board Certified Teacher; English Language Arts Teacher, Ann A. Mullen Middle School
- **Dr. Brian Osborne**, Superintendent, South Orange-Maplewood Schools
- **Mr. Richard Panicucci**, Assistant Superintendent of Curriculum – Vo-Tech, Bergen County Technical Schools/Special Services
- **Ms. Meredith Pennotti**, Principal, Red Bank Charter School
- **Ms. Judith Rattner**, Superintendent, Berkeley Heights Public Schools
- **Dr. Vivian Rodriguez**, Assistant Superintendent, Perth Amboy School District
- **Dr. Sharon Sherman**, Dean, School of Education, Rider University
- **Ms. Peggy Stewart**, Chair, Professional Teaching Standards Board; History Teacher, Center for Teaching and Learning
- **Ms. Belinda Stokes**, Principal, Henry Snyder High School
- **Dr. Dorothy Strickland**, New Jersey State Board of Education; Samuel DeWitt Proctor Professor of Education, State of New Jersey Professor of Reading, Emerita, Rutgers University
- **Mr. Bruce Taterka**, U.S. Teaching Ambassador Fellow; Lead Teacher of Science and Technology, West Morris Mendham High School
- **Ms. Patricia Wright**, Executive Director, NJ Principals and Supervisors Association

Appendix D: EPAC Meeting Background Reading, 2011-2012

1. Task Force Report: <http://www.state.nj.us/education/educators/effectiveness.pdf>
2. A Practical Guide to Designing Comprehensive Teacher Evaluation Systems: <http://www.tqsource.org/publications/practicalGuideEvalSystems.pdf>
3. The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness: <http://widgeteffect.org/downloads/TheWidgetEffect.pdf>
4. Measures of Effective Teaching (MET) Project Preliminary Report: Learning About Teaching Brief: http://www.metproject.org/downloads/Preliminary_Finding-Policy_Brief.pdf
5. Rethinking Teacher Evaluation in Chicago: <http://ccsr.uchicago.edu/publications/rethinking-teacher-evaluation-chicago-lessons-learned-classroom-observations-principal>
6. MET Project Second Preliminary Report: Gathering Feedback for Teaching: Policy Brief: http://www.metproject.org/downloads/MET_Gathering_Feedback_Practitioner_Brief.pdf
7. MET Project Policy Brief: Gathering Feedback for Teaching: <http://www.metproject.org/>
8. Teacher Evaluator Training and Certification: <http://www.teachscape.com/resources/teacher-effectiveness-research/2012/02/teacher-evaluator-training-and-certification.html>
9. Merit Pay or Team Accountability by Kim Marshall, , Education Week Commentary, August 30, 2010: <http://www.edweek.org/ew/articles/2010/09/01/02marshall.h30.html?tkn=NNPFt0BdGgxqWE+0mVYQ2wckC0Adi2uVBwxU&cmp=clp-edweek>
10. It's Time to Rethink Teacher Supervision and Evaluation by Kim Marshall, Phi Delta Kappan, June 2005: <http://www.marshallmemo.com/articles/Kappan%20Superv.%20article.pdf>
11. Implementation of the National SAM Innovation Project: A Comparison of Project Designs, Wallace Foundation, August 2011: http://timetrack.jefferson.kyschools.us/psa_sam_models_2011.pdf
12. Creating a Comprehensive System for Evaluating and Supporting Effective Teaching, Darling-Hammond, L., Cook, C., Jaquith, A., and Hamilton, M. Stanford Center for Opportunity Policy in Education, 2012: <http://edpolicy.stanford.edu/sites/default/files/publications/creating-comprehensive-system-evaluating-and-supporting-effective-teaching.pdf>
13. Ensuring Accurate Feedback from Observations: Perspectives on Practice, Craig Jerald, Bill and Melinda Gates Foundation, March 2012: <http://www.gatesfoundation.org/college-ready-education/Documents/ensuring-accuracy-wp.pdf>

Appendix E: EPAC Presentations

TOPICS	PRESENTER	PRESENTATION OVERVIEW
Videotaping lessons in educator observations	Mark Atkinson Founder, Teachscape	Teachscape provided technology to collect video of and score lessons in the MET study. Teachscape is developing next generation software to enhance this tool.
Evaluating educator effectiveness using multiple measures; Framework for Teaching	Charlotte Danielson The Danielson Group	The Measure of Effective Teaching (MET) study is a large scale investigation of teacher evaluation methods. Danielson's Framework for Teaching is an evaluation instrument used in the MET study.
Implementing a new evaluation system in a large urban district	Jason Kamras Chief in the Office of Human Capital, DC Public Schools	DC IMPACT is one of the first large district evaluation systems that uses multiple measures, including student performance, to make personnel decisions including dismissal and awarding a merit pay.
Providing feedback and support for teachers to improve the quality of their practice	Jason Lange Founder and CEO, Formative Teaching	Bloomboard's (formerly Formative Teaching) online platform provides individualized professional development. It offers video-based teaching strategies, planning tools, and support and coaching.
Adjusting observation frequency and length to be effective and efficient	Kim Marshall New Leaders for New Schools	Frequent, unannounced mini-observations provide an efficient and effective way to observe and coach teachers .
Ensuring fair and accurate teacher observations	Catherine McClellan Principal Scientist, Clowder Consulting	The MET study provides insight into how observer calibration and checks for inter-rater reliability can reduce bias and increase the fairness and accuracy of teacher observations.
Using student surveys to measure teacher practice	Rob Ramsdell Vice President, Cambridge Education	The MET study finds that student perception correlates well with other measures of teacher practice. The Tripod survey tool was used in this study for purpose of measuring student perception.
Time management for principals moving towards instructional leadership	Mark Shellinger Director, National School Administrator Manager Project (SAM)	The SAM Project has developed a process to help principals move their focus from school management to instructional leadership.
Measuring effective teacher-student interactions to support student achievement	Ginny Vitiello Director, Teachstone	The Classroom Assessment Scoring System (CLASS) is a teacher observation instrument that measures effective teacher-student interactions. In the MET study, it is shown to be predictive of social and academic gains in students.
Shortcomings of most traditional educator evaluation systems	Daniel Weisberg Executive Vice President, The New Teacher Project (TNTP)	The Widget Effect report analyzed teacher evaluations and student achievement in four states. It concluded that a teacher's observation ratings were good or great even if their student's performance was not.

Appendix F: EPAC Subcommittee Chairs, 2011-2012

- **Early Childhood Teachers**, Laura Morana
- **Teachers of English Language Learners**, Raquel Sinai
- **School Leaders**, Pat Wright
- **Professional Development and School Culture**, Peggy Stewart
- **Special Education Teachers**, Peggy McDonald
- **Summative Ratings**, Robert Fisicaro
- **Teacher Practice**, Eileen Matus

Appendix G: Subcommittee Reports, 2011-2012

Early Childhood Teachers

Recommendation 1: The Department should establish “Early Childhood” as a sub-group of NTGS and provide guidance specific to Pre-K and K & 1st through 3rd on collecting data in student portfolios for the following domains of child development: physical development, language and literacy, mathematical/scientific thinking, approaches toward learning, and personal and social development . Furthermore, the guidance should outline performance assessment criteria that include a series of options with rubrics for choosing instruments as well as a comprehensive list of acceptable instruments. The criterion would include information on collecting data that allows for comparisons of children in three ways: in relation to him/herself, in relation to the class, and in relation to a standard.

Recommendation 2: The Department should provide guidance for school districts on adapting observation systems for early childhood to take into account key components of Department-approved developmentally appropriate curriculum models, especially at the Pre-K and Kindergarten levels. Examples of these components include center time, morning meetings, and other parts of the daily routine that are not typical in later grades. Furthermore, the guidance should include training protocols that ensure reliability-training specific to early childhood. The Department should also provide a list of instruments that may inform training protocols.

Recommendation 3: The Department should set a ceiling of 10 percent for SGP as it continues to conduct and reviews research on the use of SGP scores within PreK-3rd grade.

Teachers of English Language Learners (ELLs)

Recommendation 1: The ELL subcommittee considers current teacher effectiveness observation protocols as not effectively measuring the additional competencies that educators of ELLs must possess and responsibilities they must address. The Department should include the strategic methodologies that teachers of ELLs use to help students develop English language skills and meaningfully engage with content according to their English language proficiency level.

Recommendation 2: For ELLs, achieving proficiency on state content assessments is dependent on both content knowledge and proficiency in English. Thus, ELLs’ achievement on state tests is not reflective of their mastery of the content, but in part, reflective of their mastery of the English language. As their English skills increase, ELLs’ performance on content assessments is more reflective of content knowledge than of language proficiency. It is recommended that the Department:

1. Use growth on the ACCESS for ELLs English language proficiency assessment and apply to all teachers in a school that enrolls ELLs to determine teacher effectiveness. Growth should be measured based on students’ English language proficiency improvement, and applied to all teachers in the school because such growth should be addressed across the curriculum.

2. Apply growth targets on state assessments that are based on the language proficiency levels of students. An analysis of ACCESS for ELLs and state assessment data would have to be conducted to determine what growth/achievement on state assessments can be expected of ELLs at each language proficiency level.

School Leaders

Recommendation 1: The 10 percent component identified as *Human Capital Management Responsibilities* in the Principal Pilot Notice of Grant Opportunity (NGO) should measure the effectiveness in the **quality of** and **opportunities provided** to improve teacher effectiveness and practice. For this component, those evaluating principals will be expected to seek evidence of the principal's effectiveness in:

1. Fulfilling the requirements of district policies for the supervision and evaluation of teachers
2. Observing and rating teachers consistently and accurately; and
3. Conducting pre- and post- observation conferences and providing teachers with feedback that will support them in improving their practice.

Other sources of evidence relating to this component of professional practice that could be included are documentation of the principal's effectiveness in:

1. Recruiting and/or retaining effective teaching staff;
2. Developing and monitoring teachers' required individual professional development plans;
3. Managing the implementation of the required school level professional development plan;
4. Providing opportunities for providing time and resources for collaborative job- embedded professional learning and collaborative work time; and
5. Providing high quality professional learning opportunities to meet both individual and collaborative team goals resulting from reflection and analysis of both teacher evaluation data and student performance data.

Recommendation 2: The Department should provide guidance to districts to support their selection of a principal and teaching practice evaluation instrument and data management system.

Recommendation 3: The Department should clarify the criteria recommended in the ESEA waiver regarding incorporating feedback from teachers and other stakeholders to inform a principal's evaluation.

Recommendation 4: The Department should clarify that evaluators and principals should have the flexibility to determine which collaboratively developed and mutually agreed upon goals, either academic achievement or other measures of student achievement, should comprise the 15 percent component titled School-Specific Student Performance Goals.

Recommendation 5: The Department should encourage districts to create consortia to share costs and collaborate in the training and implementation of the principal practice instruments.

Recommendation 6: The Department should expand the present EPAC to include more administrators and stakeholders with school leadership expertise and experience.

Recommendation 7: The Department should support the development of current principals and principal preparation programs by providing statewide professional development for principals, building level leaders and central office leaders to enhance school leaders' instructional leadership capacity.

Recommendation 8: The Department should clearly state the year in which the principal evaluation system will be implemented statewide.

Recommendation 9: The Department should provide information as to how the principal and teacher evaluation systems and practice evaluation instruments are aligned.

Recommendation 10: The NJ Department should design the principal evaluation system to take into consideration the principal's level of experience and the school context.

Recommendation 11: The Department should create communication and guidance documents that explain the principal evaluation system.

Recommendation 12: The Department should create several principal EPAC sub-committees to provide recommendations on key areas around principal evaluation. The present sub-committee recommended the following topics for principal EPAC sub-committees to explore for the 2012 – 2013 year: stakeholder surveys and their use as formative feedback, professional development planning for principals, the impact of a principal's experience and school context on the evaluation process and an analysis of the school leadership career continuum in light of the current evaluation system.

Recommendation 13: The Department should clarify the Notice of Grant Opportunity language explaining the aggregate student performance for high school principals. In the pilot year the principal EPAC should further study the aggregate student measures used to evaluate high school principals.

Professional Development / School Culture

Recommendation 1: The Department should require all schools and school districts to establish structures for teams to have sustained collaboration focused on teaching and learning several times per week. This includes requiring districts to build the capacity of principals to serve as instructional leaders able to support and guide the implementation of learning communities and the efficacy of collaborative teams to the end of improving teacher practice and student achievement. The Department should also require that principals be evaluated on their effectiveness as instructional leaders.

Recommendation 2: The Department should require teachers and principals to collaborate in creating goals for the Teacher Professional Development Plan (PDP) and to use multiple types of evidence to inform each teacher's goal setting. The teacher should lead the PDP discussion except when a teacher's is rated "ineffective." For those teachers rated "ineffective" in one or more evaluations, the principal should be required to guide the PDP decisions. In addition, the Department should require more observations using the teaching practice instrument for those teachers rated "ineffective" than for other teachers. When setting goals, the principal and teacher should analyze the extent to which previous professional development opportunities (a) focused on student learning and (b) addressed the teacher's needs (pedagogy, classroom management, content knowledge, etc.). Generally, PDP individual goals should be consistent with school-wide goals and district-wide goals. However, in order to address a teacher's specific need, a goal(s) may be focused on an area not tightly aligned with district or school goals.

Recommendation 3: The Department should require teachers to engage in professional reflection as an ongoing process, consistent with standard 10 of the New Jersey Professional Standards for Teachers and School Leaders. Throughout the school year, teachers should be given opportunities to collect and reflect on formative evidence, beyond evaluative data, of their planning/preparation, instruction, student learning, and professional learning—both for self-reflection and for purposes of demonstrating professional growth. Such artifacts are over and above evidence collected through required evaluation procedures. Examples of formative evidence that should be collected include: classroom student assessment data; student work products; teacher-specific or team-specific artifacts related to curriculum, instruction and assessment; and records of collaborative team goals, actions and outcomes. Principals should have opportunities to examine such evidence, when it relates to an identified weakness (e.g., via conferences) prior to completing the annual performance report.

Recommendation 4: The Department should hold district leaders and school principals accountable for ensuring that teachers receive needed support. All teachers should be held responsible for meeting their professional goals by taking advantage of opportunities to grow. The Department should also require that districts provide a mechanism guaranteeing opportunities for ongoing coaching, including "refresher" training, in the district's teacher practice instrument and the principal practice instrument for those needing or wanting such support. Follow-up training and coaching in the teaching practice instrument must not replace research-based professional learning as set out in standards 8, 9 and 11 of the Professional Development Standards for New Jersey Educators and pursuant to N.J.A.C. 6A:9-15. In particular, districts should guard against using isolated, "remediation" training models that do not effectively build a teacher's content knowledge or pedagogical skills (e.g., viewing videos of lessons with little or no peer collaboration).

Recommendation 5: Although teachers who are rated "highly effective" must be responsible for addressing any identified weaknesses, the Department should encourage districts to permit them more latitude in professional development choices than other teachers. In addition, highly effective teachers should have opportunities to serve in leadership roles that enrich and expand professional learning for their colleagues. The Department should require districts to create an infrastructure to support teachers who elect to serve in one or more hybrid roles, i.e., instructional coaches and content coaches who are certified in the teaching instrument, non-evaluative observers, facilitators of school-based collaborative teams, turnkey trainers and a host of other roles, such as those set out in the Model Teacher Leader Standards.

Recommendation 6: The Department should require School Professional Development Committees (SPDCs), when creating the school professional development plans, to include in their needs assessments an analysis of data in the aggregate showing trends in teaching practice and student achievement. These data should include anonymous classroom observation data, teacher collected evidence of teaching practice, documentation from collaborative team(s), survey data, and other teacher evaluation measures, including evidence currently required in school-based plans. SPDCs should be required to use these sources to help identify where teaching practice is effective and where improvement is needed in the school. The Department should require the Local Professional Development Committees (LPDCs) to do the same type of analysis and, further, should require that the school-level needs assessments inform the districts' goals, priorities and professional development opportunities.

Special Education Teachers

Recommendation 1: The Department should allow districts to utilize multiple assessments for teachers who are teaching in non-tested grades, subjects and programs. These multiple measures may include; statewide assessments, progress towards accomplishment of IEP objectives and student learning objectives (SLOs) as well as external measures including SMART goals designed by the district or criterion-referenced or evidence-based assessments that are valid and reliable. To provide districts guidance, the Department should develop guidelines for the selection of formative assessments that will measure progress towards the SLOs as well as identification of links to websites that provide information regarding information on commercial assessments.

The Department should develop a system to apply a weighted value to each of the multiple measures which will be aggregated in support of student progress and define a range for the relative weights of each of the multiple measures. The committee recommends that this range shall include the recommendation that no single factor shall exceed 50 percent value of the student achievement section of the teacher evaluation .

The Department should provide professional development on the development of IEP goals and objectives that support the achievement standards and SLOs developed by the state with formative and summative assessment measures to evaluate student progress.

Recommendation 2: The Department should provide effective special education teacher and related services effectiveness of standards from nationally recognized organizations such as the Council for Exceptional Children (CEC, American Board for Certification of Teacher Excellence and others.) To accomplish this, the Department should identify a group of stakeholders including general/special education teachers, related services, special education directors, supervisors, principals, university faculty and parents to identify the standards of effective special education teachers and related services staff. The Department should develop a bank of performance indicators that align with each of the standards that districts can select from to meet their schools staff particular job responsibilities. They could reach out to pilot districts and disseminate special education performance standards and indicators to get feedback from districts and use feedback to modify or make changes if necessary. The committee recommends the Department provide guidance to districts on adapting their existing observation protocols based on INTASC standards to include the special knowledge and skills needed by special education teachers and related services that is defined by multiple levels of effectiveness similar to other indicators on the observation protocol used by the districts. Once standards and indicators are identified, districts should consult with authors of their approved protocols to identify a process for validating modified protocols to protect the validity and reliability of the product.

The Department should also develop guidance for school districts to identify those individuals at the school and district level to evaluate special education teachers and related services staff. They should also develop a training module on the special education standards and indicators, conduct professional development using a training of trainers' format to inform school district personnel on the standards and indicators for the evaluation of special educator's and related service personnel's unique skills and knowledge.

Summative Ratings

Recommendation 1: After examining multiple methodologies in regards to a summative rating, the group's recommendation is to establish a two way matrix prior to reaching the summative evaluation. A two way matrix will serve to demonstrate performance levels on the teacher practice side and the student achievement side prior to combining these performance levels to establish one overall summative rating as part of each teacher's annual performance report.

Recommendation 2: The Department should consider one of two options necessary for the purpose of providing guidance to school districts in regards to how the observational data for the teacher practice side of the evaluation will correlate to a Summative Rating to the four levels of Highly Effective, Effective, Partially Effective, and Ineffective. The Department should also provide school districts with a definition for these terms.

- **Option 1:** The Department should consider requiring approved evaluation frameworks, whether they be commercial or home grown, to include formalized processes for how a summative rating is formulated on the teacher practice side according to the framework's design. All frameworks should also demonstrate alignment to the INTASC standards and once adopted school districts should be required to follow these procedures accordingly. The Department should develop guidelines and requirements for teacher practice framework vendors to submit procedures that outline how the observational data that is captured as part of the framework will interface with the four level summative rating scale on the teacher practice side of the evaluation.

- **Option 2:** The Department should consider developing an agnostic summative rating teacher evaluation rubric that could serve as a crosswalk between approved teacher evaluation practice frameworks and the INTASC standards. By requiring all observational data to filter through a common state summative rubric, the state of New Jersey will be in an improved position in regards to developing guidelines that can detail how a summative rating on the teacher practice side is formulated while seeking comparability among various approved framework models. Approved framework providers should be required to develop procedures and guidelines for cross walking the observational data collected into the statewide rubric and the Department should develop guidelines for approving and supporting the crosswalk procedures.

Recommendation 3: : The Department should develop guidance for school districts to follow for the purposes of reaching a summative rating on the student achievement side of an evaluation. This guidance should be developed in three areas; Tested grades (4th-8th grades), Non Tested Grade Levels and Subject Areas, and other School Wide Measures.

- In the area of tested grades, more guidance is needed for districts that explains how students' aggregate performances on the NJASK assessments can interface with the SGP methodology for the purposes of linking evidence of student growth to teachers' evaluation. Cut Scores and Scales that correspond to the student growth percentile scores are also needed.
- For non tested grades and subject areas, school districts should be required to develop and utilize assessments that are standards aligned and include specific student learning objectives. Common assessments should be developed within school districts and grade levels and the Department should provide collaboration and training opportunities at the county or regional level. Each assessment aimed at linking evidence of student learning to teacher evaluation should include a rubric and/or SMART goals identifying how students' performances can be quantified. Parameters and guidelines for linking student achievement to teacher evaluation should be further developed through recommendations by the EPAC, lessons learned nationwide, and at the Department.

Teacher Practice

Recommendation 1: Department should review and refine criteria for the selection of teaching practice evaluation instruments to include the following:

- Has a research-base and is shown to be valid and reliable;
- Aligns to and addresses each of the 2011 InTASC Model Core Teaching Standards that identify and describe effective teaching practice;
- Includes classroom observation as a major component with multiple observations for each teacher;
- Requires collection of evidence-based data on the following areas of teacher practice:
 - o The learning environment
 - o Planning and preparation
 - o Instructional practice/classroom strategies and behaviors
 - o Professional responsibilities and collegiality, inclusive of collaborative practice and ethical professional behavior
- Includes a component or process that provides ongoing opportunities for teacher self-reflection of practice;
- Includes rubrics for assessing teacher practice that have a minimum of four levels of performance ratings with guidance on the weighting for the domains of practice;
- Provides a differentiated evaluation procedures for novice and veteran teachers;
- Provides rigorous and deep training for observers that includes video exemplars (across grade levels and subjects) for practice and calibration; Provides for a mechanism for certification or proof of mastery of observers: Certification would be conferred on candidates who have successfully completed additional training and have passed a performance-based test to validate certification, ensuring a high level of competency and inter-rater reliability;
- Provides ongoing coaching opportunities to support observers in implementing observation protocols and providing meaningful and actionable feedback to staff;
- Provides rigorous and deep training for teachers in the framework and its implementation that includes video exemplars that also support professional growth;
- Provides ongoing supports for teachers (i.e., coaching, professional learning opportunities on the framework and its implementation, exemplar videos, etc.); and
- Provides access to a system or process to build capacity at the district/school level (i.e., train-the-trainer module, refresher courses for district trainers, access to updated rubrics, videos, etc.)

Recommendation 2: The Department should create common vocabulary by providing explicit definitions of the words and terms used for the teacher evaluation system; an immediate need is to provide definitions for the terms used in the criteria for selecting a teaching practice evaluation instrument.

Recommendation 3: The Department should develop a list of approved teacher practice evaluation instrument providers which meet the specified Department criteria. Districts would be permitted to choose the State teacher practice evaluation instrument or one of the other approved instruments. Providers not on the approved list may apply to Department for inclusion, but must meet the required criteria. Districts may also develop an instrument, but must meet all the required criteria.

Recommendation 4: The Department should provide explicit guidance for observer training that ensures all observers:

- Are able to demonstrate the required knowledge and skills to accurately assess teacher practice based on the performance elements of the teaching practice evaluation instrument;
- Are able to provide feedback that results in the continuous improvement of teaching practices;
- Have ample opportunities to maintain all skill levels annually;
- Engage in an ongoing calibration exercises; and
- Can articulate and calculate summative evaluation ratings based on teacher practice and student measures.

All observers should be able to show competence in the following skills before beginning formal observations:

- Overview of evaluation reform and the evaluation system
- Types of evidence and methods for evidence collection and analysis
- Identification and articulation of the difference between evidence and opinion
- Facilitation of pre- and post-conference discussions
- Recognition of and reduction of bias in observations
- Practice in understanding and rating differentiated levels of performance
- Use of rubrics for feedback and growth
- Use of a performance management data system
- Calibration, inter-rater agreement, and accuracy

In preparation for beginning the observation and evaluation cycle, all observers should:

- Calibrate, at a minimum, twice a year. Observers should calibrate at the beginning of each school year and at least one other time during the year. If observing teachers who are rated ineffective or partially effective, it is recommended that calibration take place immediately prior to the observation and double scoring be used as an option for ensuring accuracy. The district should have a process for ongoing calibration, as needed.
- Engage in coaching to ensure accuracy and inter-rater agreement with master scorers or experts in the evaluation system.

Recommendation 5: The Department should provide explicit guidance for teacher training that ensures teachers:

- Have an understanding of how effective teaching is defined by the standards
- Have an understanding of the processes that will take place during the observation and evaluation cycle
- Have formal and informal opportunities during the year to practice new skills learned and to receive feedback from colleagues and observers focused on a continuous cycle of improvement
- Teacher leaders should actively participate in observer training for the purposes of support within schools and building a training cadre of leaders and teachers.
- All teachers will be trained in the following prior to being formally observed:
 - o Definition of effective teaching, as defined by the district and the teaching practice evaluation instrument
 - o Expectations of the teaching practice evaluation instrument
 - o Using evidence to inform observation self-reflection and conference discussions
 - o Use of the rubrics with exemplars of performance to inform self-reflection and conference discussions

Recommendation 6: The district should provide adequate training for teachers and observers on the development, implementation and documentation for the additional teacher practice measures. These could include such measures as portfolios of teacher practice or student work samples, student surveys, team log expectations and activity logs, etc. Additionally, teachers and observers should have access to and protocols for use of exemplars of teacher practice across grade levels and subject areas, as well as across differentiated levels of performance.

Recommendation 7: The Department should ensure that no observer may begin formal observations before completing the required initial training and can show that all skills have been mastered through a certification process.

Recommendation 8: The Department should ensure that no teacher is observed for a formal observation until they have completed the required initial training.

Recommendation 9: The Department should consider developing guidance to ensure that novice teachers or teachers who have been identified as partially effective or ineffective have opportunities for additional supports which could include additional observations with pre- and post-conferencing, coaching and mentoring by accomplished teachers in the grade and/or subject area, etc. The observation feedback and ratings should acknowledge the growth of teaching practice over time, particularly in the case of novice teachers or struggling teachers.

Recommendation 10: The Department should engage in additional research on setting appropriate weights for teacher practice. Cohort 2 will be held to 40 percent (minimally) of the weight being placed on the evaluation instrument. The Department should consider allowing districts more flexibility in assigning the additional weights for teacher practice measures which include such measures as portfolios, self-reflection, and student surveys. Teacher portfolios would be developed based on specific goals to be accomplished. The portfolios should incorporate student work samples, unit plan design, and other examples of practice. Self-reflection should be based evidence collected on the observation tool and be used as a basis for conversation at least twice during the year (pre-conference and summative conference). Student surveys should be evidence or research-based.

Recommendation 11: The Department should promote observation practices that include formative processes such as walk-throughs and peer observations. Such processes would include either verbal or written feedback to the teacher for the purposes of growth and support.