There is much about the current attention to teacher evaluation that is to be applauded. The evaluation systems within some New York State Districts were stagnant due to, among other factors, negotiated restrictions that hampered the effective supervision of teachers. For example, some contracts limited how often a teacher could be observed. Others prohibited unannounced observations. Some of the principals with whom I have spoken believed that it was virtually impossible to remove incompetent teachers from the classroom. Often, those who tried had their efforts thwarted by school boards who were dismayed by the exorbitant costs of the 3020a process.

Of course, many other New York State districts had vibrant and effective evaluation systems that reflected a collaborative partnership between teachers and administrators. In such districts, ineffective teachers were dismissed and others received effective professional support in order to improve.

School leaders are committed to the continual improvement of their schools. The principals of New York State historically have neither embraced the status quo nor been resistant to change. Principals do expect, however, that school reforms be effective, sound, and evidence-based.

By recommending and then securing change in the evaluation process in one fell swoop, the New York State legislature brought the potential for a renaissance in the way that principals engage in the supervisory and evaluative process. That potential, however, can only be realized if the system is an improvement over the former evaluation system of each district and, if the unintended negative consequences of the change do not outweigh the benefits.

Unfortunately, the way that APPR is presently designed does not result in an improved system. It will, in the opinion of over 1500 New York State principals, more likely result in diminished learning opportunities for students. Our principals’ letter, which can be found at www.newyorkprincipal.org, outlines why we believe this to be true.

Principals are not alone in their concerns. The Center for Research, Regional Education and Outreach (CRREO) recently released a brief critiquing how Race to the Top has been playing out here in New York, characterizing it as a “grand and costly experiment” that may take education in the wrong direction. The author, Kenneth Mitchell, a district superintendent and executive
committee member of the Lower Hudson Council of School Superintendents, includes in the brief the exorbitant costs to districts of APPR, along with concerns regarding the possible negative effects of APPR on student learning and educational innovation. He calls for the postponement of the use of student achievement data until the preponderance of research—which he cites in the brief—supports the use of student achievement data in teacher evaluation. In addition, the brief calls for a restoration of local flexibility in teacher evaluation.

Principals strongly agree with the concerns expressed in the CRREO brief. We, who are responsible for the supervision and evaluation of our teachers, have serious doubts regarding the efficacy and fairness of the tool we have been forced to use. I am sure that you are aware of those doubts and concerns. The question before your committee, however, is what to recommend to the Governor regarding Teacher and Principal Quality. Do you recommend that he “stay the course” with APPR, or that he follow the advice of the CRREO brief, and work with the legislature to make a mid-course correction?

At the October 11, 2012 hearing in Long Island, two members of the committee asked me to describe how APPR might be changed to ameliorate some of its negative consequences and to provide greater flexibility to districts within the confines of the Race to the Top Grant and the ESEA waiver. What follows is my best attempt to provide a response to that inquiry based on the recommendations of the Principals Letter signed by 1,515 New York State principals and on the input of colleagues and scholars across the state.

Recommendation 1: Reduce the role that student test scores play in the evaluation of teachers.

The American Institute of Research, which prepared the growth scores for the State Education Department, concludes “the model selected to estimate growth scores for New York State represents a first effort to produce fair and accurate estimates of individual teacher and principal effectiveness based on a limited set of data” (p. 35). The carefully chosen words—“a first attempt”—speak volumes. The New York State model is, as described by the CRREO brief, not ready to be incorporated into teacher evaluation. We are not alone; state after state has struggled to use statistical models in a fair and valid way. Let’s withhold judgment on whether these attempts will ever approach what we need. What we should all agree on right now is that we are not there yet.

The usual response from the State Education Department that student growth scores constitute, “only 20%” is disingenuous. Because of the design of the overall point structure (which requires that a teacher’s score exceeds 64 out of 100 to get out of the “ineffective” range), those 20 points have an oversized role in evaluation. They will also be perceived by the public as the “objective” part of the evaluation, calling into judgment the professionalism of the principal who recognizes the talent of the teacher despite the flawed growth score.

Moreover, even if a valid and reliable measure that successfully isolates the effects of an individual teacher on student learning growth were available, these policies that link teacher evaluations to student test scores would still drive negative unintended consequences to student learning opportunities, such as test-prep curriculum, reduced time for the arts and physical
education, and teacher worries regarding the assignment of students with special needs to classrooms and courses. Perhaps the best evidence of this can be found in the schools that have long had a data driven focus—charter schools. The low numbers of students with learning disabilities or who are English language learners in these schools indicate how educators begin to view students when test scores become the focus.

We do know that we must begin with, the ‘given’ that student learning growth must, according to the Race to the Top grant and the ESEA waiver, play a significant role in teacher and principal evaluation. Yet there is no more specificity in either document, nor does either one expressively state what the growth data should be or that individualized growth data must be used to evaluate individual teachers. The following outlines specific suggestions for change within the guidelines.

Presently, there are seven New York State teaching standards. Student achievement should not be a standalone category. Instead, it should be incorporated in the evaluation rubric so that achievement comprises 30%, not 40%, of the evaluation. Student growth would still comprise 20%, (two thirds of the 30%). However, the policy should allow the flexibility for student growth to be included in a way that would better reflect the needs of individual schools and students and that would reduce the unintended consequences as described in the principals’ letter.

Specific recommendations for the inclusion of student growth and achievement in evaluation:

- 10%: A growth measurement of progress based on a growth goal that would be set by the teacher and principal. That goal would be based on an assessment of student need. For example, if analyses of scores or other learning evidence show a need to improve the quality of student writing, a writing growth goal would be determined. This reflects the essence of what a Student Learning Objective should be. However, because schools have been tied to the strictures and restrictions of the State Education Department’s approval process, schools (generally speaking, principals) do not have the freedom and flexibility to create and describe meaningful goals for their teachers if the course ends in a state examination, such as a Regents exam. This measurement should be approved at the local educational agency.

- 10%: A measurement of growth on State examinations. This test-score-based element has been the focus of a great deal of concern, and deservedly so. As a supervisor in charge of teacher evaluation, I am concerned that the numbers themselves lack validity. As an educator, I’m extremely concerned that the incentives created by these policies undermine some of the most important goals I set for my school, its teachers and its students. And as a school leader, I am concerned about damage to the learning and teaching culture I have worked very hard to create. Yet I also know how important it is to many lawmakers that these test scores be included in the evaluation system. The principals’ letter endorsed the idea that the score be shared collectively by a group of teachers, rather than be assigned to an individual teacher. As noted in that letter, student achievement reflects the contribution of the many teachers and other school professionals who contribute to a student’s learning. At the same time, we know from the research of scholar, Michael Fullan, that “the drivers (of improvement) that work motivate teachers to engage in instructional improvement with peers,” not those that encourage teachers to compete. Including a group score would reduce incentives felt by teachers to focus
inordinately on high-stakes tests and would create incentives for teachers to work collaboratively to improve their practice. In addition, which student is on the teacher’s roster would not be a source of worry. The strongest teacher in the group would be incentivized to teach the students who are the most difficult to reach.

- 10%: An achievement measurement to be determined through local collective bargaining, as described in the present APPR plan. This could be a school-wide index based on scores, a grade level index, or an achievement measure. The flexibility exists in the law and we support that flexibility.

**Recommendation 2: Provide greater latitude in the choice of an evaluative rubric for the measurement of the performance of teachers and principals.**

Presently there are a limited number of approved rubrics for the evaluation of teachers. They are authored by vendors, NYSUT and the District of Columbia school system. In some cases, the rubrics must be purchased. All of the vendors sell training services connected to their proprietary rubrics.

Many of the approved rubrics are overly complex and difficult to use. For example, the NYSUT rubric contains over 90 indicators which must be considered. Although numerous districts with excellent student achievement, such as the Great Neck School District and the Bellmore Merrick School District, submitted their rubrics for approval, approval was denied by the State Education Department. Why any successful school district would be denied permission to use their thoughtfully prepared rubric, yet permitted to use the rubric developed by the Washington DC School district, one of the lowest performing districts in the nation, is difficult to understand.

I urge the committee, therefore, to permit the use of district rubrics, developed by local teacher associations and school leaders and which reflect the goals and needs of the local district. Teachers and school leaders deserve the right to work together to develop teacher practice rubrics if they choose. This will spark creativity and innovation while allowing districts to focus on the most needed improvements to educator practice.

**Recommendation 3: That the APPR system move from a point-based system to a rubric based system.**

As noted in Commission member Sara Mead’s recent paper on teacher evaluation, there are “strong tensions” between using an evaluation system to improve teaching and using it as a system to dismiss teachers. As Ms. Mead and her colleagues note, “For better or worse, the design focus in most 2.0 evaluation systems has been on building systems that are sufficiently fair, valid, and reliable to hold up in court as a basis for dismissing low-performing teachers.”

This suggests a very important truth: Even if a small percentage of ineffective educators were dismissed due to APPR, there would be little benefit to students if the practice of those teachers who remained in the classroom did not improve. The current point-based system, in which a teacher must accrue the majority of points (65 or better out of 100) to avoid being labeled ‘ineffective’ and subject to dismissal, does not foster the collaboration and reflective practices needed for professional improvement. Again, as noted in the CRREO brief, the “anxieties of
low scores” will propel teachers (and principals) to abandon new creative practices that will benefit students and their professional growth as they seek a “narrow professional agenda” designed to keep their score in the effective range and maximize their points.

Please allow me to stress that this is not guesswork. We now have a decade of experience with No Child Left Behind. We see how it has worked—both positively and negatively—to change schooling practices. School leaders and other educators are human; we respond to incentives. The higher the stakes (e.g., job security), the more likely we are to change our practices in whatever ways are necessary to garner the necessary outcomes. Sometimes these changes benefit students; often they do not.

A rubric-based design that blends professional practice (70%) and student achievement (30%) would shift the focus from scores to practice. The overall ratings could be based on the preponderance of ratings on the rubric. Consider a design, for example, with ten categories—seven based on the seven teaching standards with the remaining three based on achievement as described in Recommendation 1. Rules such as “a teacher is rated highly effective overall if rated highly effective in seven out of ten categories, with no developing or ineffective ratings in any category” could easily be developed, and the improvement of practice would become more meaningful and achievable for teachers.

I thank you for the opportunity to provide input. I would again emphasize the above does not reflect my thinking in isolation. It is my attempt to provide concrete alternatives based on the research on organizational change, the research on the use of test scores for high stakes evaluative purposes, and the opinion of more than one thousand five hundred principals across New York State. Ignoring the professional wisdom and practice of those who are mandated to carry out APPR will ensure its failure as a tool to improve instruction. Our teachers and our students deserve better.

Sincerely,

Carol Burris, Ed.D.

---

4 The BETA report of AIR discusses the influence of student ‘peer effects’ on teacher scores. In addition, two recent studies show how the track that a teacher teaches influences her growth score. Teachers of low-track classes are disadvantaged. See http://www.edweek.org/ew/articles/2012/10/24/09tracking_ep.h32.html


7 Presently, that component comprises 20%, to be reduced to 15% when the state moves to a VAM measurement

