

# Five Ways to Improve Teacher Evaluations in New York

By Thomas J. Kane

Walter H. Gale Professor of Education

May 5, 2015

## 1. Allow teachers to accumulate a long-term track record. Especially for those with a distinguished track record, one bad year should not end a career.

As in many professions (including higher education), a past history of success signals that a teacher has the talent and accumulated skill to be successful in the future. The only reason to place greater than proportional weight on their most recent performance is to preserve teachers' incentive to maintain effort and meet new challenges. It's inappropriate to continue asking an accomplished teacher, "What have you done for us lately?"

For both the student achievement and the classroom observation categories, all available years of observations should be given the same weight, unless the weight placed on the most current year would be less than 20 percent (that is, if there are more than 5 years of data). In such cases, the prior years should be weighted at 80 percent and the current year weighted at 20 percent. In other words, someone teaching for three years would have 33 percent of their student achievement or observation category score determined by their most recent year; in contrast, someone teaching for 11 years would have 80 percent of the weight placed on their performance during the previous 10 years (roughly 10 percent per year) and 20 percent of the weight placed on their eleventh year.

Of course, in many districts, past classroom observations have not meaningfully differentiated between strong and weak teaching practice. Therefore, for classroom observations, observations prior to 2014-15 should not count toward one's longer term track record. All teachers—no matter where they are in their careers-- would be in the same position starting next year. Given the changes in instruction required by the Common Core standards, that may be an entirely appropriate signal to send. However, 5 years from now, experienced teachers would have a five-year track record to be judged by and would be less vulnerable to a single year's observation.

In contrast, NY State began using an assessment aligned to Common Core in the spring of 2013. The state should allow teachers to include their average student growth percentiles from the past two years as well as the current year for the student achievement category. (To facilitate comparability across years and across tests, NYSED should propose a method for standardizing the student achievement growth measures, either in terms of standard deviations or percentiles.)

## 2. Define "ineffectiveness" for probationary teachers.

Any teacher evaluation system must both support high stakes personnel decisions and provide professionals with the feedback they need to improve. However, the appropriate balance between accountability and supporting continued development is likely different for a

probationary teacher on the cusp of earning tenure and an experienced teacher who has already gained tenure. A tenure decision is one of the most consequential decisions a school system makes. For a teacher approaching the end of the probation period, the evaluation system should generate the best possible judgement of a teacher's likely future success in helping students achieve. For that purpose, a teacher's past success in promoting student growth is the single most powerful predictor. However, after tenure, the need for prediction is less pressing, and the primary goal should be to allow professionals to see where they need to continue to improve.

Therefore, parameters such as the cut score for ineffectiveness, the total number of classroom observations, the number of observations by external observers and the weight placed on observations by external observers could differ for probationary and experienced teachers.

Under the new statute, any teacher designated as "ineffective" in his or her fourth year will be prohibited from earning tenure. The state should carefully define the cut score for ineffectiveness, rather than leave it entirely to the subjective judgements of principals.

We cannot forget that every tenure decision involves two teachers-- the probationary candidate and a novice teacher who would be happy to teach their class next year. Commentators tend to focus on the rights of the probationary candidate, while ignoring the interests of the anonymous potential replacement and the students either teacher would teach next year. The definition of "ineffectiveness" should serve as a reminder of the implicit choice being made every time a teacher earns tenure.

Every teacher struggles in their first year of teaching, but most improve during their second through fourth year of teaching. That implies a criterion-based way of defining "ineffectiveness": a teacher should not earn tenure if he or she has not surpassed the effectiveness of the average first-year teacher with no prior teaching experience by their fourth year. In other words, a fourth year teacher is "ineffective" in the student achievement category if, by their 4<sup>th</sup> year of teaching, their average student growth percentile is lower than that of the average novice in the district. Likewise, a fourth year teacher should be designated "ineffective" in the classroom observation category if their average classroom observation by an external observer is less than the average observation of novice teachers in the district.

For probationary teachers, I would recommend placing disproportionate weight on the scores of the external observers. Why? A tenure decision carries huge implications for thousands of future students, parents, colleagues and supervisors. The process should explicitly recognize their interests, by ensuring that external observers ratings play a major role in evaluations. That said, the building supervisor will know about aspects of a teacher's contribution not visible to outsiders. Therefore, they should be allowed to override an "ineffective" rating and convert it to "developing"—but, when principals do so, the burden of argument should rest on them. They should be required to document their reasons and gain the approval of their local district whenever they seek to relabel an otherwise "ineffective" teacher as "developing".

Such a standard would have a number of advantages: First, it reminds everyone that a promotion decision involves a choice (albeit usually implicit) between two teachers—the probationary teacher and an anonymous novice. Linking the standard for tenure to the

effectiveness of the average first-year teacher highlights the implicit choice made in every tenure decision.

Second, it would be a self-adjusting standard: if classroom observation scores become inflated over time, if different districts choose different rubrics, or if the quality of those willing to enter teaching were to decline (or rise), the threshold for tenure would adjust accordingly.

Third, by relying on the scores given by external observers, the tenure decision would no longer be at the sole discretion of the local principal. A tenure decision involves thousands of future students as well as future colleagues and supervisors at other schools in a district where a teacher might work. It makes no sense to leave the decision in the hands of their current supervisor alone.

If tenure protections were reserved only for accomplished teachers, just imagine how different our schools would be. If such a system had been in place decades ago, when it was already clear that there was evidence on the importance of individual teachers on student achievement, we might not be having this debate today.

### **3. External observers will make the system more honest (and fair).**

We cannot expect better student outcomes without major changes in teaching. But that's adult behavior change, and we should not underestimate how difficult that will be. No one would launch a Weight Watchers club without any bathroom scales or mirrors. Student achievement gains are the bathroom scale, but classroom observations must be the mirror.

However, in order to be helpful, the process must be honest. In a recent research project, we sent videos of teachers' lessons to their own principal, as well as to other principals and peer observers to score on their observation rubric.<sup>1</sup> We found clear evidence of that teachers enjoy a large "home field" advantage when their principals are doing the observing—with principals rating their own teachers higher than principals from other schools. (Every principal in the study were rating videos from their own teachers as well as those of teachers from other schools.)

A number of districts-- such as Washington DC and Hillsborough County (Tampa), Florida—already use external observers in their observation process. In the initial years of implementation, external observers in DC tended to score teachers lower than the teachers' own principals. However, over time, those ratings have converged. (In fact, teachers in Washington and Hillsborough demanded external observers, to ensure that they would be treated fairly.)

A second opinion from an external observer will help in other ways as well. Even with trained raters, the typical rubric requires judgement, and different raters will form different judgements. For that reason, there is a much larger improvement in reliability from adding

---

<sup>1</sup> Andrew D. Ho, Thomas J. Kane *The Reliability of Classroom Observations by School Personnel* (Seattle, WA: Bill & Melinda Gates Foundation, 2013).

observations from different observers than from adding more observations by the same observer.<sup>2</sup>

However, because of the time commitment, external observations are also costly, and they should be used wisely. During the 2015-16 school year, the highest priority should be given to external observations for probationary teachers. For instance, all probationary teachers could be required to have at least two full-lesson observations by external observers during 2015-16. Second priority should be given to already tenured teachers with poor student achievement growth in the prior year. Once observers are trained and processes are in place, the remaining experienced teachers could be added in subsequent years, perhaps with one full-lesson observation by an external observer.

External observers could be administrators or teachers at other schools, retired administrators or teachers, etc. In order to be certified as an external observer, the candidates should be trained. At the end of their training, they should score a set of pre-scored videos of lessons. Those with more than a maximum number of discrepancies should not be allowed to serve as an external observer without receiving additional training. Districts should track the scores given by external observers and administrators and look for discrepancies between external observers and administrators. The district should look into cases where an observer seems to be far off from all the supervisors he or she works with or when principals give ratings with are far off from the external observers. For instance, when they differ by more than one category, the district might assign a second external observer.

#### **4. Reduce the logistical barriers to using external observers.**

As valuable as they might be, external observers will also present significant logistical challenges. A lot of time could be wasted as observers drive from school to school.

NYSED could ask for additional funding and invite applications from districts with innovative ideas for implementing external evaluations. For instance, some districts may allow teachers to submit videos to an external observer in lieu of in-person classroom observations. The use of video would have a number of advantages. For instance, teachers usually struggle because of the clues they are not noticing. Yet, it is impossible for teachers to remember something they did not notice in the first place. Therefore, it is difficult for teachers to recognize their mistakes simply by reading an observer's written notes. It would be much more powerful to allow teachers to see themselves (and their colleagues) teach.

But how could that be done? The National Board for Professional Teaching Standards has been allowing teachers to submit videos for more than 20 years. Giving teachers control of a camera, the opportunity to watch themselves teach, and allowing them to discuss their videos talk with

---

<sup>2</sup> Thomas J. Kane and Douglas O. Staiger *Gathering Feedback on Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains* (Seattle, WA: Bill & Melinda Gates Foundation, 2012)

external observers, peers and supervisors will provide be a more effective mirror than any observer's written notes in helping teachers transform their teaching.

The use of video would save observers' travel time. Moreover, it would make it much more feasible for observers working at other schools to participate. There would be other advantages as well. Harried principals could do their observations during quieter times of the day or week. And when principals do not have sufficient content expertise, they could solicit the views of content experts. Finally, video evidence would level the playing field if a teacher ever has to defend their teaching against a principal's written notes at a dismissal hearing—a teacher's video vs. an observer's written notes.

Video is now widely used to coach improvements in activities such as athletics and dance and public speaking. Why not teaching?

## 5. Incorporate student feedback.

Although it would require a change in the law, the Regents should recommend the inclusion of student feedback in teacher evaluations. Although it would be unwise to rely solely on student feedback, the results of student surveys have been shown to be related to a teacher's student achievement gains. If one is concerned about placing too much weight on student test scores—or about the lack of student achievement measures in certain grades and subjects—there is only a limited number of alternative sources to rely upon: classroom observations are one, but student surveys are another.

There are other advantages as well: First, student surveys provide an opportunity to measure other student outcomes—such as their engagement in the subject, their future plans to study the subject, the amount of test-prep they did during the semester, etc. Student surveys should be an especially valuable addition to those most concerned about “teaching to the test”. Second, like the classroom observation, a student survey can provide actionable feedback on specific teaching practices, such as time management and the quality of feedback to students.

Student evaluations are ubiquitous in higher education. In fact, in many universities, student evaluations are the only measure of teaching effectiveness used. They have been shown to be valid and reliable measures, at least in upper elementary, middle school and high school grades.<sup>3</sup> Recall that one of the best ways to improve reliability for classroom observations is to average over multiple observers. Even if a trained adult is more discerning than the typical student, it is less costly to average over 25 students in elementary school or 75-100 students in middle schools and high schools than it is average over the same number of adult observers.

---

<sup>3</sup> Thomas J. Kane and Douglas O. Staiger *Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project*, (Seattle, WA: Bill & Melinda Gates Foundation, 2010)