

TO: Beth Berlin, Acting Commissioner; Ken Wagner, Senior Deputy Commissioner; Julia Rafal-Baer, Assistant Commissioner – State Education Department

FROM: TNTP

DATE: May 13, 2015

RE: Section 3012-d: Implementation of Independent Evaluators

Introduction

On March 31, 2015, new requirements for the evaluation of teachers and principals were passed as part of the 2015-16 New York State Budget, which added a new section, 3012-d, to the existing education law. Under the new requirements, teachers will continue to be observed by their building administrators, but they will also be observed by impartial, independent evaluators.

TNTP supports the decision to include multiple observers in evaluation as one strategy to combat the inflated ratings distributions across the state and to provide additional, important feedback regarding teacher practice. In a study of the peer evaluator system in Hillsborough County, MET researchers found that that increasing the number of observers more than doubles the reliability of a teacher's rating compared to increasing the number of observations performed by the same observer.ⁱ If implemented well by districts, independent evaluators have the potential to provide an important, non-biased perspective on teacher performance that, when used in combination with principal observations, can lead to more accurate ratings.

This memo presents key considerations for the implementation of independent evaluators in teacher evaluation and examines five school districts which have implemented independent evaluators in recent years.

Key Considerations

The following considerations are based on the multiple evaluator systems we looked at in Washington DC, New Haven, Newark, Denver, and Hillsborough and best practices we have identified from designing and implementing teacher evaluation systems across the country. With the exception of Hillsborough, which was featured in the Gates MET study, to our knowledge, none of the other programs have been formally studied so as to provide conclusive evidence of their effectiveness. Because of this, and because of the many considerations listed below that are specific to local context, we recommend that the SED set some broad criteria, while leaving the majority of design and implementation decisions around independent evaluators to districts.

1. **While the primary rationale for implementing an independent evaluator system may be to bolster the reliability of observation ratings, independent evaluators can also serve as an important source of content-specific feedback for teachers.** In all five of the districts examined, administrators cite reliability as a key driver behind the formation of a peer/third party evaluator system. However, in at least three of the districts below, being able to provide teachers with additional sources of feedback – often content-specific - along with development opportunities also represented a tremendous value-add to the existing evaluation system. Where possible, we recommend framing this role as an additional, and where possible, content-specific perspective that supplements that of the principal.
2. **Serving as an independent evaluator is most commonly a full-time position held by former teachers, or current teachers taking a temporary leave of absence.** While all of the districts we examine use different recruitment methods to find and hire their independent evaluators, the one constant is that all evaluators are hired in a full-time capacity to allow evaluators enough time to observe teachers and provide meaningful feedback. Most districts recruit Effective or Highly Effective teachers for this role. If targeting this population, it is important to consider a salary for independent evaluators that is at least equivalent to the teacher's current position, which may be higher than average if a pay-for-performance compensation system exists. Further, we recommend requiring some performance-based screen at the end of the hiring process and not relying solely on past ratings of Effective or Highly Effective, as we know that 1) many systems are prone to inflation, and 2) being

a strong teacher does not mean that someone will automatically be a strong independent observer or deliver high-quality feedback.

3. **The average caseload for an independent evaluator appears to be around 100 teachers or fewer depending on the level of support and development they are expected to provide to teachers.** In three of the districts below, caseload of teachers per independent evaluator hovers around 100 or just below. Keeping caseloads at a manageable size allows evaluators to provide higher quality feedback and support to teachers in the field and, thus, improve teacher practice across the district. Real-time feedback is a crucial element to improving teacher practice, which is why we recommend that evaluator caseloads be kept small enough to support this type of immediate feedback loop. Districts who consider such a caseload to be unmanageable or unaffordable should consider the use of video observations.
4. **Ensuring that independent evaluators are normed correctly is time intensive and crucial to the reliability and success of these programs.** Based on our own experiences, training and norming evaluators well takes a large upfront investment in district time and resources, as well as regular, ongoing norming sessions throughout the year. All five districts cited in the case studies below required their evaluators to engage in comprehensive training and norming practices, with DCPS's six week summer orientation standing out as the most extensive. In addition, four out of the five districts contracted with an outside consultant to assist in the creation, implementation and management of these training programs.
5. **There are some commonly applied cost-saving strategies for districts implementing these programs.** Districts should be provided with flexibility in how they implement this component in order to reduce costs. For example, in the MET study in Hillsborough, video observations were frequently used to provide multiple evaluators access to the same lesson without evaluators having to travel to schools. Additionally, many systems have chosen to reduce costs by only having a subset of teachers observed by independent evaluators. We recommend districts consider this specifically for more novice or probationary teachers, teachers at the high or low end of the performance spectrum, or a rotating sample of teachers.
6. **When included as a weighted measure in a teacher's evaluation, independent observations typically count for between 10-30% of a teacher's rating.** When setting a weight for any evaluation component it is important to consider its overall reliability and value as well as the other components included in the system. It will also be important to consider the frequency and duration of independent evaluator observations. We believe that the weight of independent evaluator observations should not exceed the weight of principal observations. If principals are to be the instructional leaders and talent managers in their schools, their determination of teacher performance should hold considerable weight. Based on our experience, weighting the independent evaluator component at 10-15% should be substantial enough to affect the final rating and begin to change practice.

Case Studies

Please note that the information below was gathered via public documentation on each system, as available. Although we attempted to include the most recent information on each system, some public documents date back to the initial years of program implementation, and thus, may not include details that have changed over time.

DCPS's Master Educator Program

Background: The District of Columbia Public Schools' (DCPS) Master Educator Program was introduced in 2009 as a part of IMPACT, the district's teacher evaluation and compensation system. It is the direct result of teacher feedback received from over 50 focus groups, in which teachers expressed a desire to be evaluated by an objective content expert in addition to their school administrator. The program currently consists of over 40 Master Educators spanning 13 content areas, each with a caseload of about 75-80 teachers.ⁱⁱ

Role of the Evaluator: The Master Educator (ME) serves as an additional, impartial evaluator for all teachers, conducting two out of a teacher's five required observations each year. All observations are weighted equally, with the lowest score (ME or administrator) discarded prior to final rating calculation. In addition to their evaluative duties, MEs provide limited coaching and support to teachers, including a one-on-one developmental conversation following each observation.ⁱⁱⁱ

Selection and Training: DCPS recruits nationally for its ME positions and targets candidates with a variety of content and grade level expertise. All candidates must have had at least five years of teaching experience. Candidates are required to undergo a five stage interview process which includes 1) submitting a resume and four short essays, 2) a performance task, 3) a phone interview, 4) an in-person interview with the ME Selection Team, and 5) a one-on-one interview with the Director of the ME Program. If selected to participate in the program, MEs must complete a six-week summer orientation, which focuses on the Teaching and Learning Framework as well as norming and ratings calibration.^{iv} DCPS now offers an extensive video library, which supports with norming across MEs and principals.

Results: Data from DCPS demonstrates a stronger ratings distribution than in almost any other state or district. During the 2013-14 school year 77% of teachers in DCPS were rated as Effective or Highly Effective (the top two rating categories) through the IMPACT program.^v While the ratings distribution may not be completely attributed to the ME program (there are other factors such as cut scores that are critically important when looking at ratings distributions), it is possible that this program helped to control for inflation in classroom observation ratings.

New Haven's Third-Party Validator Program

Background: In 2010, as part of contract renegotiations, New Haven Public Schools (NHPS) reformed its evaluation system to include third-party validators to be used as a norming mechanism, focusing on performance outliers (i.e. teachers receiving either the lowest or highest ratings).^{vi}

Role of the Evaluator: The primary responsibility of the validator is to provide additional observations to teachers on-track to be rated either Exemplary (E) or Needs Improvement (NI), New Haven's highest and lowest ratings.^{vii} These additional observations are conducted jointly with the teacher's school administrator, during which both the validator and administrator observe the lesson and complete a validation review form. Teachers who have been identified as likely to receive an E receive two additional observations – one announced and one unannounced – while teachers on track to receive an NI receive three additional observations – one announced and two unannounced.^{viii} In the event that the validator and administrator cannot agree on an observation rating, the final decision is made by the Assistant Superintendent in collaboration with the local union president.^{ix}

Selection and Training: Third-party validators must be former teachers who have demonstrated effectiveness in the classroom and are not currently affiliated with the district. Currently, NHPS contracts with ReVision Learning Partnership LLC to recruit, train and manage its validators, who include: administrators from other districts, retired

teachers and administrators from the district, and retired teacher prep program professors. All validators undergo a four-hour intensive training that focuses on providing low-inference feedback to teachers.^x

Newark's Peer^{xi} Validator Program

Background: Newark Public Schools (NPS) first implemented its Peer Validator Program in the spring of 2013 as a means of ensuring a fair and valid evaluation system, while simultaneously supporting teachers by providing an additional source of feedback.

Role of the Evaluator: NPS uses peer validator observations as a norming mechanism and as an additional development resource for teachers. Teachers who have received various ratings, with an emphasis on Ineffective and Effective, are selected to receive peer validator observations throughout the year. If a teacher receives an Ineffective rating on their Annual Evaluation or if a School Improvement Panel (SIP) feels a teacher needs additional support, the SIP may specifically request a peer validator visit. Once the need for a peer validator observation has been identified or requested, the peer validator conducts an unannounced observation of the teacher and schedules a post-conference conversation with the teacher to review the evidence collected during the observation. The teacher's principal will also receive a copy of the completed observation form from the peer validator.^{xii}

Selection and Training: Peer validators can be current teachers, former teachers, administrators from NPS or other school systems, academics, or outside experts.^{xiii} Currently, NPS contracts with ReVision Learning Partnership LLC – an organization comprised of mostly former teachers and administrators who specialize in classroom observation – to conduct the peer validation process. In addition, Assistant Superintendents and Special Assistants for Teacher Quality (SATQs), both of whom are NPS administrators, are also authorized to conduct validation visits. While the Superintendent is encouraged to consult with the local union president on peer validator candidates, the ultimate authority over selection criteria, process and peer validator management remains with the Superintendent.^{xiv}

Program Oversight: The Peer Validator Program is overseen by the Peer Oversight Committee (POC), which is comprised of five representatives from the union and five representatives from the district, and meets quarterly. The POC is responsible for providing guidance and recommendations regarding the implementation of the peer validation process and offering suggestions for improvement in the following year.^{xv}

Denver's Peer^{xvi}/Mentor Evaluation Program

Background: Denver's Peer Observation (PO) Program was piloted in the 2011-12 school year as part of the Leading Effective Academic Practice (LEAP) evaluation program. The program was first used for stakes during the 2013-14 school year. Denver Public Schools (DPS) currently employs 49 POs, each with a caseload of about 90-100 teachers.^{xvii}

Role of the Evaluator: POs are used to conduct two unannounced teacher observations each school year for the following groups: 1) novice teachers new to DPS and 2) teachers who do not have sufficient data to receive a Professional Practice rating. All other DPS teachers, regardless of their rating in the prior year, are divided into two alternating groups and receive peer observations every other year.^{xviii} PO ratings are weighted equal to and averaged with administrator ratings for a total of 30%-35% of a teacher's overall rating (in off years, this component only consists of administrator observation ratings). Following an observation, the PO will conduct a post-observation conference with the teacher in which they provide feedback and professional development resources.^{xix}

Selection and Training: All POs must be recognized as highly effective teachers (through recommendations, CSAP growth results, etc.), have recent experience in an urban classroom setting, and content expertise in the subject area of the teacher they are observing. Candidates for the PO position must go through a four-stage interview process, which includes: 1) submitting a resume and cover letter, 2) a panel interview, 3) giving an observed lesson and mock teacher feedback, and 4) a one-on-one interview with either the Director of POs or the Executive Director of Educator Effectiveness. While all POs are required to be certified to teach in the state of Colorado, prior experience in DPS is not required. Before assuming their role, POs participate in over 100 hours of training that focus on norming practices, giving low-inference feedback, and bias reduction.^{xx}

Hillsborough's Peer/Mentor Evaluation Program

Background: The Peer/Mentor Evaluation Program was implemented in Hillsborough Public Schools (HPS) in 2011 as part of a larger strategy to change the evaluation and compensation structure in the district and create a teacher career ladder. There are currently about 100-200 Peer and Mentor Evaluators working in HPS.

Role of the Evaluator: Peer Evaluators (PEs) and Mentor Evaluators (MEs) are former HPS teachers who have agreed to leave their schools to serve a two to three-year term as a full-time evaluator before returning to the classroom. PEs are assigned to experienced teachers in HPS based on aligned content-area and are responsible for conducting between two and eight formal observations (depending on the teacher's prior rating and level of performance) each year, along with pre- and post-observation conferences. MEs are generalists assigned to novice teachers in HPS and are responsible for meeting with new teachers weekly, providing targeted support, and conducting evaluations for novice teachers outside of those that they mentor directly. While PEs have a higher caseload of approximately 100 teachers, MEs carry a lighter caseload of 15 teachers to provide them with more time for thoughtful feedback and coaching.^{xxi} Peer evaluations now comprise 25% of a teacher's overall rating.

Selection and Training: PEs and MEs are chosen by a selection committee from a pool of teachers with strong evaluations and at least five years of experience. The initial screening committee is comprised of a teacher, an administrator, a curriculum and/or instructional specialist, and the Director of the Peer/Mentor Evaluation Program, but there is a secondary committee consisting of a large and diverse group of teachers, administrators, and representatives from the Hillsborough Classroom Teachers Association to further narrow down the applicant pool prior to the interview stage. The selection process consists of submitting a resume, cover letter, and essay-style application and then if screened through, an in-person interview and simulation activity. Before assuming their new roles, PEs and MEs complete a six-hour prerequisite online course followed by a week of in-person training which includes rubric calibration and norming, classroom observations and practicing delivering teacher feedback. In the past, HPS has contracted with Cambridge Education and the New Teacher Center to assist in the creation of this training program. Cambridge also provides ongoing calibration and data reporting support to evaluators throughout the year.

Results: As mentioned earlier, the MET study in Hillsborough found that increasing the number of observers more than doubled the reliability of a teacher's rating compared to increasing the number of observations performed by the same observer.^{xxii}

Conclusion

Overall, we believe that the State's decision to include multiple evaluators as part of teacher development and evaluation is a good one. Not only does the addition of multiple evaluators have the potential to boost the rigor and reliability of existing systems as the MET study has demonstrated, but it can also provide teachers with valuable opportunities for feedback and support in the field.

However, while TNTP supports the use of multiple evaluators in teacher evaluation, we also recognize that the use of independent evaluators as an additional observation source requires a significant lift on the part of districts. The State Education Department can play a vital role in the roll out of independent evaluators by furnishing districts with the guidance they will need in order to develop effective systems of their own, while still allowing districts the flexibility to implement these systems according to their specific needs and limitations. Once established, the SED will be able to ensure these programs' continued improvement by sharing best practices and lessons learned from across the state.

ⁱ Ho, A.D. and Kane, T.J. (2013). The reliability of classroom observations by school personnel. *The MET Project*. Retrieved from http://www.metproject.org/downloads/MET_Reliability%20of%20Classroom%20Observations_Research%20Paper.pdf.

ⁱⁱ DC Public Schools. (2015). An overview of IMPACT. Retrieved from <http://dcps.dc.gov/DCPS/In+the+Classroom/Ensuring+Teacher+Success/IMPACT+%28Performance+Assessment%29/An+Overview+of+IMPACT>.

-
- iii DC Public Schools. (2015). Master educator position overview. Retrieved from <http://dcps.dc.gov/DCPS/About+DCPS/Career+Opportunities/Lead+Our+Schools/Master+Educators/Position+Overview>
- iv DC Public Schools. (2015). Application process. Retrieved from <http://dcps.dc.gov/DCPS/About+DCPS/Career+Opportunities/Lead+Our+Schools/Master+Educators/Application+Process>
- v Chandler, M.A. (2015, February 24). IMPACT score ward-by-ward in DC. *The Washington Post*. Retrieved from <http://www.washingtonpost.com/news/local/wp/2015/02/24/impact-scores-ward-by-ward-in-d-c/>
- vi New Haven Public Schools. (2010). Teacher evaluation and development – Introduction to the process. Retrieved from http://www.nhps.net/sites/default/files/1_NHPS_TEVALDEV_Introduction_-_Aug_2010.pdf
- vii These determinations are made by November 1st using available teacher observation data and ratings.
- viii New Haven Public Schools. (2010).
- ix New Haven Public Schools. (2010). NHPS Teacher Evaluation and Development Process Teachers and Administrators Guide. Retrieved from http://www.nhps.net/sites/default/files/2_TEVAL_-_Teachers_and_Administrators_Guide_100825.pdf
- x New Haven Public Schools Teacher Evaluation and Development (TEVAL). Retrieved from <http://www.shaker.org/Downloads/NHTEVAL.pdf>
- xi It should be noted that the use of the word “peer” in this context is not meant to signify a colleague within the same school, but rather an independent evaluator with an instructional background (i.e. a current or former teacher or administrator).
- xii Newark Public Schools. (2015). Teacher talk: Peer validators. Retrieved from http://www.nps.k12.nj.us/?wysija-page=1&controller=email&action=view&email_id=44&wysijap=subscriptions#peer-validators
- xiii Newark Public Schools. (2015). Framework for teaching: Newark public schools teacher performance evaluation. Retrieved from <http://www.nps.k12.nj.us/wp-content/uploads/2014/08/NPSTeacherEvaluationGuidebook2014-15.pdf>
- xiv Newark Public Schools. (2015). Teacher talk: Peer validators.
- xv Newark Public Schools. (2015). Memorandum of agreement between the Newark State Operated School District and Newark Teachers Union: October 18, 2012. Retrieved from <http://www.nps.k12.nj.us/wp-content/uploads/2014/08/TeacherContractMemorandum.pdf>
- xvi Again, “peer” here does not refer to a teacher colleague within the same school, but rather former teachers who have become full-time independent evaluators.
- xvii Denver Public Schools. (2015). Peer observation: Third-party feedback with first-hand knowledge. Retrieved from <http://leap.dpsk12.org/Professional-Practice/Observation/Peer-Observation>
- xviii Ibid.
- xix Denver Public Schools. (2015). LEAP Handbook: 2014-15. Retrieved from http://leap.dpsk12.org/LEAP/media/Main/PDFs/2014-15-LEAP-handbook-master_1.pdf
- xx Denver Public Schools. (2015). PO experience and training. Retrieved from <http://leap.dpsk12.org/Professional-Practice/Observation/Peer-Observation/Experience-and-Training>
- xxi Marietta, G. (2011). The roles of peers and mentors. *Case Study for the Hillsborough County Public Schools*. Retrieved from <http://communication.sdhc.k12.fl.us/eethome/casestudies/Roles%20of%20Peers%20and%20Mentors%20in%20Hillsborough%20County%20Public.pdf>
- xxii Ho, A.D. and Kane, T.J. (2013).