

Maryland Teacher and Principal Evaluation Field Test

Statewide Summary Report

*Results of the LEA Site Visits
March, 2013*

Megan Dolan, Ed.D
April 23, 2013

Maryland Teacher and Principal Evaluation Field Test

LEA Site Visit Focus Groups

Statewide Summary

Megan Dolan, Ed.D.

Background: Race to the Top and Educator Evaluation

In September 2011, as part of Maryland's Race to the Top initiative, seven of the state's LEAs began to pilot a variety of educator evaluation practices including both professional practice and student growth. At the same time, educational leaders at the Maryland State Department of Education (MSDE) worked with the Maryland's Council for Educator Effectiveness and other stakeholder groups to develop guidelines for LEA evaluation systems and a state default model that LEAs would use if they could not come to agreement with their local bargaining unit on an LEA model. Representatives from the seven pilot LEAs (Baltimore City, Baltimore County, Charles County, Kent County, Prince George's County, Queen Anne's County, and St. Mary's County) met each month to discuss emerging issues and possible promising practices, and examine the potential benefits and challenges of various evaluation approaches. As the year progressed, state and LEA leaders recognized that stakeholders at all levels would benefit from additional time to continue their investigation of evaluation practices and to address the vast logistical issues raised by the new systems. The state also wanted to ensure that the LEAs that were not participating in the pilot had adequate time to prepare, so that when the systems were fully implemented, they would be fair, credible, and useful. While all of Maryland's LEAs had some form of educator evaluation system in place, some were more intensive than others, and the requirements under RTTT represented a dramatic conceptual shift for many LEAs. Maryland was granted an additional year to conduct a field test of TPE activities and determined that a critical part of this process was an on-the-ground review of LEA activities.

Data Collection: Focus Groups

Between February 24 and April 3, 2013, 66 focus groups were conducted in the 22 LEAs participating in the Teacher and Principal Evaluation (TPE) Field Test. In each LEA, the Field Test Monitor (FTM) met separately with three focus groups consisting of central office personnel, school level administrators (including principals, assistant principals, and content supervisors), and teachers. On February 21, 2013, the FTM made initial contact with the LEA representatives at the monthly field test meeting and provided an overview of the data collection process for the monitoring visits. The local Field Test Points of Contact for each LEA were asked to arrange the meetings as follows:

- One hour with 3-6 individuals from the central office (human resources, executive officers, curriculum specialists, data managers, etc.)
- One hour with 3-6 principals , which could include assistant principals who are conducting evaluations or training
- One hour with 6-9 teachers, including a representative sample of subjects, tested areas, and levels when possible

Other than this guidance, the local POCs had discretion regarding which participants were selected, and they were sometimes hampered by scheduling issues. The data collection window was short and coincided with MSA testing and Spring Break activities in many LEAs. There were also two weather-related incidents that impacted scheduling. Even within these limitations, almost all of the focus groups included a wide range of stakeholders and seemed to present a good sample of the local field test participants. Most of the LEAs included at least one participant from a local union or association, and in some cases association leaders attended. It should also be noted that all of the POCs were extremely helpful in making arrangements for the visits, communicating with the FTM and the local participants, and providing additional information on request.

At the start of each focus group, participants were assured that all data would be collected and reported anonymously. They were also provided with contact information for the FTM in case they had additional comments, questions, or concerns, and encouraged to share that information with their colleagues. All of the discussions were scripted in real time with as much detail as possible. Many participants also provided documentation and artifacts to support these discussions. In addition to this statewide summary report, the data will also be analyzed and synthesized into 22 LEA-specific reports to be shared with the local superintendents.

The focus group conversations centered on the activities occurring in each of the LEAs, the roles played by the focus group participants, and their perceptions of the process. Each participant had the opportunity to share his or her experiences, ideas, concerns, questions, and suggestions. Many had also gathered input from their colleagues in anticipation of the meeting and were able to share those additional perspectives. Most of the participants were extremely engaged, eager to contribute, and very interested to hear what their peers had to say. The nature and demeanor of their comments suggests they were being very candid and honest about their experiences, observations and perceptions, both positive and negative, and that there was little selection bias in the focus groups.

Data Analysis

The data collected during the focus groups and any additional data passed on to the FTM were analyzed in multiple ways. An initial review was conducted to classify each comment by the primary, and if applicable, secondary topic it referenced. These categories included teacher evaluation, principal evaluation, student learning objectives (SLOs), School Performance Index

(SPI), student growth, MSAs, other assessments, Common Core, training, ratings, rigor, sample instruments, consequences, communication, technology, data management, inter-rater reliability, non-pilot participants, rubrics, fairness, etc. A second review was conducted to classify the comments as predominantly reflecting descriptions, perceptions, or questions. Additional reviews were done to determine what other elements emerged, such as trends, patterns, commonalities, inconsistencies, and idiosyncrasies within and across focus groups and LEAs.

The varied compositions of the focus groups offered two interesting additional data points. Because the focus groups were conducted separately with educators from three different levels within the LEAs (central office staff, school-based administrators, and classroom teachers), the data could be compared across these levels and the descriptions of activities and perceptions analyzed for consistency. For example, when central office administrators described outreach activities being done to ensure that educators were knowledgeable about various evaluation components, participants in the principal and teacher groups might offer feedback that demonstrated the degree to which school level educators were actually aware of such activities.

The levels of familiarity within each of the groups also allowed insight into the probable accuracy of certain perceptions. The focus groups for central office personnel typically included individuals who knew and worked with each other, especially since they had all been part of the field test work. The stories and perspective within these groups were usually fairly similar. Participants in the principal groups, which included principals, assistant or vice principals, and school-based content supervisors or specialists, were generally more familiar with one another in smaller LEAs and less so in larger ones. These participants tended to offer similar views about shared experiences, such as common meetings with central office staff or MSDE events, and more varied descriptions of activities within their own schools. Conversely, most of the people in the teacher groups did not know one another, or only knew a few from working in the same school or attending common LEA activities. For many of these participants who did not regularly work together, this was one of the first chances for them to talk about their involvement in the field test and to hear what others were doing. The stories they shared about the activities in their schools had much greater variety. Some variation would be expected, especially when it related to differences in content area or grade level, while other instances were more surprising. For example, when the teachers within one LEA described the kinds of student measures that could be used for SLOs, it was clear that they each had a very different understanding about what was required and what was considered appropriate, even when they taught similar grades or content areas. These participants also expressed a wide range of views regarding the origin of the differences. Some thought they initiated the state, some at the LEA, and some at the school level.

Through this complex analysis, a picture began to emerge regarding the degree to which each LEA was progressing toward being ready to effectively implement a full teacher and principal evaluation system during the 2013-2014 school year. Much of the data collected are specific to individual LEA activities, policies, and practices, and these are detailed in the individual LEA reports. However, there are several important observations that appear to be applicable in most, if not all, of the LEAs, as well as in a statewide context.

Findings

Field Test Process. Above all, it is clear that stakeholders in all of the LEAs are working very hard to establish teacher and principal evaluation systems that will be ready for full implementation starting in August, 2013. However, the most common request from focus group participants was for an additional no-fault year in 2013-2014 so they would have an opportunity to allow all personnel to experience the new components and implement lessons learned from the field test year, but the majority of participants understood that this was unlikely. During the field test year, they have engaged in numerous activities, ascertained existing and potential challenges, and tried to identify and employ practical solutions to make the evaluation process effective and meaningful. There has been tremendous progress across the state in the creation, roll-out, and initial implementation of the new systems, but there have also been significant developments throughout the year that could impact successful implementation next year. The data collected suggests that the LEAs will be able to report evaluation ratings at the end of next year, but they may not all reflect good evaluation practice. The following sections highlight the other prominent topics that emerged from the focus groups, and the issues, questions, and promising practices associated with them.

The Conversations. The most notable idea that was reported consistently in every focus group was the exceptional value of the conversations between teachers, administrators, and curriculum staff. All of the components of the new systems, including both student growth and professional practice are prompting educators at all levels to engage in more critical conversations about instruction, curriculum, and assessment. Even participants who tended to focus primarily on the challenges of the field test spoke highly of this development. School level educators in particular noted that the more formal emphasis placed on this kind of discourse made their discussions richer, more robust, and centered on more sophisticated elements of instruction than in previous years. Many also indicated that discussions were more data-driven than they had been, although some reported that their LEAs had already been emphasizing data use. Among most of the focus groups, participants suggested that the processes involved in conducting the evaluations, such as in-depth analyses of data, selecting and creating student assessments, developing SLOs, cooperative planning among content and grade levels, conferencing before and after classroom observations, and these critical conversations were already helping them to be more reflective about their instruction and become more thoughtful and better teachers.

Making All Teachers Better. When the most recent national reform efforts switched their emphasis from highly qualified teachers to highly effective teachers, a primary part of the political rhetoric was that more rigorous evaluation systems that included student growth as a significant factor would make it easier to ‘get rid of bad teachers.’ Participants in the focus groups noted several potential issues with relying on this foundational idea as a driving force for

improving educator evaluations practices. For example, until sufficient student measures can be identified that meet an adequate standard for validity and reliability, teachers who are rated as ineffective or even developing in LEAs that include that category are likely to contest ratings that can influence their professional placement, assignment, promotion, compensation, or termination. There are also some models that do not make allowances for exceptional circumstances, such as a teacher who switches grades or content areas at the request of the administration. Educators at hard-to-staff schools or those who work with more challenging student populations are very concerned about the impact of an evaluation system that appears to focus on punitive rather than supportive action, and that teachers and principals may be less willing to take on these assignments. This issue was frequently raised by teachers of students in special education programs and by teachers involved in similar co-teaching situations. At the local level, this concern appeared to be much more prevalent in LEAs in which educators perceived TPE systems as tools for compliance rather than improvement. Responses were far more positive in LEAs that focused on the goal of making all teachers better.

Local Teacher and Principal Evaluation Committees. All of the LEAs had some form of TPE group, including task forces that engaged in intensive, hand-on work developing policies, practices and materials, ad-hoc groups that focused on specific targeted issues, and less formal groups that met primarily to share ideas, questions and concerns. Some LEAs had large committees representing multiple stakeholders, while others relied on existing teams to continue and monitor the field test work. LEAs with more active and engaged committees tended to provide more consistent feedback across the three focus groups and participants seems to have a clearer understanding of the activities taking place and the expectations for the process.

Considerations for Improvement: Throughout the 66 hours of focus group discussions, participants shared a wide range of perspectives reflecting both the benefits and challenges related to the current field test activities and their expectations for full implementation in the future. Some of the issues were specific to individual LEAs and not directly related to the TPE process, but rather could present challenges to any new initiative. This does minimize the impact of these issues, but rather provides a context through which the challenges that are related to the new TPE systems and their unique circumstances can be viewed, For example:

1. LEA size, access to funding, and central office capacity, especially regarding data management systems.
2. The confluence of multiple new initiatives that require extensive investment of resources at all levels, including the transition to the Maryland Common Core Curricula and the anticipation of the new assessments from the Partnership for Assessment of Readiness for College and Careers (PARCC),
3. The degree to which the LEA is developing and/or implementing a new system, and how well this system aligns with previous versions.

4. The role played by local bargaining units.
5. The existence of locally developed common assessments.
6. Local turnover at the central office and school administrator level.

In addition to these ongoing concerns, many of the participating LEAs also encountered challenges that were tied more directly to the new TPE systems.

1. The opportunity to participate in the initial TPE pilot during the 2011-2012 pilot year and the LEA's capacity to effectively use that time. It should also be noted that MSDE was undergoing significant changes in leadership while the initial pilot was being conducted, including the state superintendent.
2. The evolving research base related to expanding educator evaluation and the inclusion of student growth as a factor in evaluation ratings.
3. The complexity of models for calculating teacher and principal ratings. Several LEAs expressed concerns about being able to effectively articulate the intricacies of these models to their stakeholders, especially as they were often different for different positions.
4. The demand for professional development materials and the limited capacity of LEA central offices to devote time and resources to create them, and experienced facilitators to conduct such trainings.
5. The wide range of approaches to the roles principals, assistant or vice principals, and curriculum supervisors or specialists at both the central office and school level played in teacher evaluation.
6. Periodic changes, or perceived changes, in deadlines for submitting data to the state.
7. An apparent misalignment between the availability of student assessment scores and the schedule for making personnel decisions that are dependent on these scores. This prompted some LEAs to rely on a mixture of professional practice data from the current school year and the results of student scores from the previous school year.
8. Data management technology that is not currently designed for the kind of access and analysis necessitated by the new TPE systems. Several LEAs have been working with Performance Matters, but it was sometimes difficult to get the necessary attention for their LEAs' specific needs.
9. The perception that the new TPE systems will require significantly more time on the part of teachers, principals, and central office personnel, and the complexities presented by adjusting schedules and responsibilities to accommodate those changes.
10. The emphasis on teacher evaluation rather than principal evaluation. As it also did in many other states, this led some LEAs to devote the bulk of their resources to developing and implementing teacher evaluation, while far less attention was paid to principal evaluation, and some of those models are not yet complete.

11. The reluctance of some LEAs to make certain local policy decisions without specific, official guidance from MSDE. This means that some LEAs have not addressed certain issues, such as the impact of student attendance on SLO attribution and how ratings will be assigned to teachers with very small student sample sizes.
12. LEA stakeholders who were not included in the field test process. There will be many stakeholders engaging in the full TPE systems for the first time next year when official ratings are assigned. Principals and specialists were particularly concerned about assigning ratings to teachers while they are still learning the system.
13. The limited time frame of the field test year. This will not allow all of the LEAs to adequately analyze data and lessons learned from this year such that they can make adjustments before the 2013-2014 school year.

The following examples illustrate some of the challenges encountered by LEA in greater details.

Teacher Professional Practice Evaluation. Some LEAs were field testing systems that closely aligned with evaluation practices that were already in place, while others were making dramatic changes that represented tremendous paradigm shifts for their educators. Many LEAs had already been using systems that were based on Charlotte Danielson's *Framework for Teaching* or similar models, so their transitions to their new models were less stressful. However, even these LEAs had to engage in additional work to ensure all participants had sufficient training and to address issues related to more intensive implementation practices. For example, several had been using rubrics based on their model for their classroom observations, but had not been holding the in-depth pre-observation meetings. These meetings added time to the process for both the teacher and evaluator, and also presented scheduling challenges, as it became more critical that the evaluator observe the specific lesson discussed in the meeting. Many educators were also challenged by the change in the final ratings, as most LEAs had been using ratings that were essentially satisfactory or unsatisfactory, and the new ratings offer a greater range in performance. Of particular concern were the indicators for the highly effective rating in the Danielson-type models, because they tend to focus on student behavior, over which teachers feel they have less control. Early childhood teachers also noted that many of these student behaviors are not developmentally appropriate for younger students and that some of their curricula are scripted and do not allow for lessons that support these behaviors. Many teachers felt it was unreasonable to include a rating that was not attainable within certain contexts. LEAs are also working to increase inter-rater reliability to ensure that the process is as fair as possible within and across schools.

Teacher and Principal Student Learning Objectives. While there were some LEAs that had been using goals related to student measures in their evaluations already, and virtually all educators develop and assess student objectives in some manner, SLOs in the format presented by MSDE were new to most participants. Both teachers and principals reported frustration and

anxiety related to developing, reviewing, and scoring SLOs for their evaluations. A major concern across the state, even in LEAs that had been using SLOs already, was establishing practices that ensure consistent rigor and equity in the setting of targets and scoring of SLOs.

Some of challenges related to a lack of guidance on the process, driven by the kind of training participants had, and there was tremendous variety in the training being developed and presented throughout the state. Some LEAs relied primarily on MSDE for both actual training and training materials, which had both advantages and disadvantages. It did reduce the amount of time, funding, and staff resources LEAs had to invest in the development and presentation, but it also made the LEAs dependent on MSDE's schedule and availability. All of the LEAs offered very positive feedback on MSDE's training related to SLOs, and appreciated the degree to which the team tried to accommodate their needs and requests. Unfortunately, it was impossible for the MSDE team to provide the individualized, intensive training that the LEAs needed early enough in the year so that all participants had adequate experience before writing and reviewing their SLOs. As the SLO process for evaluation is continuing to evolve, both in the national context and within Maryland's system, it also meant that the requirements, guidelines, and parameters for SLOs changed somewhat during the field test year, which was frustrating for many participants. Other LEAs developed their own process, materials, and training, which brought its own challenges regarding scheduling and resources.

Recommendations

While some of the challenges reported by the LEAs are primarily related to the field test year and initial implementation, there are some issues that will require ongoing attention, and there are a variety of ways MSDE can provide support. These recommendations reflect both logistical and substantive issues with which most of the LEAs are struggling.

1. Establish clear communication plans for the TPE process that emphasizes common and consistent messages, and include opportunities for stakeholder feedback. This is a critical issue to ensure fair and credible practices throughout the state.
2. Identify and support additional opportunities to increase LEA and school administrator capacity to act as instructional leaders.
3. Support LEAs as they prioritize their work and make appropriate connections between the TPE systems and other work they are already doing.
4. Clarify what decisions will be made by MSDE and which decisions should be made at the local level.
5. Emphasize the opportunities the TPE process offers to improve instructional practice and student learning, especially exemplified by the more in-depth conversations among teachers, administrators, and content specialists.
6. Ensure the data systems at the central office, school, and classroom level can facilitate collection, analysis, retrieval, and reporting of the necessary data and are compatible with MSDE's systems.

7. Facilitate opportunities for LEAs to collaborate with other LEAs to address emerging issues.

Summary

As previously noted, educational stakeholder throughout the state have invested an exceptional amount of time, energy, and resources to get the most out of this field test year and prepare their personnel for a full, formal implementation of their local teacher and principal evaluation systems. While there is no doubt that TPE participants would benefit from an additional no-fault year to ensure all of their personnel have adequate training and experience, and to fully address other logistical issues, the majority of the LEAs will be able to implement the components of their TPE systems by August 2013. However, while a few LEAs have solid systems in place, the degree to which the other systems will promote practices that actually impact classroom instruction and student learning during this first year is questionable. LEAs may also have to struggle to obtain and maintain buy-in from their stakeholders if the first year does not proceed smoothly. MSDE can expect all of the 22 RTTT LEAs to report evaluation ratings for their teachers and principals at the end of next year, but they may not represent meaningful reflection and professional growth across the board.