Putting the Pieces in Place: Charlotte-Mecklenburg Public Schools' Teacher Evaluation System

By Rachel Curtis

March 2012

THE ASPEN INSTITUTE
EDUCATION & SOCIETY PROGRAM

Putting the Pieces in Place: Charlotte-Mecklenburg Public Schools' Teacher Evaluation System

By Rachel Curtis

March 2012

Acknowledgments

About the Author

Rachel Curtis works with school systems, foundations, and education policy organizations on teacher and principal human capital issues. In 2006, as assistant superintendent of the Boston Public Schools, she developed the system's teaching standards and aligned new teacher induction support and teacher evaluation to them. Her publications include the books *Teaching Talent, Strategy in Action*, and *The Skillful Leader II*.

Overview

The first two goals of *Teaching Our Way to the Top*, the *Charlotte-Mecklenburg Public Schools* (CMS) Strategic Plan 2014, – Effective Teaching and Leadership and Performance Management – make clear that the district's top two priorities are ensuring excellent instruction district-wide and creating a performance culture organized on the principles of continuous improvement and accountability for results. Critical to reaching these goals is the commitment to evaluate and compensate all teachers in the system based on their performance as measured by classroom observations, student test scores, and additional metrics. This work is part of a larger effort to ensure equity and excellence for all students by shifting the district's focus from managing inputs – the amount of money and services provided to schools – to focusing on equity of outcomes: student achievement and graduation rates.

CMS is developing a system that will measure teacher performance and linking compensation to it by 2014. The district began to develop its own value-added formula in 2009, with the goal of ultimately being able to provide a value-added score for every teacher. In 2010, CMS adopted the new state teacher evaluation standards, augmenting them with indicators aligned to the district's priorities and beginning to use it to assess teachers' classroom practice. The system tackled the issue of developing value-added scores for teachers whose classes are not included in the state assessments by piloting summative assessments in non-tested grades and subjects in the spring of 2011. This work on summative assessment will be expanded in 2012, working in collaboration with the North Carolina Department of Public Instruction and its Race to the Top plan. The district is also exploring other metrics it will use in measuring teacher effectiveness, such as student surveys, leveraging what it is learning through implementation of its Teacher Incentive Fund grant and as a partner district in the Bill & Melinda Gates Foundation's *Measuring Effective Teaching* (MET) study.

In the midst of defining teacher effectiveness and how to measure it, CMS is acting on the informa-tion it is generating along the way. For example, the system developed its *Strategic Staffing Initiative* (SSI) in 2009 to create incentives for principals and teachers who are identified as highly effective, through their value-added data, to work in the district's highest-needs schools. And in each of the years since the spring of 2009, the district has relied on teacher-performance data to determine which teachers to lay off, after the economic crisis forced reductions in the workforce. Throughout all its efforts, CMS has sought to take time to learn from early efforts and refine the overall strategy based on these learnings, while also using the best data available to guide important decision-making.

The district's commitment to both learn from and act on its early teacher effectiveness efforts is guided by a district-wide commitment to continuous improvement in all of its work, codified in its *Cycle of Continuous Improvement* (See Appendix A). To ensure effective implementation, CMS has developed systems and structures to engage a broad group of stakeholders, track implementation, and surface critical learning to inform improvement. This orientation makes system leaders anticipate that the design and application of teacher effectiveness measures will evolve in the coming years. The 2014 timeline and the steady march toward it offer the system a chance to build broad and deep understanding and buy-in and ensure the integrity of the evaluation system before aligning teacher compensation to it.

DISTRICT FACTS	
SCHOOLS	189
TEACHERS	8,890
STUDENTS	138,012
STUDENT DEMOGRAPHICS	
ASIAN	5%
BLACK	42%
HISPANIC	17%
MULTI-/OTHER	3%
WHITE	33%
ELIGIBLE FOR FREE OR REDUCED PRICE LUNCH	54%

Table of Contents

Background

From 2006 to 2011, CMS had steady leadership and demonstrated success in addressing equity issues. Peter Gorman, the superintendent during this time, garnered strong support for fundamental reforms aimed at ensuring equity in a system that has both a strong white, middle-class base and a very culturally and socio-economically diverse community. Ann Clark, the Chief Academic Officer, has a storied 28-year history in the district, serving in a variety of leadership roles and was a former National Principal of the Year. During Gorman's tenure, the academic performance of black and Hispanic students outpaced that of their counterparts across the state in reading and math at all school levels, and achievement gaps narrowed, in some cases substantially. The achievement gap narrowed between black and white students in reading and math at all levels, while the gap between Hispanic and white students' performance narrowed in math at all levels and in middle and high school reading.[1] This progress has garnered CMS national attention and contributed to their winning the Broad Prize in 2011.

In 2005, the state of North Carolina set the stage for the district's efforts to measure teacher performance and use the data in compensation decisions. That year, the state became one of the first to develop a simple growth metric based on the state assessments given in grades 3 through 8 in literacy and math. In 2007, CMS won a Teacher Incentive Fund (TIF) grant to explore pay-for-performance in twenty of its high-poverty, low-performing schools. In the first year of the grant (school year 2007-08), teachers were able to earn bonuses based on simple growth data, provided by the state. Teachers who demonstrated "high growth" as defined by

the state, earned 15% bonuses while teachers who earned "better than expected growth" earned 10% bonuses. Teachers in tested grades and subjects were the only people eligible for this bonus.

The district continued the program, but the student performance measures used to calculate bonuses evolved. The state moved from a simple growth model to a value-added model. The consultant North Carolina partnered with to develop its model would not share his formula and methods so the calculation could not be made transparent to teachers. Knowing that this metric would eventually factor significantly into teachers' performance rating and impact their compensation, CMS was uneasy about relying on calculations it could not explain. In response to this, CMS developed its own value-added metric in 2009-10 using a formula it could explain to teachers. The CMS metric served as the basis for future pay-for-performance payouts.

In 2008, CMS expanded the pay-for-performance experiment to include teachers in non-tested grades and subjects. For those teachers, the district introduced student learning objectives (SLOs). Teachers could set up to three SLOs and earn $1,400 for each goal met, for a total possible bonus of $4,200. Teachers in tested grades and subjects could set up to two goals to supplement their potential value-added bonus.

Over the last three years, CMS has worked to strengthen the rigor of SLOs and to use them to leverage teacher collaboration and align teachers' work to school goals. It has been important to ensure that teachers are setting goals focused on appropriate "power standards" from the expected course of study and setting high, "stretch" goals. To

---

[1]   The Broad Prize Charlotte-Mecklenburg Schools, NC Profile, http://www.broadprize.org/asset/1579-tbp%202010%20charlotte%20fact%20sheet.pdf

that end, the district developed a template teachers are expected to follow, which guides teachers through the process of choosing two to six objectives from the standards to focus on and defining detailed growth goals. Teachers can develop whole-class or targeted groups SLOs depending on their students' needs. These SLOs can also be part of a team SLO where teachers collaborate to generate common goals and assessments because grade/class/student data identifies common needs.

Developing, implementing and refining strategies to achieve the goals then becomes the focus of the Professional Learning Communities (PLC) work. Where this is being done most successfully (i.e., with the greatest gains in student achievement), SLO goals clearly align with PLC goals, which strategically support goals of the School Improvement Plan.

After several years of refining the SLO structure, CMS developed two additional ways teachers in non-tested grades and subjects could earn additional compensation. A school value-added measure was introduced that provided a bonus to all teachers in a school that demonstrated student growth in the top 40 percent of district schools. Additionally, teachers in non-tested grades and subjects who felt they made a significant contribution to student learning in a tested grade or subject could "tie into" that test to gain recognition for student growth. For example, a middle school social studies teacher who focuses on developing her students' content-area literacy would have the opportunity to tie into her students' English language arts test scores.

All of these efforts to develop performance measures for teachers reflect CMS's focus on performance management, one of the goals of Strategic Plan 2014. Then-superintendent Gorman took the lead on this work when he negotiated a performance-based contract for himself with the school board. Part of his compensation was tied to student achievement and the graduation rate. This set the stage for everyone in the system to eventually have performance-based metrics that influence compensation and that are tied to both their individual job responsibilities and students' achievement and attainment. (The superintendent requires every employee in the system to have a goal tied to raising the graduation rate, one of the priorities in Strategic Plan 2014.)

The district originally proposed to expand the system of tying compensation for all employees to student performance over time. Beginning with the superintendent in 2010-11, the system was expected to expand to the district's executive team members and senior managers the following year, to principals and all other district staff in 2012-13, and, eventually, to teachers. Gorman wanted teachers and principals to see his commitment to holding central office employees accountable for outcomes before pursuing a performance-based, outcome-oriented approach with front-line educators.

However, a state law enacted to enable North Carolina to qualify for the federal Race to the Top competition overrode Gorman's intentions. Under the law, the state introduced a new principal evaluation system in school year 2009-10, followed by the new teacher evaluation in school year 2010-11 and a new central office evaluation in 2011-12. Additionally, in August 2011, the North Carolina Board of Education voted to add a sixth standard to its teacher evaluation focused on measuring teachers' contributions to the academic success of students which requires three years of student performance data. This will have implications, which are still being determined, for CMS's evaluation and its plans to tie compensation to student performance.

Evaluation

Value-Added

To develop value-added data, CMS partnered with the Center for Educational Policy Research at Harvard University to recruit Strategic Data Fellows, who work in CMS full time as part of a two-year fellowship. These fellows became the talent effectiveness work and produced CMS's first value-added measure in the summer of 2009, based on 2008 and 2009 state assessment data. This value-added score was calculated for teachers in tested grades and subjects who worked in the twenty high-poverty, low-performing schools participating in the TIF grant. Having established the value-added model, CMS then calculated value-added for previous years for these teachers, going as far back as 2002 to look for patterns.

The early analyses showed little correlation between value-added and principals' evaluation ratings: the correlation between value-added scores and principals' ratings of teachers was 0.1 to 0.3,[2] depending on the grade and subject level. Teachers regularly received the highest possible evaluation score while their students were achieving less than a year's worth of growth in their classroom. And when the district examined value-added and compensation, it found a correlation of less than 0.01, which meant there was virtually no relationship between teachers' performance, as measured by student achievement growth, and how much teachers were paid.

These early analyses convinced the CMS leadership that they would need to strengthen principals' skills of observing and analyzing instruction. They also suggested that in case there wasn't a high correlation between observation and value-added scores

(after evaluators were well trained) it would be important to explore a variety of measures in an effort to develop the most holistic assessment of teachers' practice and not rely too heavily on value-added.

Observations

In 2009, a year before North Carolina won the federal Race to the Top competition, it introduced new standards for teacher practice and an evaluation process that every school system would be expected to use. This expectation is in keeping with North Carolina practice, in which the state plays a powerful central role, rather than allow district discretion in education policy. (The state also sets a formula that guides the allocation of teaching positions and implements a statewide salary schedule.[3])

In January 2010, just after the state released its new teaching standards, CMS began to focus squarely on strengthening principals' capacity to evaluate teacher performance. This work began with an exercise at a monthly principals' meeting in which the Chief Academic Officer (CAO) asked principals to sort their teachers into four performance quartiles, placing 25 percent of their teaching force in each quartile. The CAO then took those rankings and mapped them to the teachers' previous evaluation ratings and student growth data, where available.

The findings highlighted several problems. Principals placed teachers they had previously evaluated as meeting standards or better in all four quartiles, showing a serious misalignment between what principals thought about their teachers' performance and how they formally rated them. Additionally, 97

---

[2] In social sciences a correlation of .3-.5 is considered moderate while 1.0 is a perfect correlation and 0.0 demonstrates no correlation.

[3] The state salary schedule can be supplemented with local revenues, which CMS does.

percent of the teachers had evaluation ratings of meets or exceeds standards, while the high school graduation rate is 69.9 percent, suggesting that principals thought teachers were doing a great job while student performance languished. In a moment everyone in attendance at the meeting recalls quite clearly, the superintendent expressed his disbelief that a system could rate almost all of its teachers as meeting standards when so many of the students they are responsible for educating weren't meeting theirs. This shared experience set the stage for the district to raise expectations for teacher performance and to introduce a variety of measures to assess effectiveness.

NORTH CAROLINA'S TEACHING STANDARDS

1. Teachers demonstrate leadership
2. Teachers establish a respectful environment for a diverse population of students
3. Teachers know the content they teach
4. Teachers facilitate learning for their students
5. Teachers reflect on their practice

CMS then needed to decide how it would implement the evaluation system North Carolina had introduced. The system included five standards (see sidebar) and twenty-five sub-standards. (See Appendix B for the list of standards and sub-standards.)

The new system has a five-point rating scale as compared to the four-point scale of the old evaluation. The terminology of the ratings is different in other significant ways. The old four-point ratings were unsatisfactory, below standard, at standard and above standard. Under the new scale's ratings – not demonstrated, developing, proficient, accomplished, and distinguished. In the new system there are two ratings above proficiency (as compared to one in the old system). These ratings emphasize growth and development beyond proficiency and allow for evaluators to make distinctions among high performers.

However, the single rating below proficient – "developing" – fails to distinguish between persistently poor performers and teachers who are, in fact, developing (particularly important for teachers in their first and second years of teaching). The language suggests an assumption that most teachers are performing at or above a basic level of performance. The consequences associated with the rating clarify this distinction, though. Under the state policy, non-tenured teachers must be performing at the level of proficient in all standards at the end of their third year of teaching to pursue tenure (awarded at the end of the fourth year of teaching). Tenured teachers who receive a rating of "developing" in any standard go into corrective action which requires additional professional development to address areas of deficiency. North Carolina has given school systems the authority to determine how long they will support teachers rated as "developing," and the level of improvement required to maintain employment.

Ultimately, the effectiveness of the new state rating system will be determined by the alignment of teacher ratings to student performance. If student performance suggests lots of room for improvement, so too should teacher ratings.

Along with the new four-point rating scale, the state provided a rubric, which includes descriptors of what performance on each sub-standard looks like at the four levels of performance to support observers in their assessment of teaching practice. CMS decided to customize the state-developed rubric by creating additional indicators for each standard and sub-standard. By making the state descriptors more robust, CMS gave teachers and principals more detail about what the standards actually look like in the classroom. This gave teachers a richer foundation on which to reflect and refine their practice and supported evaluators in making their assessment of teachers' instruction.

At the same time, the indicators helped align the evaluation to district priorities. Examples of priorities the district had been working on that CMS wanted to explicitly call out in its indicators include: the use of data to drive instruction; teacher collaboration and professional learning communities; differentiation in instruction; and closing the achievement gap. (See Appendix C for an additional excerpt from the NC rubric that illustrates CMS's adaptations.)

Standard II: Teachers establish a respectful environment for a diverse population of students.

A. Teachers provide an environment in which each child has a positive, nurturing relationship with caring adults. Teachers encourage an environment that is inviting, respectful, supportive, inclusive, and flexible.

OBSERVATIONS	DEVELOPING	PROFICIENT	ACCOMPLISHED	DISTINGUISHED	NOT DEMONSTRATED (Comment Required)
STATE DESCRIPTORS	☐ Appreciates and understands the need to establish nurturing relationships.	☐ ...and Establishes an inviting, respectful, inclusive, flexible, and supportive learning environment.	☐ ...and Maintains a positive and nurturing learning environment.	☐ ...and Encourages and advises others to provide a nurturing and positive learning environment for all students.	
POSSIBLE CMS INDICATORS	▪ Teacher is working to establish classroom rules and procedures to facilitate an orderly and nurturing learning classroom environment. ▪ Teacher/student interactions are respectful, positive and appropriate. ▪ Teacher's strategies for affirming positive student behavior are observed in student/teacher interactions.	▪ Classroom rules and procedures are well established that facilitate an orderly learning environment. ▪ Discipline concerns and issues are handled promptly, appropriately and effectively. ▪ Students are supported and protected from harm, abuse, bullying, and neglect. ▪ Instructional time is maximized for learning.	▪ Student/student relationships are consistently positive. ▪ Classroom rules and procedures have become internalized. ▪ Classroom is student-centered, interactive and highly engaged. ▪ Instructional time is maximized for learning.	▪ Students are encouraged to take leadership roles in the classroom. ▪ Students and teachers work together as a classroom "PLC", creating positive synergy.	

The above excerpt from the rubric illustrates the state's descriptors and CMS indicators for one sub-standard.

In the sample rubric, the CMS indicators highlight the system's commitment to: clearly established classroom rules and procedures; classrooms that are student-centered, interactive and highly engaging; student leadership and modeling the PLC concept at the classroom level.

CMS recruited two retired principals with extensive training and demonstrated expertise in classroom observations and instructional analysis to write the indicators. By the spring of 2010, CMS had finalized its indicators and was ready to prepare teachers and principals to implement the new evaluation in the 2010-11 school year.

Putting The Pieces In Place

Observation Training

In the spring and summer of 2010, CMS focused on training all teachers and principals on the new standards and evaluation process. The district identified fifty of its most skilled observers and evaluators of teacher practice and sent them to a three-day state training on the standards and evaluation provided by the North Carolina Department of Public Instruction. The state training provided participants with a script and slide presentation, in an effort to ensure consistency in the message being communicated to teachers. The people CMS sent to the state training returned to the district and led a similar three-day training to introduce their teacher and administrator colleagues to the new standards and evaluation process. Teachers and principals received ten hours of introductory training in the standards by June 2010 and were required to participate in ten additional hours in August before administrators began using the new evaluation instrument in the fall of 2010.

CMS leaders knew that school administrators would need ongoing support to effectively implement the new evaluation. In the context of a third year of budget cuts, this was hard to provide. The CAO found herself in the situation of having to choose between hiring a person to oversee all aspects of implementing the new evaluation and having someone available to train principals in the use of the new tool. She chose the latter, hiring the two retired principals who had developed the CMS rubric-indicators to work part-time to design and lead the ongoing training of principals and assistant principals. This meant that the CAO, who is responsible for the entire academic program of the district and the indirect supervision of all 178 principals, had to serve as the coordinator and champion of implementing the new teacher evaluation system.

Beyond the initial three-day training, principals received additional training for an hour at each monthly principal professional development day. The focus was on building principals' skills of scripting observations and note-taking in an effort to build a common, best practice. Through these sessions, principals' depth of knowledge about instruction was perceived by the two retired principals charged with developing principals' skills to be "all over the place," reflecting the principals' historic emphasis on management, rather than instruction, and the fact that 120 principals had been hired in the last five years.

The state training focused on the process of how to use evidence from classroom observations to assess teacher performance on the standards. This approach assumed that school administrators understood the standards and knew what evidence to look for relative to each standard. This proved not to be universally true across all standards but, most particularly, for the two standards – *Teachers demonstrate leadership* and *Teachers reflect on their practice* – that were entirely new to CMS teachers and the administrators who evaluate them.

The challenge became how to find the time to provide school administrators the training they needed and how to ensure all principals had equal access to the training. This issue of equal access and treatment was particularly important for CMS to be able to demonstrate at the end of the year. It anticipated needing to make an additional teacher Reductions in Force (RIF) based on the budget and intended to use administrators' ratings of teachers using the new evaluation as the basis for these dismissals. In case of challenges made to the RIF process, CMS needed to be able to demonstrate that all administrators received the same training in the evaluation process.

Some principals received additional support through their Professional Learning Communities (PLC). A number of the teams used the implementation of the new evaluation system as a starting point for their work. Some PLCs chose to deeply explore one of the standards from the evaluation. Others discussed issues that emerged as needs, such as the issue of rigor

New principals received additional training and development through an induction program. Experienced principals were recruited to work with first- and second-year principals as "consulting coaches." These coaches led monthly professional development for the new principals and followed it up with school visits to provide individualized support. The trainers responsible for all principal professional development on the new evaluation developed the curriculum for these monthly meetings, which focused on implementing the new evaluation instrument.

Observation Implementation

The state requirements of the new evaluation system are differentiated based on teachers' experience level. "Non-career" (i.e., non-tenured) teachers are observed three times a year, culminating in a formal evaluation. "Career" teachers are observed at least once a year and receive a full evaluation, which includes three observations and a formal write-up, once every three years.

Principals are expected to provide written feedback for each observation and conduct post-observation conferences. Early indications suggested that principals were overwhelmed by the time required for the post-observation conferences. In the first round of observations, they reported that conferences were 45 to 90 minutes long, with much of the time spent on building teachers' understanding of the standards. In addition to the observations and the related conferences, principals are responsible for completing a written summative evaluation rating in which they include and synthesize data sources beyond the observation, such as lesson plans, student assessments, and walk-through data.

The district's Human Resources department reviewed principals' work to ensure they followed procedures – e.g., signed the evaluation, followed the timeline, made comments in standard areas rated as needing improvement. The findings from this review informed a mandatory, three-hour training all evaluators were required to participate in during the spring of 2011, which outlined the expectations for end-of-year summative evaluation write-ups. While principals received feedback on their implementation of the evaluation process, they received little feedback on the quality of their analysis of instruction and feedback provided to teachers.

Non-Tested Grades and Subjects

While the principals were implementing the new observation system, CMS was working to determine how it would measure student learning in non-tested grades and subjects. Early in the TIF-funded work, CMS considered student learning objectives (SLO) as a strategy to measure value-added in non-tested grades and subjects. As CMS piloted and refined its SLO work, district leaders became increasingly convinced that SLOs were a practice they wanted to make a way of doing business in the system. For that reason, the district was hesitant to scale-up the TIF-funded strategy of providing bonuses for teachers for achieving their SLOs. It didn't want to set the precedent of paying for something it expected as a core practice of effective teaching.

At the same time, the district was developing end-of-year and end-of course assessments for non-tested grades and subjects. Those tests could be factored into a value-added measure for courses and grades that don't administer the state assessment or Advanced Placement or International Baccalaureate exams.

CMS hired an external vendor, Measurement Inc., to help develop the end-of-semester/course test questions. In the spring of 2011, CMS piloted seventy-five tests, focusing specifically on Social Studies (K-8), Science (K-4 and 6-7) Math (K-2), and English Language Arts (K-2). As part of the test development, CMS created adapted assessments for students with special learning needs. After making refinements to the tests based on the learning from the pilot and defining cut points, the system plans to administer these assessments district-wide in spring 2012.[4]

---

[4] In February 2012, CMS agreed to use state-developed summative assessments in place of the assessments the district developed. This agreement allowed CMS to conserve resources and avoid duplication, given the state's commitment to develop a battery of end-of-course tests. While CMS will use the state assessments in most subject areas, it will continue to use the assessments it created in fine and performing arts, physical education and world languages.

In the fall of 2011, teacher design teams in the fine and performing arts and physical education (e.g. art, chorus, band, physical education, drama) collaborated with Discovery Education to develop authentic assessments of student performance and work. Discovery Education trained teacher teams on meaningful performance assessment and the structure of rubrics for scoring the performances. As part of the partnership, Discovery Education developed a distributed scoring tool (to CMS's specifications) that allows teachers to design their own assessments and the district to provide structured performance assessments. Ultimately, student performances will be uploaded into an electronic system and distributed to teachers for scoring, ensuring more than one teacher scores each performance to ensure validation. The assessments are being designed to produce enough variability in score that a growth model from year to year in continued study in the content area should be able to be produced.

With a team of three people managing this work internally, CMS has a schedule for rolling out all of the tests for non-tested grades and subjects. Once the tests have been fully refined and implemented CMS's value-added team will determine how they will fit into the CMS value-added model.

Initially, many teachers reported anxiety about these new assessments. Some consider the tests a constraint on their practice, because they feel compelled to teach what is tested. For teachers in the early grades (K-2), these tests will mark the first time their students are being assessed so formally. Teachers are concerned about the time it takes to administer the assessments and are feeling the pressure of accountability. The system also encountered pushback from parents who are concerned about too much time being spent assessing students and how those assessments are being used to inform high-stakes decisions for students and teachers. However, the district considers the tests essential for equity, because they set, for the first time, common expectations for curriculum and student performance across all schools in Charlotte-Mecklenburg.

Other Measures

The combination of the work of the TIF grant and CMS's participation in the Gates Foundation-sponsored *Measuring Effective Teaching* (MET) study immersed the system in an exploration of multiple measures of teacher effectiveness. Through TIF, CMS piloted value-added and SLOs. The MET study focuses on exploring different ways of measuring teacher effectiveness and determining which measures have the greatest correlation to value-added. Four hundred CMS teachers were involved in the study, which included videotaping, evaluation of their content knowledge, and colleague and student surveys. CMS anticipates it will receive data from these analyses and benefit from the cross-site review of various measures of effectiveness. While the MET study emphasizes alignment of different measures to value-added, CMS is trying to also align different measures to one another to see what it can learn and to avoid setting up value-added as the gold standard when it has not yet proven itself.

Given the dynamic nature of the work on multiple measures, in 2010 CMS decided to engage teachers in researching a variety of measures – many of which are being explored through TIF and/or MET – to get their best thinking about how the system might pursue each. Teacher Working Teams (TWTs), made up of teachers from across the district, explored a variety of measures including:

1. Value-Added Measures (school, team, individual)

2. Professional Learning Community (PLC)

3. Student Learning Objectives

4. Student Survey

5. Teacher Observations

6. Hard to Staff Schools and Subjects

7. Teacher Work Products

8. Content Pedagogy

Their research into each of these measures will lead to a series of pilots, which will inform the ultimate choice and weighting of the measures CMS will include in its overall effectiveness metric. The goal is to define, pilot, and assess all possible measures by June 2013, so that they can be fully implemented and integrated into teacher evaluation ratings in school year 2013-14.

It is not assumed that all of the measures TWTs are exploring will be included in the calculation of teacher effectiveness. The intention is to explore a variety of ways to both support teacher development and assess their performance and figure out which ones have the most potential.

Year One Data And Results

After implementing the new observation tool for a full year, CMS has a variety of data that offer both general and nuanced information about teacher performance. The following chart illustrates teacher evaluation ratings at the end of the 2010-11 school year by each standard and each of the four performance levels.[5]

CMS provided teachers their evaluation results by standard rather than by a single aggregate rating. This was done in recognition of the fact that CMS is moving towards a more robust evaluation which will ultimately include a variety of other measures. Which measures will be included and how they will be weighted has not yet been determine. As CMS phases in the elements of the new evaluation it is focused on providing teachers transparent data about their performance. While teachers are not getting a single rating, the state has established

accountability for teacher performance by requiring that any teacher who receives a rating of "developing" in one or more standard be placed in corrective action, a status which focuses on requiring teachers to pursue professional development and other support to address identified areas of weakness.

While it is not an apples-to-apples correlation, it is instructive to compare the 2010 summative teacher ratings distribution and the 2011 ratings distribution. There are several things worth noting about these data. First, the ratings for teachers appeared to have dropped between 2010 and 2011. In 2010, more than 90 percent of the teachers were rated in the top two performance levels. The following year, more than 90 percent of the teachers were rated in the middle two performance levels. It is hard to discern if this reflects a raising of the bar because the titles of the new four categories have changed; as noted

STANDARD	NOT DEMONSTRATED	DEVELOPING	PROFICIENT	ACCOMPLISHED	DISTINGUISHED
1. Teachers demonstrate leadership	0.09%	5%	59%	33%	3%
2. Teachers establish a respectful environment for a diverse population of students	0.32%	5%	54%	38%	3%
3. Teachers know the content they teach	0.07%	5%	65%	28%	2%
4. Teachers facilitate learning for their students	0.27%	6%	66%	26%	1%
5. Teachers reflect on their practice	0.18%	6%	61%	31%	2%

---

[5] Because CMS teachers are on a bi-annual evaluation cycle, these data reflect evaluation ratings for half of the system's teachers and those teachers who are on an annual evaluation cycle due to performance concerns.

| SCHOOL YEAR 2010 | |
| PERFORMANCE RATINGS USING OLD STANDARDS AND OBSERVATION TOOL | |
TEACHER PERFORMANCE LEVEL	% OF TEACHERS
ABOVE STANDARDS	40%
AT STANDARDS	56%
BELOW STANDARDS	4%
UNSATISFACTORY	0%

above, Level 2 is now "proficient." It appears that the shift does reflect a clear focus on growth and development and a desire to give teachers a performance rating they can stretch towards. At the same time the percent of teachers scoring at the lowest level has risen from 0 to 5 - 6 percent, which represents a dramatic increase. Conversely, the percent of teachers performing at the highest level has dropped from 40 percent to 3 percent – a substantial decline, suggesting that principals are reserving the "distinguished" rating for the most exceptional teachers.

Digging into the data at the level of the five teaching standards surfaces some important differences in performance. Teachers are more likely to be rated "proficient" – and less likely to be rated "accomplished" – on Standards 3 and 4, the two standards that most relate to the content knowledge and delivery of instruction, than on the standards related to teacher leadership and classroom environment. The ratings for the ability of teachers to reflect on their practice sit between these two groupings. These data suggest that content knowledge and the delivery of instruction are the areas where teachers need the most support.

The next level of analysis looks at sub-standards in each standard area where performance ratings are significantly above or below the other related sub-standards. A number of things stand out in CMS's data:

- In Standard 1, 71% of teachers were rated level 1 or 2 in sub-standard: Teachers advocate for schools and students

- In Standard 2, 55% of teachers were rated level 3 for the sub-standard: Teacher establish a respectful environment for a diverse population of students

- In Standard 2, 71% of teachers were rated level 1 or 2 in sub-standard: Teachers embrace diversity in the school community and in the world

- In Standard 4, 76% of teachers were rated level 1 or 2 in sub-standard: Teachers help students develop critical-thinking and problem-solving skills

- In Standard 4, 78% of teachers were rated as level 1 or 2 in sub-standard: Teachers help students work in teams and develop leadership qualities

- In Standard 4, 76% of teachers were rated level 1 or 2 in sub-standard: Teachers use a variety of methods to assess what each student has learned

The first three findings taken together provoke questions for further inquiry. On the one hand, teachers are getting high marks for creating a respectful environment for a diverse group of students. On the other hand, they're being rated lower for their advocacy for students and the extent to which they embrace diversity. Understanding how evaluators see the relationship between these three factors and the differences in scores would likely provide the system insight into how to better support teachers in these areas, how to continue the process of calibrating evaluators' scores, and how to best pursue its goal of closing the achievement gap.

The last three findings relate specifically to instruction and suggest areas where teachers need additional support. Looking across the three findings raises questions about how each is related to the other. For example, to what extent do the assessments teachers use require students to think critically, solve problems, and work collaboratively? The work of the system now is to inquire about the root causes of these issues to inform its response.

In the spring of 2011, then-superintendent Gorman reflected on the evaluation system and its usefulness in improving teacher practice and making it easier to remove ineffective teachers. At the end of the first year of implementation he had fundamental questions about the extent to which the same evaluation instrument can serve multiple, essential roles: supporting novice teachers as they develop their skills, appropriately addressing more experi-

enced teachers as they continually improve their practice, and dealing with persistently poor performers who may need to exit the system. There is a sense that the tool may be most helpful for teachers who perform in the middle two ratings.

While new teachers are observed more frequently, it is quite likely that they will perform at the developing level in many areas. Because the standards are not prioritized or considered along a developmental continuum and the rubric is not differentiated to address the particular needs of novice teachers, there is concern about how well this tool can be used to guide and support new teacher development.

Principals have raised similar concerns about the usefulness of the new evaluation for the teachers they describe as "F-level players." The issues of

classroom management and planning that so often plague weak teachers are reflected in just a few of the 25 sub-standards. Yet for poor performers, these issues dwarf their ability to address the other standards.

It is hard to distinguish to what extent the concerns relate to the tool versus the infrastructure that CMS needs to put in place to support teachers in meeting the standards. Such an infrastructure could prioritize specific standards and/or adapt the rubric to meet the needs of different groups of teachers. This is a real challenge the system faces as it holds steady on its commitment to ensuring educator effectiveness and a significant shift in evaluation practices in the context of a fiscal crisis and decreased resources.

System As A Learning Organization

In talking with central office leaders responsible for executing different aspects of the teacher effectiveness agenda, it is striking to hear each of them talk about how Strategic Plan 2014 is clearly a living document that guides their work. For example, one of the six priorities of the strategic plan focuses on increasing the graduation rate. Every employee has a personal goal that connects her work to raising the graduation rate and part of their performance assessment relates to their achievement of it. This is one way the system keeps employees focused on the most important priorities of the district.

Similarly, there is shared understanding of the teacher effectiveness work as part of the system's equity agenda. There is widespread agreement that providing every child access to a teacher with demonstrated success in raising student achievement and ensuring that the most effective teachers have incentives to work with the most struggling learners will drive improvements in student achievement and raise the graduation rate. This clarity and shared purpose sets the foundation for the system to function as a learning organization. To make this happen, CMS has put in place various structures to support adult learning and collaboration.

Professional Learning Communities (PLC)

In the 2008-09 school year, the superintendent introduced the idea of Professional Learning Communities as a vehicle to support adult learning and collaboration to address common problems of practice. The district set the context for PLCs by requiring all of them to set a SMART (strategic, measurable, attainable, relevant, and timely) goal that connected each PLC's work to the district's strategic plan and the goals for gains in student achievement and the graduation rate. PLCs exist at every level of the system, with teachers, principals,

principal supervisors, central office staff and senior managers participating in them. CMS hired a consultant who is an expert in PLCs and works closely with Rick DuFour, a nationally recognized expert on the topic, to educate everyone in the system about the concept of PLCs as well as how to practically set them up for success. This work has been supported throughout the system for the last few years.

In the first year, the emphasis was on ensuring that PLCs had norms and protocols to support their functioning and established SMART goals tied to student achievement and the graduation rate goal. Every principal PLC submitted its SMART goal for review by the CAO, the supervising zone superintendent, and the PLC consultant. This team provided feedback on goals and ensured that the goals were rigorous and appropriate. With that as a foundation for PLC functioning, the second year of work focused on integrating the PLC work with the district's major improvement efforts: using data and an improvement cycle to drive instruction and the implementation of tiered instructional supports for students as the system implemented Response to Intervention (RTI).

To see how these foci played out in PLCs, consider the example of a principal PLC focused on narrowing the achievement gap with students who are English language learners (ELL). The group of principals who came together in a PLC all had a significant ELL population in their schools and an achievement gap they knew they needed to address. The principals first set a SMART goal related to narrowing the gap. They then researched strategies for supporting English language development and scaffolding ELLs' learning. The participating principals committed to implementing the strategies in their schools and the focus of the PLC then shifted to talking about the nuts and bolts of implementation, sharing progress, and tracking results.

This group's work followed the Datawise cycle CMS had adopted; it also leveraged the tiered intervention strategies of RTI.

The zone superintendents' PLC provides an image of this work at another level in the organization. Ann Clark, the CAO, facilitates this PLC, which consists of all the zone superintendents who supervise principals. Backwards-mapping from the district's graduation goal, this PLC decided to focus its attention on ninth-grade retention, which is shown to be a significant contributing factor to the graduation rate. The PLC set a goal of realizing a district-wide ninth-grade promotion rate of 90 percent (up from 82 percent) and a graduation cohort rate of 75 percent (up from 69%). Meeting weekly, the PLC began by reading about the impact of ninth-grade retention on graduation and looking at CMS data that provided a sense of the scope of the issue. The PLC disaggregated data to compare the performance of subgroups of students and to assess trends within and across schools. After researching different approaches to improving ninth-grade promotion, the PLC eventually developed a ninth-grade credit-recovery strategy for implementation in the high schools. Once that work began, the team looked at rosters by school to set goals for students and assess how credit recovery was going. It compared progress across schools with similar student populations to see what could be learned about the impact of different approaches.

Teacher Working Teams (TWT)

As CMS embarked on the messy work of defining measures for assessing teacher effectiveness, it developed TWTs as a process for engaging teachers in the inquiry and decision making. TWTs would give teachers a voice in setting the direction of a significant change in practice that has implications for their work, the assessment of their effectiveness, and, ultimately, their compensation. Teachers are invited to participate in design teams that study research and best practices associated with a proposed initiative and identify potential challenges and ways to address them, using a common process. Teachers run the TWTs and get support from a central office staff person assigned to TWTs. The teachers regularly bring their work to larger groups of their colleagues for feedback and refinement. This structure puts teachers firmly in the role of defining the problem to be studied and shaping the solution.

The format of TWTs' work is to methodically answer three questions: 1) Does this measure matter? 2) If it does matter, how might we measure it? and 3) Who should measure it? TWTs follow a consistent process as they answer these questions and work through a process of engagement and vetting that leads to a proposal. (See Appendix D for the template teams use to guide their work.) The steps include:

1. Research the issue and best practices and develop initial recommendations;

2. Hold focus group of teachers to share work and solicit responses to specific questions the design team is struggling with;

3. Refine thinking and recommendations;

4. Present research, findings and recommendations at a Town Hall meeting to which all key stakeholders are invited; solicit high-level feedback and agreement or disagreement with team's proposal;

5. Present recommendations to CMS executive team;

6. Executive team responds to design team's work in writing;

7. School board has access to design team recommendations and executive team's response before making final vote on issue.

The seven Teacher Working Teams (TWTs) include:

1. Value Added Measures (school, team, individual): focused on modifying the value-added statistical model for both state tested grades and subjects and non-tested grades and subjects based on learning from early implementation

2. Professional Learning Community (PLC): focused on measuring school's level of functioning as a PLC and individual teachers' contribution to the school's PLC

3. Student Learning Objectives: focused on how to apply the learning from the student learning objectives work initiated through the TIF grant into the evaluation

4. Student Survey: focused on developing a student survey of classroom experience; has chosen Tripod Project survey and is working on designing a pilot to test implementation

5. Teacher Observations: focused on how to make classroom observations a growth opportunity for teachers through the use of peers and videotaping and recognizing teachers' participation in this as part of their evaluation score

6. Hard to Staff Schools and Subjects: focused on how to create incentives for the most effective teachers to take these positions and earn additional compensation

7. Teacher Work Products: focused on how to include analysis of the tests and assessments teachers develop in their evaluation rating; developing a rubric to assess teacher-developed assessments

8. Content Pedagogy: focused on defining the specific pedagogical skills required to teach specific content and assess teachers' skills in this area

The example of the work of the Teacher Working Team 2 – focusing on PLC-measures – brings these steps to life. First, the team agreed that it was important to create a measure of teacher collaboration with peers and their contribution to the overall mission of the school, because they believed these things contribute to improving instructional practices. So then the question became, *How would we measure performance in this area*? In trying to answer this question, the design team realized that it needed to think about this issue at two levels: the school's level of functioning as a PLC; and individual teachers' level of collaboration and contribution to the school's functioning as a PLC. To assess the school as a PLC, the team is developing a survey, drawing from the work of Rick DuFour. The survey will ask teachers to rate the school on the extent to which it has things in place required to support the school as a PLC: e.g., common planning time, time for teachers to observe one another, and vertical and horizontal teams to address instructional issues within and across subjects.

In thinking about measuring a teacher's individual contributions to the school functioning as a PLC, the team drew on an article that outlined ten different ways teachers can contribute to their school's PLC culture. The list included things like: share practice, conduct action research in your classroom, serve as a mentor to another teacher, and serve as a technology specialist for the school. The

design team envisions that every teacher would select one way to contribute to the school and work on it. Two colleagues the teacher selects and two other teachers, identified by the principal, would assess the teacher's work in this area.

Finally, the school PLC rating as calculated by the survey responses and the individual teacher's rating would be combined to give the teacher her PLC rating. The team felt strongly that individual teachers are responsible for fostering a school-level PLC culture so they should be held accountable for the school's PLC rating as well as their individual work. The next step for this team is to pilot the survey and individual PLC design in a small number of schools to test the model. The learning from the pilot will help determine what needs to be refined, and will assess the overall value and viability of the measure.

The spirit that CMS has been able to create in this work was illustrated when one of the central office staff who supports several of the TWTs explained, "Teachers sitting on these committees don't necessarily agree with everything about value-added and multiple measures, but they're engaged in the work to ensure the approach the system pursues is thoughtful and well-conceived."

District Senior Leadership Team

At the most senior level of the system there is a similar commitment to sharing ownership and responsibility for the district's work. The superintendent has an executive committee made up of the cabinet – the CAO, the Chief Financial Officer (CFO), the Chief Operating Officer (COO), the Chief Human Resources Officer (CHRO), the Chief Accountability Officer, the Executive Director of Communications, and the General Counsel – the five zone superintendents, the chief information officer, and the assistant superintendents for exceptional students and auxiliary services. This group meets weekly and functions as the Project Management Oversight Committee (PMOC), which is charged with tracking implementation of the strategic plan and problem solving issues that compromise effective strategy execution.

Each week, managers responsible for implementing specific tactics of the strategic plan report progress and impediments to the committee. They are encouraged to present problems of practice they

are facing that the PMOC can try to help solve. The analogy used to describe the PMOC's work is that of air traffic control. It knows how each of the tactics of the strategic plan is progressing, where the problems are, and who is involved in which tactics. It tracks where there are opportunities for coordination across tactics, collaboration of central office staff in the implementation of the strategic plan, and how to strengthen cross-functional work.

At the same time, the senior leadership team has several structures in place to support its own learning. Periodically, the superintendent invites in thought leaders in areas the district is focused on to share the state of the work and best practices identified nationally. Recent visitors included Jon Schnur, the founder of New Leaders for New Schools, who talked about the direction of principal leadership, and Dan Weisberg and Karla Oakley from the New Teacher Project, who shared the most promising practices for measuring teacher effectiveness. In addition, the team explored the Common Core State Standards to help everyone understand the shift in expectations they represent. Twice a year the team has a full-day retreat off site to explore an issue that is critical to Charlotte's work. Recently, the topic was poverty and privilege.

Voices From The Field

With a year of implementing the new teaching standards under their belt, both teachers and principals have important perspectives on both the new evaluation and its implementation.

Principals

Principals agree that the instrument is beginning to drive instruction and that this is a positive development. Most significantly, principals see that the emphasis in classrooms has shifted to what students are doing rather than what teachers are doing.

Principals like how this new emphasis reinforces the idea that teacher actions shouldn't be measures independent of students, as if they are an end in themselves but, instead, as a means to the most important end: student achievement. Principals described the new standards as "much clearer about instructional expectations" than the old instrument and explained that they have had "some of the most powerful, good conversations with teachers" using it. There was less agreement on whether the teaching framework (developed by the state) reflects CMS's priorities for student learning, such as 21st century skills, global skills, culture and diversity, and the Common Core State Standards. Some principals felt there was a disconnect between those priorities and the teaching standards.

Principals also noted that the standard related to teacher reflection has also prompted a shift in teacher practice. Teachers are characterized as "more thoughtful about what they are doing," and as having "a clearer understanding of the elements they need to work on." Principals also see the new tool encouraging teachers to collaborate. As one principal explained, "This makes teachers think

more than they ever had." Another commented, "Teachers are taking charge a little more about their work." Principals described the impact of the new evaluation instrument on teachers as differing depending on teachers' experience level. One principal made the observation, "Newer teachers are paying a lot of attention to it, but it's hard to create a sense of urgency with career-status teachers." A colleague offered a different perspective on the impact on experienced teachers, commenting, "Veteran teachers are used to getting 4s (the top rating on the old system). This has helped them stop and think. It makes teachers more reflective about their teaching practices and their impact on student learning."

While principals mostly appreciate the observation instrument, there is consensus that implementation has been challenging. Some obstacles relate to the training teachers and principals received related to the framework and their readiness to implement it. When the framework was introduced, principal received three days of training and were then expected to provide the same training to their teachers. Principals described the training that actually happened at schools as "very varied."

When thinking about their own training, principals wish they had received more in year one and that the emphasis had been different. The year one professional development was described as "process-oriented: how to fill out the forms, how many observations to do, how long to stay in the class." Principals' professional development in year two focuses on calibration of scores. They watch videos together, discuss their assessments of teacher practice, and come to a shared understanding (and rating) of the practice. Principals describe this training as "valuable," helping them "develop a common understanding of effective instruction" and

they wish it had begun in the first year of implementation. When asked about the level of principal engagement in the new evaluation work, principals' comments reflected a broad range. One principal offered, "I'm an eternal optimist. I think everyone is engaged," while another disagreed. "I'm not convinced," he said. "The system didn't dedicate the time to roll it out well."

A second set of obstacles relate to implementation. Principals described spending two class periods debriefing with teachers after a formal observation, and lamented, "and that didn't include a review of their artifacts." In year two they are getting better at figuring out how to complete the evaluations more efficiently, but they described them as "very labor-intensive." One principal captured the trade-offs associated with such a labor-intensive evaluation, explaining, "I don't get back in [the classroom] to check on implementation of my feedback."

The on-line tool, developed by McREL, for managing evaluation data was characterized as "in constant flux and not user-friendly." Principals want videos of exemplary teaching practice both to support their own learning and to offer to teachers as resources to support their improvement. Principals thought that piloting the system would have allowed many of these problems to be addressed with much less stress on the principals. One explained, "Every meeting we went to there was a new change announced."

Principals also said they experienced a variety of mixed messages, which made figuring out a new system harder. The training left them confused about the extent to which they should be focusing tightly on observation data vs. a variety of artifacts. When a couple of principals expressed this concern, another colleague said, "Wait till next month. The whole training session is going to focus on how to assess artifacts and integrate them into the rating." Several principals were particularly startled when they participated in a session with a staff member from the North Carolina Department of Public Instruction. The state staff member rated a teacher's performance higher than the CMS principals did. This made them realize that, as daunting as it is to get inter-rater reliability in a district the size of Charlotte Mecklenburg, achieving it across the state is an enormous undertaking that requires high expectations, leadership, and resources at the state level.

Teachers

"Confusion" is the word teachers most commonly used to describe their experience with both the substance and the procedures associated with the new teacher evaluation system. When the new evaluation system was rolled out to teachers in the fall of 2010, the two components of the system were classroom observations and value-added measures. While teachers understand other measures are being explored (and are participating in that exploration), they say the general sense among teachers is that the evaluation is all about summative test scores. That perception, combined with three years of budget cuts, frozen teacher wages, and RIFs, leave teachers anxious about the purpose and intentions of the new teacher evaluation system.

Teachers' level of confusion appears to vary, though, by the school they work in and the amount of training they have received. In a focus group of teachers, descriptions of the school-based training they received on the new evaluation system were quite disparate. One teacher is participating in an eight-week course on the framework that happens during her weekly grade level team meetings. For forty-five minutes a week, she and her colleagues study the framework and are quizzed on it. Another teacher had a two-hour training at the beginning of the year, where the principal walked through the state-provided PowerPoint that outlines the system, handed her a manual, and told her she needed artifacts for the evidence-based component of the evaluation. No one ever explained what artifacts are, how to select and organize them, or anything else. A third teacher was handed the PowerPoint and a handout and then the principal moved to the next item on her agenda. The handout included a link to a tutorial to the framework, but when the teacher went to use it, the tutorial instructions did not work and she couldn't access it.

Teachers' experience of being evaluated surfaced further inconsistencies. They described how some of their observations can last sixty to ninety minutes, while others are thirty minutes long. The idea of the teacher sharing artifacts is intriguing to the teachers; it provides them an opportunity to tell their story and they think that works best when the principal really knows them and their work. Some teachers, though, expressed discomfort with the skills and inclinations of their principals; as one

teacher commented, "There is a big problem with principal bias."

As teachers talked about the new evaluation system they tended towards two reactions: "I need to keep working harder" or "it doesn't matter how hard I work, I won't ever get above proficient." On the positive side, one said, "This instrument isn't bad at all. It's trying to be reflective, identify strengths and weakness and promote instructional growth." Teachers agreed that the new evaluation system is altering the conversation for some teachers, but as one veteran explained, "I'm too old and too stubborn to focus much on these new standards. I'm just going to run my grade-level team like a Professional Learning Community."

Another teacher joked about the framework's emphasis on teachers as leaders and how that's leading to burnout, commenting, "It's only November and we've already had five teachers resign this year." One teacher worried that everyone is so "stressed out and overwhelmed" that new teachers are getting less support from their experienced colleagues when they need it the most. When one teacher summed up her feelings, saying "I feel like apathy has been created," all the other teachers in the room sat silently, perhaps illustrating her point.

Communication And Messaging

There were a number of vexing communications challenges in the first year and a half of implementing the new evaluation system in Charlotte. Some of these challenges can be attributed to the fact that the district sits awkwardly in the middle, between the state, which sets much of the evaluation policy, and teachers, who want clarity about this high-stakes endeavor.

Many teachers weren't able to distinguish between the decisions the district was making and those the state was passing down over which Charlotte had no control. Teachers, by and large, interpreted everything as originating from the district, which meant that the state's missteps cost the district credibility and good will. The contentious issues for teachers and principals that North Carolina has authority over include: the teaching framework to be used (Charlotte customized indicators to reflect its priorities); the number of observations each teacher would receive; the timeline in which the observations have to be completed; the expectation that any teacher rated "developing" in a standard is defined as "ineffective" and has to engage in corrective action; the forms that have to be used throughout the process; and the information management system into which evaluators record their observation results.

An example of the confusion caused by the state surfaced in the fall of 2011 as year two of implementation began. The state introduced a sixth standard – tying teacher performance to student achievement – to the existing five standards measured in the observation framework. The language of the standard is: *"Teachers contribute to the academic success of students. The work of the teacher results in acceptable, measurable progress for students based on established performance expectations using appropriate data to demonstrate growth."*

Nobody in CMS could explain what the state was planning to include in this standard and rumors were running rampant that this was another way – beyond the value-added measures – to emphasize student test results in teacher evaluations. Teachers, principals, and senior central office staff alike had received no clear communication from the state about either the intention or the details of the new standard. There was no information listed on the North Carolina Department of Public Instruction's website about this standard and what it encompasses. While that language sounded reasonable to many teachers, the addition of the new standard without clear communication provoked anxiety and distrust which teachers tended to direct at the school system.

While state-level communication created challenges for Charlotte, the system put a variety of strategies in place to support strong two-way communication throughout the district about the evaluation system. The district relied on a pre-existing structure to inform teachers and administrators about the work and to receive feedback from them about its implementation. In every school, teachers elect a colleague who serves as their school's liaison to the larger system. These teachers serve on a teacher advisory committee for their zone (geographic grouping of schools, K-12), working with their zone superintendent. The zone superintendent meets with the group monthly to hear how specific initiatives are being implemented and received in the schools. These discussions surfaced differences in how work is being pursued across schools and concerns teachers in particular schools have. Using the information provided by the teachers, zone superintendents followed up with individual principals to solve problems in an effort to address concerns as quickly and locally as possible.

Each zone teacher advisory group, in turn, elects one of its members to serve on the countywide teacher advisory committee. Once a month, this group meets with district leadership for a full day, including standing meetings with the superintendent and CAO. These meetings provide opportunities for district leaders to solicit input on major initiatives and share thinking and early work underway on new priorities. All other senior leaders have the opportunity to meet with the group each month to solicit feedback on plans for major upcoming work and to get feedback on how implementation of key initiatives is going and how the work is being received at the schools. Senior leaders take this feedback back to their team to inform improvements and communication with principals.

These efforts, while promising, were undermined by what teachers and principals (and central office leaders) experienced as confusing and contradictory messages about the purpose and intentions of the system. When the evaluation system was first rolled out in 2010, the district tried to convey the view that it was a valuable tool to support teacher growth and development. Principals and teachers both remember hearing that message very clearly and felt it was a positive message that supported their work. As the year wore on, however, other messages were communicated that undermined the growth and development orientation and generated distrust.

One issue that muddled the message was the role of the evaluation system in teacher pay. Charlotte had been experimenting with pay-for-performance through its TIF grants and intended to refine and expand that work, integrating the new evaluation data. It didn't intend to formally link pay and performance until 2014, when value-added and other measures would be thoroughly vetted and everyone in the system would be well versed in the new observation tool. Yet, the plans around pay-for-performance loomed in the distance and local media outlets focused on how CMS was revamping teacher compensation to recognize teachers' impact on student learning. Then-superintendent Gorman was frustrated that the central tenets of CMS's evaluation work – supporting teachers' continuous improvement, engaging them in identifying the right mix of measures, and taking three years to get the measures right before tying them to compensation – got lost in the press. Instead, the headlines empha-

sized paying teachers for student performance and teachers' worry about taking pay cuts when the new system goes into place.

Concerns about the evaluation rating and the other measures being a vehicle for pay-for-performance were exacerbated by Gorman's efforts to change state law regarding teacher compensation in Charlotte. A 2007 North Carolina law allows Charlotte to adjust the state teachers' pay scale by securing a majority vote of its teachers. In spring 2011, Gorman, wanting additional flexibility, worked with legislators in the House to introduce House Bill 546, which would give him the authority to revise CMS's teacher compensation structure without having to obtain a majority vote by teachers. The bill passed through the House and is now "parked," on hold before going to the Senate. Predictably, teachers objected to this proposed change, reflecting a growing sense of distrust.

The final messaging challenge, which further eroded trust, related to the RIFing of 107 teachers based on their summative evaluation. This continued a policy CMS began in 2008 of laying off teachers based on performance rather than tenure in the system. On the face of it, this makes perfect sense. The system is developing a stronger evaluation system and wants to use it for consequential decisions. Yet, it was the first year of implementation, there had been no pilot of the system, the kinks were being worked out through implementation, and the messaging had been focused on growth and development. Teachers and principals had trouble holding the two messages – supporting growth and development and exiting poor performers from the system – and having a measured interpretation.

Principals described the effect of all of this in their schools as a "damaged culture," and worried that the good intentions of the evaluation system would be undermined. As one principal commented, "We HRed it. We took it from a growth tool to a HR function." His remark underscores how hard it is to effectively message that an effective teacher evaluation tool is about both growth and development and consequential employment decisions.

The final messaging problem occurred in the summer of 2011 when teachers' value-added scores (in NC state-tested grades and subjects) became

available. In 2010, when this data came available it was shared with teachers. In 2011, the decision was made to share the information with principals regarding their teachers' performance, but not to share it directly with teachers. District officials explained that they wanted to give principals the information and help them learn how to use it effectively before making the information available to teachers. This decision reflected Charlotte's historic culture of principals playing the role of gatekeeper in the school system. Given that teachers had been given their value-added the year prior, their suspicions were raised when it – the most controversial measure in the new system – wasn't shared with them in 2011.

Year Two Priorities

With the initial implementation of the new evaluation system completed, a foundation has been set for the work. It has also become clear where the next stage of implementation needs to focus to ensure the foundation is solid and built upon. In year two of implementation, CMS is digging more deeply into three priority areas: principal expertise in observing, analyzing, and discussing instruction; teacher development in standard areas; and defining the additional measures that will be included in teachers' evaluation.

The focus on principal scoring reflects the need to strengthen the validity and reliability of the ratings. CMS began year two with three validation studies underway to assess the reliability of evaluator ratings in year one. One of the validation studies CMS has embarked on engages forty principals in reviewing and rating five videos that have been normed by an expert, using the CMS four-point rating scale. Initial results suggest that principals' ratings vary widely and that there is work to be done with principals to ensure calibration to an agreed-upon rating norm and inter-rater reliability.

Principal professional development in year two is focusing on the calibration issue with time devoted each month to principals rating and discussing normed videos. Their training is also prioritizing how to look at multiple data sources. Principals are being asked to use a variety of artifacts in teacher evaluation, and they need help thinking about what artifacts are most useful, how to collect and analyze them, and how to synthesize this array of data and the observational data into a performance rating. The third area of focus for principal development in year two is around the skills of coaching conversations. Principals are being taught how to use the array of data and the calibrated scores to talk with teachers about their practice in ways that help them set growth targets, organize support to help them realize them, and measure progress.

To complement the year-two focus on principal skill development, CMS is working to build a teacher professional development system that is aligned to the teaching framework and supports growth and development. The goal for June 2012 is to have all of the system-supported professional development aligned to the teaching framework. This will help teachers choose professional development in response to their evaluation ratings and help CMS identify training gaps that need to be addressed. The system is also beginning to develop content for a digital library that will include videos, articles, lesson plans, and other resources organized by teaching standard, which teachers will be able to easily access to study a standard they want to work on.

As a district that participated in the MET study, CMS is synthesizing the learning from extensive videoing of teachers across the system to inform its thinking about how it can integrate video (footage and the act of videoing) into teacher training and support. Armed with more than fifty cameras CMS inherited from its participation in the MET study, the district is working to develop ways to use video to support teacher reflection. The ultimate goal is an online video library with footage arranged by teaching standard and other instructional priorities the system has identified.

The final priority for year two is to move the work of Teacher Working Teams (TWT) from the inquiry stage to being able to pilot the additional measures they are helping develop in fall 2012. This requires that by June of 2012 each TWT has defined what it wants to measure and how it wants to measure it, and has developed the measurement tool (most likely, rubrics for many of the TWTs) for measure-

ment. The intention is to spend the 2012-13 school year piloting the measures and determining which measures provide the most valuable complementary data and how many measures CMS can afford (both financially and capacity-wise) to implement before bringing a plan for additional measures to the executive staff and school board for a vote in June 2012. This schedule will allow a full year of implementation of these measures before the 2014 synthesis of all measures into an overall evaluation rating that will inform pay decisions.

Tracking this work to discern the relative merits of each TWT's proposal and then the interrelationships between different measures will provide CMS a rich array of information for making decisions. As the multiple measures work evolves and the final measures come into focus, CMS will have the opportunity to explore the interrelationships between these measures and the teaching standards. For example, the TWT work on a PLC measure will likely result in an assessment that measures some things that are currently included in the first teaching standard, *Teachers demonstrate leadership*. This is one example of how the use of multiple measures will afford CMS the opportunity to triangulate data about teacher performance in particular areas.

While the district is refining and expanding its evaluation system in year two, it is doing so in a changing leadership and political context. In the summer of 2011, Peter Gorman left the superintendency. Hugh Hattabaugh, the chief operating officer under Gorman, was appointed interim superintendent for one year while a national search is conducted. In this year of transition, the work continues with what some staff described as "not a strong presence regarding the talent effectiveness work." Gorman's

vocal championing of this work will be hard to replicate. The November 2011 school board election added another variable to the issue of context. The election of two new board members whose platforms range from sustaining the Gorman reforms to rolling them back will create an interesting dynamic. The board's choice of CMS's next superintendent will likely signal its commitment to the teacher evaluation and the teacher effectiveness agenda.

This changing political context also relates to the district's relationship to the state. CMS is working in partnership with the North Carolina Department of Public Instruction to share its work and learning from its pioneering efforts in evaluation to inform the state as it develops the infrastructure (particularly related to value-added measures and not-tested grades and subjects) it needs statewide. This partnership is important because there is the opportunity for CMS to leverage its early work to inform or drive the direction the state takes. The more state policy and tools align with CMS's work-to-date the easier it will be for CMS to meet state expectations, the clearer and more consistent the messaging to teachers will be, and the more the whole state will benefit from CMS's early work and learning. The most significant short-term implication of this partnership relates to CMS's original plan for developing value-added measures for every teacher in the system. CMS has paused its development of assessment except in the areas of art, physical education, world languages where the work with Discovery Education and teacher teams is underway. Beyond that, CMS is focusing on sharing its assessment work with the state and giving the state time to decide how it will approach this issue and to what extent it might want to adopt or adapt CMS's value-added calculation and the tests it has developed.

Conclusion

At this stage in the teacher evaluation work in CMS, some dynamic tensions have emerged that reflect the complexity of this work. They speak to former superintendent Gorman's observation that "this work is a lot more nuanced than we ever realized." Several of the tensions that the system is facing are worth noting:

1. Balancing the system's focus on accurately measuring performance with a commensurate focus on building the capacity within the system to observe, analyze, and discuss practice and support instructional improvement

   To drive instructional improvement, teacher evaluation systems need to sit in the context of a larger performance-management system. Such a system both tells teachers how they are performing and supports their improvement. Using data to tell teachers how they are performing and to make high-stakes decisions raises a plethora of questions – which measures to use, how to ensure fairness – that districts need to answer. And there is great pressure on systems to figure this out. CMS is in the midst of sorting through these issues and trying to develop a robust measure of effectiveness.

   Yet, for the vast majority of teachers, telling them how they are doing is most helpful when it is accompanied by very concrete, specific feedback about their practice and how they can improve and supports to help them do so. Building the capacity of evaluators to talk in substantive ways about instructional practices and their impact on student learning is a significant undertaking. It requires learning how to shift from global

characterizations of instruction – *The teacher engages students in their learning* – to specific, descriptive information about which students are engaged, how they are engaged, and when and what it looks like when they aren't engaged. Talking with teachers about their observations also requires skills that are new to many evaluators and will likely challenge some school cultures. Finally, realigning professional development is a messy undertaking that usually implicates many departments and challenges long-standing ways of doing business.

The political environment in which all of this work is happening favors a focus on measures. Knowing that the other work is what will drive improvement, school systems need to develop a multi-year strategy to build the capacity to analyze instruction and develop tailored support to foster instructional improvement. This plan needs to be as robust as the one for creating a robust effectiveness metric.

2. Supporting the varied needs of teachers through the teacher evaluation system

   An evaluation system needs to support novice teachers to develop their practice, help all teachers continuously improve, and address incompetence. Given these different development needs and the system's different interests relative to each, how do you differentiate the teacher evaluation instrument and process to be responsive? This is akin to the challenge teachers face every day in their classrooms. Do you prioritize certain standards to make them more accessible for new and/or struggling performers? Do you use the

standards as the starting point for a deeper study of content pedagogy for teachers who are ready for that work?

As significant an undertaking as it is for a district to implement a new teacher evaluation system, the real promise lies in adapting it to meet the needs of the array of teachers in the district and the district itself. Adaptation can take several forms. One relates to how the evaluation instrument is used as a development tool to assess classroom practice based on the experience level of the teacher. Another relates to how the evaluation process is differentiated based on experience and performance. Districts need to consider both.

3. Setting clear directions for the work and providing necessary support while honoring local autonomy

Within the rollout of CMS's new evaluation system there are several examples of choices the system has made to hold some aspect of the work tightly while giving teachers and principals a lot of flexibility about other pieces of the work. The work of the TWTs is an example of holding something loosely – giving teachers a general frame for their work, followed by a lot of autonomy and authority to create. Conversely, the tight oversight of the procedures principals follow relative to the process of evaluating teachers was aimed at ensuring the integrity of a process that would inform which teachers were affected in a RIF. Gorman's efforts in the spring of 2011 to gain control over the design of teacher compensation is an example of him wanting authority to tightly control something that he sees as a lever to drive equity and acceleration of student achievement.

PLCs, on the other hand, represent a way the district is trying to create space for educators to define and shape their work in hopes it leads to deep learning and innovation. Early results appear to be mixed, and the district finds itself in a moment when the new evaluation work requires more training and development than it currently has the time or resources to provide. And so there may be a need to reshape the PLC work in a way that supports implementation of the new evaluation without disempowering principals and teachers. In addition to managing the tight-loose balance,

the challenges in all of this is to be clear and consistent in the messaging to build common understanding of the decisions and the rationale behind them.

4. Ensuring that the mix of measures used to assess teacher effectiveness are complementary and well-integrated and provide both a consistent and holistic assessment

Given the high stakes associated with tying teacher performance to compensation and the nascent stage of the value-added work, CMS is working to include a mix of measures in its assessment of teacher effectiveness. As it pilots different measures to identify which ones provide the most useful information and can be manageably integrated into its effectiveness model, the system will need to combine these in a way that creates a consistent message to teachers about their performance. It is likely, in the short term, that some of the measures will assess similar things, as the example earlier in the paper about the PLC measure and the first teaching standard of the evaluation document illustrate.

There are two issues related to this that are worth considering. First is the need to triangulate different measures with one another, not just with value-added to see where there are strong relationships. Given the lack of experience using value-added in teacher evaluations, the fact that it is not yet well-enough honed to answer all the questions of its critics, and the strong reaction the whole concept provokes, triangulating more broadly to determine the relationships between different measures is wise.

The system's leaders on the assessment work envision that, over time, they may be able to figure out, in the case of different measures assessing the same things, which of them is the best measure on which to rely. Ultimately, this might reduce the number of measures or the uses of different measures. This is dynamic work and the challenge that CMS and other systems face is how to build the best possible set of measures, ensure their integrity, and do this in a way that is comprehensible to teachers and does not raise anxiety or stress to the level at which the entire system is compromised.
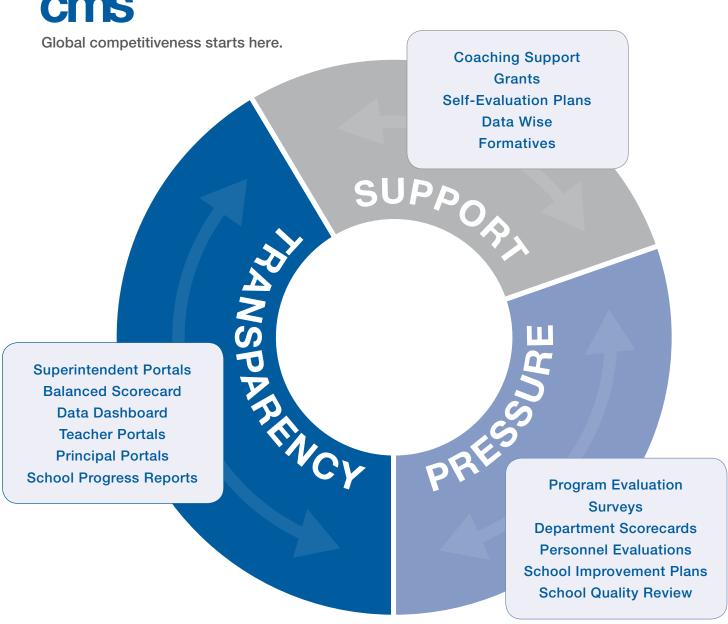
5. Messaging effectively, ensuring consistency between words and deeds, and being honest about the big picture

CMS implemented a variety of thoughtful communication strategies to engage teachers in the new evaluation system and to build their ownership of it. And from the outset this work was framed as focused on teacher growth and development. Tension arose when pay-for-performance and RIFs – two high-stakes issues related to pay and employment – were introduced in relationship to the new evaluation system. These high-stakes issues eclipsed the theme of growth and development in teachers' minds and there was a growing sense that the system wasn't being consistent or trustworthy in sharing its intentions for this work.

The very public effort to rewrite legislation that stripped teachers of a voice in compensation issues and the RIF were problematic both in terms of process and timing. Both initiatives were pursued with no input from or outreach to teachers and both were pursued in the first year of implementation of the new evaluation system. Yet the principles that lay beneath each effort – that teachers' compensation should be influenced by their effectiveness and that poor-performing teachers should be exited from the system – are ones that many school systems subscribe to and are fundamental to a performance culture. The challenge arises in how these multiple messages are communicated and sequenced in a way that builds understanding and ultimately buy-in.

> **CMS's teacher evaluation work is thoughtful and promising. It acknowledges the complexities while trying to realize the full potential of a fundamentally different orientation to defining and measuring teacher effectiveness, all in the context of declining resources. While the context of CMS, like that of every system embarking on this work, informs how the district is pursuing the design and implementation of a robust teacher evaluation system, there is much to be learned that is generalizable. CMS's willingness to share its approach, its successes, and the challenges it is encountering along the way will help accelerate the learning in the field and the sector's continuous improvement.**

Appendix A – Cycle of Continuous Improvement

cms

Global competitiveness starts here.

Coaching Support
Grants
Self-Evaluation Plans
Data Wise
Formatives

Superintendent Portals
Balanced Scorecard
Data Dashboard
Teacher Portals
Principal Portals
School Progress Reports

Program Evaluation
Surveys
Department Scorecards
Personnel Evaluations
School Improvement Plans
School Quality Review

SUPPORT

TRANSPARENCY

PRESSURE

- **STRATEGIC PLAN 2014**

- **REFORM GOVERNMENT POLICIES**

- **BOARD OF EDUCATION VISION, MISSION, CORE BELIEFS & COMMITMENTS, THEORY OF ACTION**

Appendix B – List of NC Standards and Sub-Standards

Standards 1: Teachers Demonstrate Leadership
Teachers lead in their classrooms
Teachers demonstrate leadership in the school
Teachers lead the teaching profession
Teachers advocate for schools and students
Teachers demonstrate high ethical standards

Standards 2: Teachers Establish a Respectful Environment for a Diverse Population of Students
Teachers provide an environment in which each child has a positive, nurturing relationship with a caring adult
Teachers embrace diversity in the school community and in the world
Teachers treat students as individuals
Teachers adapt their teaching for the benefit of students with special needs
Teachers work collaboratively with the families and significant adults in the lives of their students

Standard 3: Teaches Know the Content They Teach
Teachers align their instruction with the North Carolina Standard Course of Study
Teachers know the content appropriate to their teaching specialty
Teacher recognize the interconnectedness of content areas/disciplines
Teachers make instruction relevant to students

Standard 4: Teachers Facilitate Learning for Their Students
Teachers know the ways in which learning takes place, and they know the appropriate levels of intellectual, physical, social, and emotional development of their students
Teachers plan instruction appropriate for their students
Teachers use a variety of instructional methods
Teachers integrate and utilize technology in their instruction
Teachers help student develop critical thinking and problem-solving skills
Teachers help students work in teams and develop leadership qualities
Teachers communicate effectively
Teachers use a variety of methods to assess what each student has learned

Standard 5: Teachers Reflect on Their Practice
Teachers analyze student learning
Teachers link professional growth to their professional goals
Teachers function effectively in a complex, dynamic environment

Appendix C – Rubric Excerpt

Standard IV: Teachers Facilitate Learning for Their Students.

C. Teachers use a variety of instructional methods. Teachers choose the methods that are most effective in meeting the needs of their students as they strive to eliminate the achievement gaps. Teachers employ a wide range of techniques including information and communication technology, learning styles, and differentiated instruction.

OBSERVATIONS	DEVELOPING	PROFICIENT	ACCOMPLISHED	DISTINGUISHED	NOT DEMONSTRATED (Comment Required)
STATE DESCRIPTORS	☐ Demonstrates awareness of the variety of methods and materials necessary to meet the needs of all students.	☐ …and Demonstrates awareness or use of appropriate methods and materials necessary to meet the need of all students.	☐ …and Ensures the success of all students through the selection and utilization of appropriate methods and materials.	☐ …and Stay abreast of emerging research areas and new innovative materials and incorporate them into lesson plans and instructional strategies.	
POSSIBLE CMS INDICATORS	▪ Teacher is learning to address individual student's needs through differentiated instruction. ▪ Teacher includes appropriate lesson components when providing instruction (e.g. review, objective, output, modeling, guided practice, independent practice). ▪ Teacher is beginning to use available technology tools to enhance instruction. ▪ Teacher is beginning to include strategies that address students' learning styles.	▪ Teacher regularly differentiates instruction, using different teaching methods and materials. ▪ Teacher appropriately and effectively implements district and/or school initiative. ▪ Teacher regularly implements a variety of instructional strategies that address students' learning styles. ▪ Teacher regularly integrates technology into instruction and student activities.	▪ Teacher consistently utilizes various instructional strategies that effectively narrow the achievement gap for students. ▪ Teacher consistently and intentionally selects the "best practices" to match students' learning needs/styles and the lesson's objective. ▪ Teacher consistently and effectively differentiates instruction to meet all learner needs. ▪ Teacher consistently and effectively includes technology as an integral learning tool whenever appropriate and accessible.	▪ Teacher is a resource for staff, modeling and supporting the use of new and creative instructional strategies. ▪ Teacher constantly explores, researches, and effectively implements new and innovative technology into teaching, whenever feasible. ▪ Teacher's success in closing the achievement gap in his/her classroom motivates other staff to replicate these effective techniques and strategies.	

Appendix D – Teacher Working Team Template

Working Team – (insert measure team is focused on)

Charter

The (insert measure team if focused on) Working Team will first consider whether this measure would be an effective and appropriate measure of teacher effectiveness at CMS. The team will develop recommendations that reflect the team members' experience and expertise as well as research and feedback from other stakeholders. If the recommendation is approved, the team will continue to collaborate, with a focus on developing tools and process in preparation for pilot testing.

There are three guiding questions for the consideration of each potential measure: Does it matter? If so, how can it be measured? Who should measure it? Detailed responses to these questions will inform the team's direction and recommendations. During the design process, the responses will be at a level of detail sufficient to support a decision regarding moving forward with development of an initial version of measurement tools and process. At that point, the team will solicit feedback and the recommendation will be reviewed.

Definition

The XXX measure is….
CMS is considering this measure because…

Timeline

MILESTONE	TARGET DATE(S)
Design – Does it Matter?	
Design – How to Measure?	
Design – Who Measures?	
Feedback/Focus Group	
Initial Design Recommendations Complete	
Field Testing	
Final Design/Development Recommendations Complete	

Research

RESOURCE	DISCUSSION
Summary Report from Initial Design Team	
Article	
Study	

Design – Does it matter?

QUESTION	What difference to student success does it make?
DISCUSSION	
DECISION	
QUESTION	Does it indicate the quality and effectiveness of teaching?
DISCUSSION	
DECISION	
QUESTION	Does it differentiate among levels of effectiveness?
DISCUSSION	
DECISION	
QUESTION	Can teachers grow in this area? What is the connection to professional development?
DISCUSSION	
DECISION	
QUESTION	How might providing feedback on this measure support teachers?
DISCUSSION	
DECISION	
QUESTION	How might providing feedback on this measure support other stakeholders, including students, parents, principals, etc.?
DISCUSSION	
DECISION	
QUESTION	What is the connection to evaluation?
DISCUSSION	
DECISION	
QUESTION	How does this measure align with CMS vision and priorities?
DISCUSSION	
DECISION	

Design – How to measure it?

QUESTION	What instrument could be used for assessment of this measure?
DISCUSSION	
DECISION	
QUESTION	What are the critical requirements for the instrument and the process?
DISCUSSION	
DECISION	
QUESTION	Do instruments exist that could be used for this purpose? Do they meet the requirements? If not, can they be modified?
DISCUSSION	
DECISION	
QUESTION	If a new instrument needs to be developed, what should it look like?
DISCUSSION	
DECISION	
QUESTION	What would the process look like?
DISCUSSION	
DECISION	
QUESTION	How will data be compiled, analyzed, reported, etc.?
DISCUSSION	
DECISION	
QUESTION	How will measurement data be used?
DISCUSSION	
DECISION	

Design – Who should measure it?

QUESTION	Who should evaluate, assess or otherwise measure?
DISCUSSION	
DECISION	
QUESTION	Why these people and not others?
DISCUSSION	
DECISION	
QUESTION	What type of training and support will be needed?
DISCUSSION	
DECISION	

Feedback

Note - additional content will be provided to working teams to assist in completing this section of the document

[focus group planning and logistics]

[focus group questions, presentation development, etc.]

[other review and feedback]

Development – Instrument and process

Note - additional content will be provided to working teams to assist in completing this section of the document

[questions to define the instrument to be tested]

[questions to define the process to be followed for field testing – scope, support, timing, etc.]

[questions to define how a go/no-go recommendation will be reached]