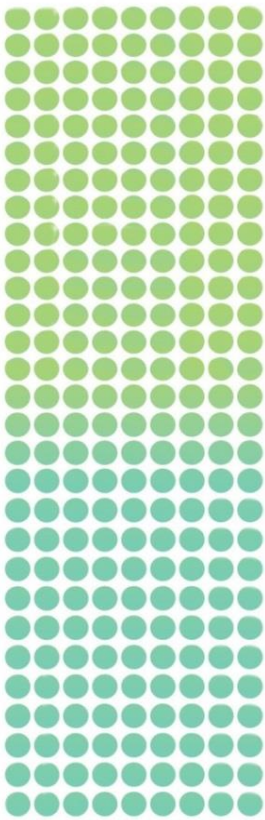
OHIO EDUCATION RESEARCH CENTER

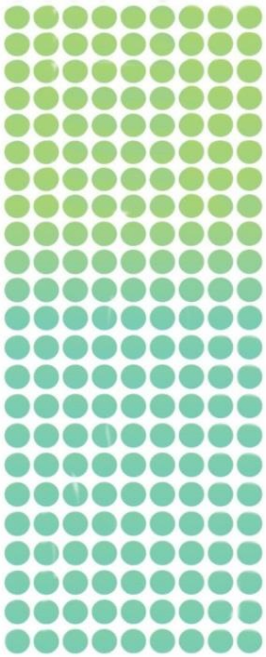
Making Research Work for Education

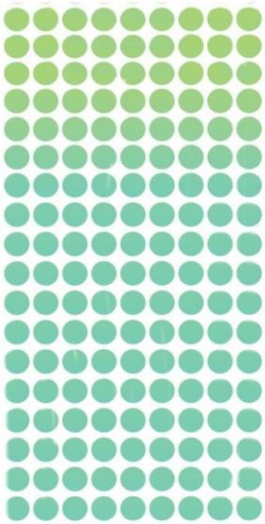
Ohio's Student Growth Measures (SGMs): A Study of Policy and Practice

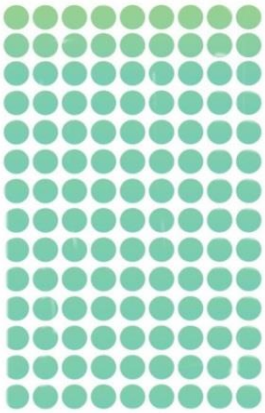
Marsha Lewis, PhD, Ohio University
Anirudh Ruhil, PhD, Ohio University
Margaret Hutzel, MPA, Ohio University

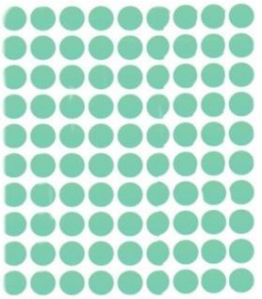


#FR-2015-16
Teachers and Leaders
oerc.osu.edu









OHIO EDUCATION RESEARCH CENTER

The Ohio Education Research Center (OERC) is a **COLLABORATIVE** of Ohio-based researchers from six universities (Case Western Reserve University, Miami University, Ohio University, The Ohio State University, University of Cincinnati, and Wright State University) and four research institutions (Battelle, Battelle for Kids, Community Research Partners and Strategic Research Group) The founding partners coordinate the work of the OERC through the Governance Committee and three standing committees (Research Agenda, Data, and Outreach). Membership includes key participants from State of Ohio agencies and partner organizations. Administratively, the OERC reports to the State of Ohio through the Policy Council. The OERC is headquartered at The Ohio State University.

The **MISSION** of the OERC is to develop and implement a statewide, preschool-through-workforce research agenda addressing critical issues of education practice and policy. The OERC identifies and shares successful practices, responds to the needs of Ohio's educators and policymakers, and signals emerging trends. The OERC communicates its findings broadly, through multiple platforms and networks, producing materials, products and tools to improve educational practice, policy and outcomes.

The **VISION** of the OERC is to be the source for cutting edge knowledge and resources regarding education and training for Ohio's educators, policymakers and community leaders creating a dynamic cycle of research and practice where the needs of practitioners drive the research agenda and high-quality research has a rapid impact upon practice in the field.

Core **FUNDING** for the OERC is provided by the Ohio Department of Education. Additional funding comes from the Ohio Department of Job and Family Services in collaboration with the Ohio Board of Regents.

ABSTRACT

The Ohio Student Growth Measures Policy and Practice study is a two-year study investigating early implementation of Student Growth Measures (SGMs) as part of the Ohio Teacher Evaluation System (OTES) and the Ohio Principal Evaluation System (OPES). For the pilot year and the first implementation year of OTES and OPES (2012-13 and 2013-14), Ohio law required that Student Growth Measures comprise 50 percent of a teacher's evaluation. Ohio law allows SGMs to be measured using a variety of data sources including: 1) teacher-level Value-Added data based on Ohio Achievement Assessments; 2) vendor-developed assessments that have been approved by the Ohio Department of Education; and/or 3) locally determined measures such as Student Learning Objectives (SLOs) or the shared attribution of Value-Added measures. This study's purpose is to provide data to inform state policy and state and local practice around the use of SGMS in teacher and principal evaluation during the early stages of deployment of the new evaluation systems.

During the 2012-13 academic year, 13 of Ohio's Local Education Agencies (LEAs) received funds to pilot SGMs with state support. As part of this study, input from these early-adopter LEAs was sought from teachers and administrators. In addition, in the fall of 2014, data from Ohio's Electronic Teacher and Principal Evaluation System (eTPES) for most Ohio LEAs were obtained and analyzed. These data comprised the SGM data and the Performance on Standards data for every teacher and principal in Ohio who participated in OTES and OPES during the first full implementation year (2013-14). This study utilizes the qualitative data on perceptions and experiences from the 13 early adopter LEAs combined with the statewide eTPES data for the first implementation year of OTES and OPES to provide a detailed description of the early implementation of OTES and OPES in Ohio.

Funding for this report was provided by the Ohio Department of Education.

The Ohio Education Research Center would like to thank the following individuals who helped make this research possible:

Lead Researchers:
Marsha Lewis, PhD, Ohio University
Ani Ruhil, PhD, Ohio University
Margaret Hutzel, MPA, Ohio University

Additional Research Team Members:
Suzanne Franco, EdD, Wright State University
Jill Lindsey, PhD, Wright State University
Lauren Porter, MA, The Ohio State University

For questions, contact the authors of this report: Marsha Lewis (lewism5@ohio.edu), Anirudh Ruhil (ruhil@ohio.edu), or Margaret Hutzel (hutzel@ohio.edu).

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Study Purpose

In most large-scale system change initiatives the early stages of implementation realize both successes and challenges. The initial implementation phase is also a time where much learning can take place that informs policy and practice to improve the new systems going forward. The Ohio Education Research Center was tasked with examining and reporting on the early implementation stages of the new Ohio Teacher Evaluation System (OTES) and the Ohio Principal Evaluation System (OPES). The research team conducted interviews and focus groups with teachers and principals and analyzed the 2013-14 statewide teacher and principal evaluation data extracted from Ohio's new Electronic Teacher and Principal Evaluation System (eTPES).

Ohio's new teacher and principal evaluation systems would be considered by any measure to be transformational organizational changes (Nutt, Burke). Therefore, it is to be expected that the systems will require some modification based on initial learning. It is also expected that such a wholesale change will garner resistance and negative feedback. The old system of teacher evaluation in Ohio and other states relied almost exclusively on classroom observation and often resulted in very little variability in the ratings of teachers within a building (Weisberg et al. 2009; Harris & Herrington, 2015). The new systems that Ohio and other Race to the Top states have started to implement—systems that include measures of student academic growth--have resulted in more discernment/variability in the ratings of teachers and principals than has been the case thus far. This study describes the resulting variability in teacher ratings under Ohio's evolving system, variability that is also evident in other states that have implemented similar systems in recent years.

Collecting and analyzing data on the outcomes of the initial stages of implementation of OTES and OPES, as well as the perceptions of teachers and principals regarding the new systems is critical in order to 1) provide information for system improvement; and 2) gauge changes in teacher and principal perceptions of the system over time. This executive summary presents an overview of the research questions that focused the overall study and provides a brief synopsis of the findings that focus on OTES. Detailed analyses of both OTES and OPES are contained in the full technical report.

Snapshot of Year 1 OTES Implementation (2013-14)

In 2013-14, the first year of full implementation, OTES was implemented by approximately 77 percent of Ohio's school districts and over 200 community schools, plus a number of Joint Vocational School Districts and Educational Service Centers.

- 86,600 Ohio teachers received an OTES rating for 2013-2014
- OTES is comprised of two "sides." Each side makes up 50% of the teacher's Final Summative Rating. One side of OTES is the *Teacher Performance on Standards*. This is assessed by an administrator through classroom observations. The other side of OTES is the *Student*

Growth Measures (SGM) score that a teacher receives based on his or her students' performance on some type of assessment(s).

For the Student Growth Measure (SGM) side of OTES, Ohio's teachers are categorized by the types of assessments or tests that are available for his or her subject(s) and grade(s) taught. Category A1 teachers receive Value-Added data for all subjects and grades taught; Category A2 teachers teach some grades/subjects with Value-Added data as well as other grades/subjects for which they do not get Value-Added data; Category B teachers do not get Value-Added data but have testing data from vendor-developed assessments that have been approved by the Ohio Department of Education; Category C teachers teach grades/subjects with no available Value-Added data or growth data derived from approved Vendor Assessments. These teachers use locally-developed measures such as Student Learning Objectives (SLOs) or the Shared Attribution of Value-Added data. Note that, for 2013-14, Category A1 teachers' SGM ratings could include local measures as well as Value-Added data.

In 2013-14:

- Approximately 6 percent, or 5,144 Ohio teachers participating in OTES were Category A1 teachers
- Approximately 14 percent, or 11,865 teachers were Category A2 teachers
- Approximately 14 percent, or 12,120 teachers were Category B teachers
- Approximately 66 percent, or 57,531 teachers were Category C teachers

Research Questions and Summary of Findings

What is the distribution of summative teacher ratings across the "Accomplished", "Skilled", "Developing", and "Ineffective" categories?

The percentage of teachers in each of the four rating categories for the *Performance on Standards* portion is indicated below. Note that fewer than one percent of teachers were rated as Ineffective on the *Performance on Standards* side of OTES.

Table A. Percent of Teachers in Each of the Performance on Standards Rating Categories

Ineffective	Developing	Skilled	Accomplished
0.4%	6%	69.7%	23.9%

The percentage of teachers in each of the three rating categories for the *Student Growth Measures (SGM)* portion is indicated below for all teachers and for teachers by their SGM type. While the distribution of *Performance on Standards* ratings was similar for teachers across all SGM types (Cat. A1, A2, B, C), the distribution of *SGM* ratings varied by SGM type.

Table B. Percent of Teachers in Each of the SGM Rating Categories by Type of SGM

	Below	Expected	Above
All Teachers	6.8%	44.2%	49%
Cat. A1	10.9%	57.2%	31.9%
Cat. A2	11.5%	46.9%	41.6%
Cat. B	5.7%	55.6%	38.7%
Cat. C	5.7%	39.9%	54.4%

In order to receive the lowest ***Final Summative Rating*** of Ineffective, a teacher had to be rated as either *Ineffective* or *Developing* on the *Performance on Standards* side and *Below* on the *SGM* side of OTES. An overall breakdown of the 2013-14 Final Summative Ratings for teachers is as follows:

Table C. Teacher (OTES) Final Summative Ratings (2013-14)

Final Summative Rating	% of Teachers in Each Category
Accomplished	46%
Skilled	44%
Developing	9%
Ineffective	1%

Breaking the OTES data out by SGM Category indicates that Category C teachers were more likely to receive higher Final Summative Ratings than teachers in Categories A and B:

Table D. Teacher (OTES) Final Summative Ratings by Student Growth Measure Type (2013-14)

Rating	Cat. A1	Cat. A2	Cat. B	Cat. C
Accomplished	31%	41%	37%	50%
Skilled	56%	46%	53%	41%
Developing	12%	13%	9%	8%
Ineffective	1%	1%	1%	1%

How do teacher-developed Student Learning Objective (SLO) measures correlate with (a) Value-Added, (b) Approved Vendor Assessment growth measures, and (c) Shared Attribution measures? Do the SLO data look different for teachers assigned to those subjects (Art, Music) that are more performance based, as compared to those teachers assigned to core academic subjects?

In general, teachers with SLOs and/or Shared Attribution were rated higher on the SGM portion of OTES than teachers with Value-Added or approved Vendor Assessment data comprising their SGM rating.

Teachers in non-core subjects had SGM and Final Summative Ratings that were somewhat higher than teachers in core academic subjects.

What are teacher perceptions of the new evaluation systems?

Teachers and principals expressed concerns about fairness regarding the use of a variety of Student Growth Measures in teacher evaluation. A chief concern was that the level of difficulty of showing growth is not perceived as equal among the various SGMs. A strong theme revealed was a perception that with an SLO, a teacher has more control in the testing outcome because he or she often develops the pre- and post-test based on the specific content taught in that classroom, whereas a teacher who has Value-Added data derived from Ohio Achievement Assessments does not know what questions will be on the test.

Are LEAs opting for Shared Attribution measures for teachers and/or principals (for example, building or district Value-Added)?

Slightly more than one-third of LEAs participating in OTES used Shared Attribution in 2013-14; Shared Attribution was used to determine all or part of the SGM rating for 31 percent (26,985) of all Ohio teachers participating in OTES. The decision to use Shared Attribution was somewhat related to the district's overall Value-Added letter grade on the District Report Cards. For example, 48 percent of districts with an A in Value-Added and 61 percent of districts with a B in Value-Added for 2013-14 opted for shared attribution as one of the SGMs, while a lower percentage of districts receiving a D or F in Value-Added opted for shared attribution. The weight assigned to Shared Attribution ranged from 10 percent (for approx. 7,000 teachers) to 50 percent (for approx. 11,000 teachers).

How would the distribution of summative teacher ratings change if the relevant weights on Value-Added and Approved Vendor Assessments were recalibrated along the 0-50 percent continuum?

The eTPES data for 2013-14 indicate that the weights selected to comprise the 50 percent SGM portion of OTES were consequential. For example, 21 percent of teachers with Value-Added making up all of their SGM rating are rated Accomplished as compared to 46 percent of their colleagues with Value-Added being used in conjunction with local measures. A similar gap is evident for Approved Vendor Assessments; only 19 percent of teachers with Approved Vendor Assessments being the only SGM component are rated Accomplished versus 47 percent of teachers with Approved Vendor Assessments being used along with other local measures.

How are SGM ratings distributed across such factors as district-type, socioeconomic/demographic profiles, etc. of LEAs?

In general, the data show notable differences in the distribution of teacher ratings in the Big 8 Urban districts (Akron, Canton, Cincinnati, Cleveland, Columbus, Dayton, Toledo, Youngstown) versus other districts. In brief, almost 19 percent of teachers in the Big 8 districts were rated as either Ineffective or Developing, compared to only 8 percent of their colleagues in other districts. These gaps are even more pronounced for administrators. Teachers and administrators in Community Schools were also more likely to be rated either Ineffective or Developing than their colleagues in other LEAs. There is also evidence of a positive association between student poverty and higher proportions of Ineffective or Developing educators.

Do the SGM-based ratings of teachers correlate highly with teacher Performance on Standards ratings?

While perfect congruence between the *Performance on Standards* side and the *SGM* side of OTES and OPES would not be expected, the 2013-14 data indicate more congruence or alignment between the two sides for teachers with Value-Added and Vendor Assessment data than for teachers with SLOs. For example, 40 percent of teachers in Category C who received one of the two lowest ratings (Ineffective or Developing) on the Performance on Standards rubric, received the highest rating on the SGM portion of OTES; only 17 percent of Category A1 teachers who received an Ineffective or Developing rating on the *Performance on Standards* side received the highest rating on the SGM portion. In other words, teachers with low ratings on the *Performance on Standards* side of OTES were "better off" overall if they were Category C teachers, as their SGM ratings were more likely to be high.

General Conclusions and Next Steps

- In the first year of OTES implementation, relatively few teachers were rated in the lowest category of "ineffective." However, the data indicate more variability in teacher ratings across the other three categories than has been the case with teacher effectiveness ratings in the old systems based solely on classroom observation. This variability was most pronounced in Ohio's largest eight urban districts that showed a higher percentage of teachers rated in the two lowest categories than the rest of the state. In general, high poverty districts, both urban and rural, showed higher percentages of teachers rated as "Ineffective" or "Developing" than other districts.

- Teachers in the thirteen pilot districts who participated in the qualitative portion of this study see fairness issues with OTES---partly because of the difference in the types of Student Growth Measures (e.g. Value-Added based on a State assessment vs. SLOs where the teacher develops the assessment). Many teachers also articulated an inaccurate or incomplete understanding of how Value-Added is calculated.

- Category C teachers were more likely to receive higher SGM and Final Summative ratings than Category A or B teachers. Approximately two-thirds of these Category C teachers used

SLOs exclusively to determine their SGM rating. It is unlikely that higher ratings for Category C teachers is a consequence of the relative effectiveness of teachers in subjects or grades where there is no current State or Approved Vendor Assessment compared to teachers in tested grades and subjects (including State assessments or vendor assessments). Modifications should be made to state and local policy regarding SLOs so that they better align the rigor of other Student Growth Measures used across grades and subjects. Ongoing research on OTES and OPES will look for evidence of these modifications.

Since Ohio will implement additional changes to OTES and OPES starting in the 2014-15 academic year, a careful analysis of the 2014-15 OTES and OPES data, as well as additional field work to gauge teacher and principal experiences and perceptions as the system matures and changes, is critical to understanding whether the system is functioning as intended in order to develop a strong teacher and principal workforce to increase the academic achievement and success of Ohio's students. Perhaps the most critical question across all states implementing the new systems is how teacher and principal perceptions change over time and how these perceptions shape changes in practice. For example, do the new systems foster unintended consequences such as reduced collaboration among teachers or demoralizing the teaching workforce, or do they provide data that focuses teachers' efforts to improve student achievement. In a recent special issue of Educational Researcher, Doug Harris and Carolyn Herrington (2015) frame the essential questions for those implementing and examining the new teacher and principal evaluation systems in Ohio and other states:

> *The main underlying theory of these policies is that teacher accountability will motivate teachers to work harder and smarter and help attract and retain only those who are successful. Does this happen in practice? Does the increased scrutiny lead educators to work harder and smarter in helping their students? Does the recognition that comes with high performance ratings encourage a stronger focus on the student outcomes on which the educator performance measures are based? Are teachers more likely to demand and seek out instructional leadership from their principals, peers, coaches, and other sources?...Do these systems increase cohesion around common goals and expectations at the school level? (p. 72)*

This study begins the work of answering these questions by detailing the results of the first year of implementation of Ohio's new systems of teacher and principal evaluation.

I. INTRODUCTION

Until the advent of No Child Left Behind (NCLB) and more recently Race to the Top (RttT), calls for holding schools accountable for student achievement had waxed and waned periodically but the spotlight, when it shone, had invariably fallen on schools rather than on teachers (Lavigne and Good, 2014). However, fairly recent research indicates that the teacher is the largest school-level influencer of student learning, and that the effectiveness of teachers varies quite a bit within individual schools (Johnson, 2015; McCaffrey, Koretz, Lockwood, &Hamilton, 2004). Both RttT and The Teacher Incentive Fund (TIF) ratcheted up the stakes for individual educator performance by tying teacher and principal professional development explicitly to weaknesses identified by each individual's evaluation, and with graduated consequences for underperforming educators (Thorn and Harris, 2013). This trend received further impetus once states could apply for, and receive if approved, waivers from some provisions of the Elementary and Secondary Education Act (ESEA) provided they put in place educator evaluation systems that identified both effective teachers as well as those who could benefit from additional support to improve their instructional practice (Pennington 2014).

This recent drive for educator accountability prompts the obvious question: How should teachers be evaluated? The authors of The Widget Effect (Weisberg et al. 2009) have emphasized weaknesses in traditional classroom observations as the sole means of evaluating teachers. This report illuminated the fact that, in the traditional teacher evaluation systems that relied almost exclusively on classroom observation, there was very little variability in the ratings of teachers and most teachers received very good ratings. Proponents of a more comprehensive teacher evaluation method, as documented in the Measures of Effective Teaching (MET) report (see Kain and Staiger 2012; Mihaly, McCaffrey, Staiger, & Lockwood, 2013; Harris 2013), support the use of composite indicators that include (but are not limited to) student academic growth measured by state assessments, classroom observations by trained observers, student learning outcomes, and student surveys[1].

By the close of 2013 some 40 states (and the District of Columbia Public Schools) required that objective measures of student learning inform a teacher's evaluation, with 35 of these states requiring that student achievement be a significant (if not the most significant) component of a teacher's evaluation (State of the States, 2013). Each state's evaluation system is fairly new, continually evolving, and in many instances yet to go into force. In Ohio, for example, almost all local education agencies (LEAs) were required to have in place no later than July 1, 2013 a state-approved standards-based educator evaluation policy. In 2012 Ohio also provided funds and technical training to 13 LEAs to pilot the development and implementation of Student Growth Measures (SGMs) – the student learning-based measures of teacher effectiveness that were to comprise 50 percent of an educator's evaluation. The academic year 2013-14 was the first year of large-scale implementation of Ohio's new teacher and principal evaluation system.

[1] See Bill & Melinda Gates Foundation (2013); Hull (2013); State of the States (2013).

In this study we present early evidence from Ohio's educator evaluation system in force statewide for the 2013-14 school year. In particular, we report the findings of a two-year mixed-methods study motivated by two overall questions: (a) What are some of the leading perceptions and concerns of Ohio's educators regarding the new educator accountability framework established by Ohio?; and (b) What does the distribution of educator ratings look like, both in the aggregate and when disaggregated by specific attributes of the LEAs?

In the sections that follow we begin with a brief overview of Ohio's teacher and principal evaluation systems in place for the academic year 2013-14. Thereafter, we describe our approach to gathering both the qualitative and the quantitative data used in this study, as well the light shed by each of these datasets on particular research questions. Where necessary we disaggregate the qualitative/quantitative data by particular criteria (for example, by specific LEA attributes such as District typology, whether it is a Community School or District, and so on). We conclude by highlighting key findings and their implications for educator accountability policy.

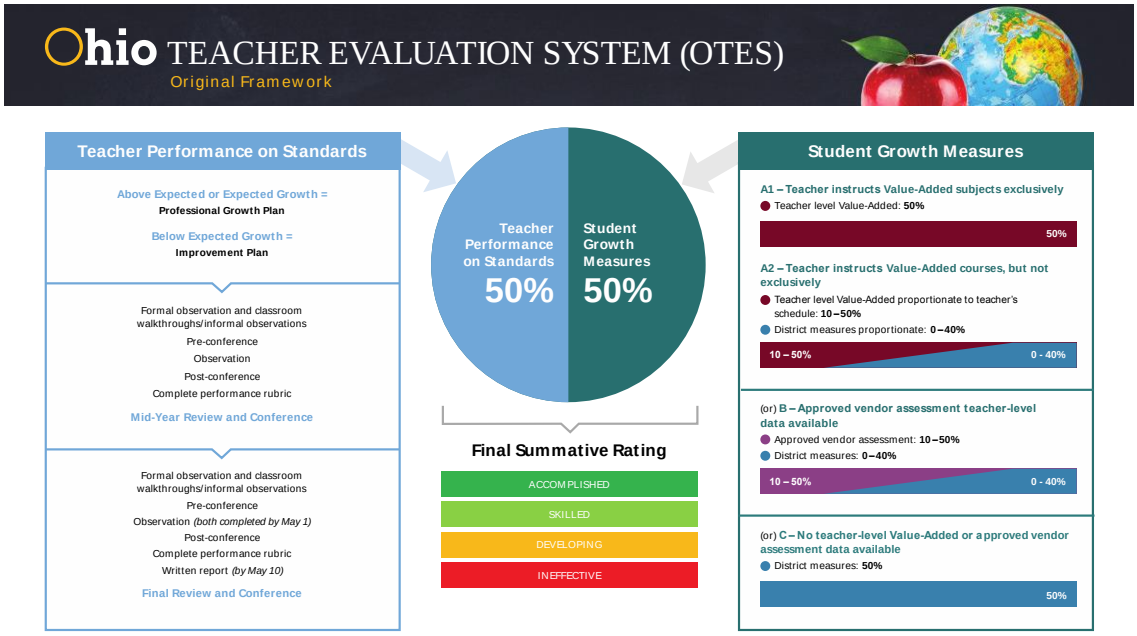
Ohio's Educator Evaluation System

Today, educator evaluation in Ohio is driven by the Ohio Teacher Evaluation System (OTES) and the Ohio Principal Evaluation System (OPES). The evaluation system originated in the Ohio General Assembly, Amended Substitute House Bill Number 59, which was signed into law in June 2013. Both systems rely on two key evaluation components, each weighted at 50 percent: (i) A rating of teacher (principal) performance on state-specified performance standards, and (ii) A summative rating based on student academic growth. This approach of multiple inputs to educator accountability has pretty much come to be accepted as the safest approach (see Mihaly et al. 2013; Harris 2013).

A Brief Overview of the Ohio Teacher Evaluation System (OTES)

Figure 1 summarily describes the key elements of the Ohio Teacher Evaluation System as designed for the 2013-14 academic year.

Figure 1. A Schematic of the OTES Framework



Below is a brief description of the two components that feed into the overall rating.

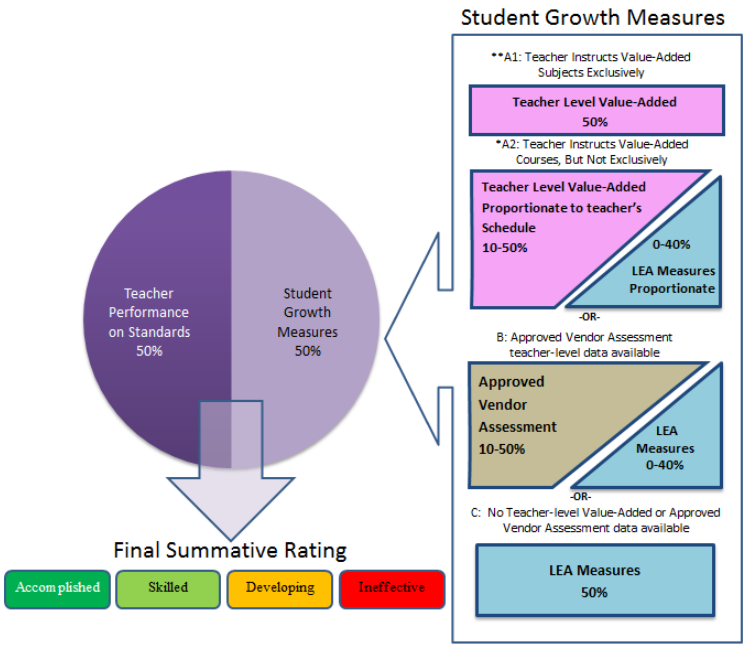
(A) Teacher Performance on Standards
Ohio's performance standards, established by the state's Educator Standards Board, span seven components that include

(i) Students -- Teachers understand student learning and development and respect the diversity of the students they teach;
(ii) Content -- Teachers know and understand the content area for which they have instructional responsibility;
(iii) Assessment -- Teachers understand and use varied assessments to inform instruction, evaluate and ensure student learning;
(iv) Instruction -- Teachers plan and deliver effective instruction that advances the learning of each individual student;
(v) Learning Environment -- Teachers create learning environments that promote high levels of learning and achievement for all students;
(vi) Collaboration and Communication -- Teachers collaborate and communicate with other educators, administrators, students and parents and the community to support student learning; and
(vii) Professional Responsibility and Growth -- Teachers assume responsibility for professional growth, performance, and involvement as an individual and as a member of a learning community.

(B) Student Growth Measures (SGMs)
SGMs can be determined by a combination of state-approved Vendor Assessment results, Value-Added results from standardized testing, and measures developed by the Local Education Agency.

LEA measures include, but are not limited to, student learning objectives (SLOs). Figure 2 depicts how the various SGM components mesh to populate a teacher's SGM-based rating for the 2013-14 academic year.

Figure 2. A Schematic Description of OTES' SGM Components



We next describe the three SGM components – Value-Added, approved Vendor Assessments, and LEA measures.

1. Value-Added
   If available, teachers must include Value-Added data (calculated by SAS, Inc. through their EVAAS method) in the student growth measure. Currently, EVAAS calculates teacher-level Value-Added based on Ohio Achievement Assessments (OAA) results for grades 4 through 8 in reading and mathematics. Starting with the effective date of HB 555 (3/22/13), the majority (of the SGM 50%) of the student academic growth factor of the teacher's evaluation must be based upon Value-Added data. Beginning July 1, 2014 and before more recent changes have allowed for an alternative framework, for those teachers who have only Value-Added assessed subjects, the entire student academic growth factor will be based upon Value Added (all of the SGM 50%). Because the teacher-level Value-Added ratings for any academic year are not available until the following academic year, the Value-Added scores from the previous year will be brought into the teacher rating in the evaluation year. For teachers whose entire teaching schedule is not comprised of reading and/or mathematics grades four through eight, the local education agency may also use local student growth measures in percentages commensurate to the teaching schedule.

2. Approved Vendor Assessments
   If Value-Added data is not available, districts or schools can use other assessments provided by national testing vendors and approved for use in Ohio[2]. The vendors on the approved list provided evidence that the assessments meet these fundamental requirements for measuring student growth:

   a. Be highly correlated with curricular objectives;
   b. Have enough "stretch" to measure the growth of both low-and high-achieving students;
   c. Meet appropriate standards of test reliability; and
   d. Have specifics on relating assessment growth measures to the established Teacher Effectiveness scale.

3. Locally Determined Measures
   For subjects where Value-Added or Vendor Assessments are not an option, districts or schools are encouraged to establish a process to create locally determined or developed measures, including Student Learning Objectives, to measure student progress. Types of locally determined measures include:

   a. Student Learning Objectives (SLOs)

   A Student Learning Objective is a measurable, long-term academic growth target that a teacher sets at the beginning of the year or semester for all students or for subgroups of students. SLOs have recently become a popular tool for untested subjects and grades (see Gill et al. 2013 for a summary studies investigating SLOs). SLOs demonstrate a teacher's impact on student learning within a given interval of instruction based upon baseline data gathered at the beginning of the course. Each SLO must be approved as per a state-specified process and includes:

   - The baseline and trend data;
   - The student population or sample included in the objective;
   - The period of time covered by the SLO;
   - The standards the SLO will align with;
   - The assessments that will be used to measure student progress;
   - The expected student growth; and
   - The rationale for the expected student growth.

   b. Shared Attribution

   Shared Attribution involves the concept of "we all contribute to student learning," whether that be a classroom teacher and an intervention specialist, or all teachers at a grade level or in a building, e.g., self-contained classroom teachers and the "specials" teachers (e.g., art, music, physical education) who teach the same students. There are circumstances where it might be reasonable to do this (e.g., facilitates collaboration; acknowledges every teacher's

[2] Vendor Assessments are instruments that are approved by the Ohio Department of Education to function as the Student Growth Measures component of an educator's OTES rating, in the absence of value-added data from Ohio Achievement Assessments, which are administered in grades 3, 4, 6 and 7 for reading and math, and reading, math and science for grades 5 and 8. As of September 2013, there were 28 approved Vendor Assessments that are available for purchase by districts. See Porter and Mauck (2013) for an overview and analysis of Ohio's approved Vendor Assessments for the 2012-2013 school-year.

contributions; eases the identification of multiple measures). Some examples of Shared Attribution include school- or district-level Value-Added data, a composite value-added estimate calculated for school-level teams (by content area, for example), or school\district-level SLO.

c. Vendor assessments not on the approved list may be used as a local measure.

A Brief Overview of the Ohio Principal Evaluation System (OPES)

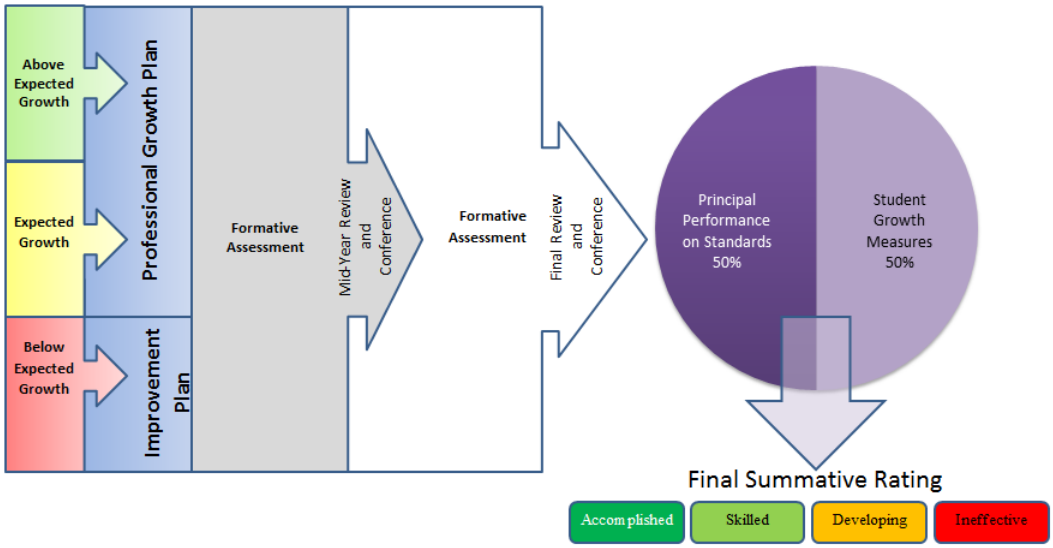
Mirroring the system for teachers, Ohio's principals will also be evaluated on the basis of (i) their Performance on Standards (50%), and (ii) Student Growth Measures (50%).

(A) Principal Performance on Standards
A principal's rating on state-specified performance standards is determined by using a rating rubric consisting of indicators based on the elements and standards from the Ohio Standards for Principals. The evaluation process requires the evaluator to use evidence gathered in a variety of avenues (professional growth/goal-setting, formative assessments, observations, communication and professionalism, conferences) to determine a principal performance rating. Under this approach Ohio's effective principals:

- Help create a shared vision and clear goals for their schools and ensure continuous progress toward achieving the goals;
- Support the implementation of high-quality standards based instruction that results in high levels of achievement for all students;
- Allocate resources and manage school operations in order to ensure a safe and productive learning environment;
- Establish and sustain collaborative learning and shared leadership to promote learning and achievement of all students; and
- Engage parents and community members in the educational process and create an environment where community resources support student learning, achievement and wellbeing.

Figure 3 shows how performance on standards feed into the overall rating for principals.

Figure 3. A Schematic Description of OPES' Performance on Standards Components



(B) Student Growth Measures

Again, much like the system developed for teachers, Ohio principals' student growth measures components span value-added, approved vendor assessments, and locally determined measures (which include Student Learning Objectives, Shared Attribution, and Vendor Assessments).

1. Value-Added

If school-level Value-Added data are available for grades 4 through 8 in Reading and Mathematics, then these data must be used.

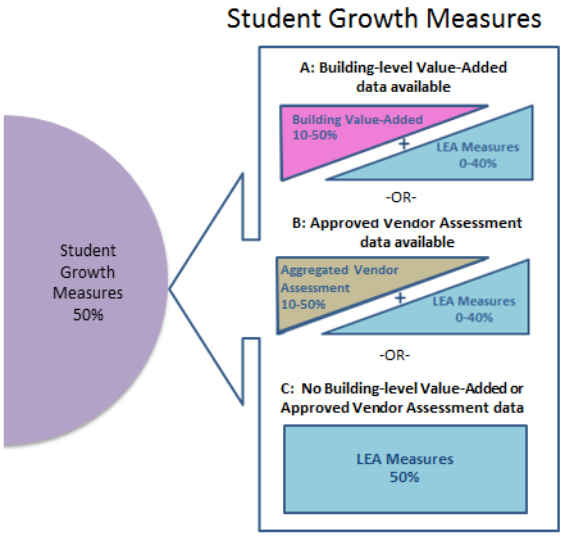
2. Approved Vendor Assessments

These are the same set of assessments that apply to teachers, albeit here a composite growth measure at the school-level is calculated. This composite measure may well be a simple aggregate of approved vendor assessment scores.

3. LEA-determined measures

These measures may include an aggregate of all teachers' student growth scores, school-based student learning objectives that tap school-wide goals and priorities, student achievement trends, locally\regionally used subject-specific test results, progress on school improvement plans, student course-taking patterns (for example, more students taking advanced courses, etc.), district-based SLOs, or district Value-Added. Figure 4 captures the SGM side of the accountability equation for principals.

Figure 4. A Schematic Description of OPES' SGM Components



In the case of both teachers and principals, eTPES takes the evaluation components input by the evaluator and calculates, based on the weights set by the local school board on the SGM elements, the Final Summative Rating of teachers and principals. Figure 5 maps the summary lookup table that shows how the two halves of the evaluation system combine for the 2013-14 academic year. The vertical axis of the lookup table represents student growth measures, and the horizontal axis on the table represents teacher or principal performance on standards.

Figure 5. Combining Teacher\Principal Performance and SGMs for 2013-14

Teacher (or Principal) Performance on Standards

SGM Rating	4	3	2	1
Above	Accomplished	Accomplished	Skilled	Developing
Expected	Skilled	Skilled	Developing	Developing
Below	Developing	Developing	Ineffective	Ineffective

It is important to note that the Ohio General Assembly, in House Bill 362, has enacted modifications to OTES that went into effect beginning in the 2014-15 academic year, including the elimination of the summary lookup table (Figure 5). The modifications include less frequent evaluations of teachers receiving skilled or accomplished ratings, an option for schools and districts to use other metrics to evaluate teachers, including student surveys and peer reviews, as well as a change in the methodology for calculating the Final Summative Rating. The Final Summative Rating will now be calculated using a formula-based approach and a 600 point scale. The Student Growth Measures

component will now have five categories instead of three categories. These changes will likely impact the distribution of OTES ratings at some level, but the magnitude and direction of the changes are unknown at this point.

II. METHODOLOGY

In this study we employ a mixed methods design that includes the analysis of statewide data from Ohio's Electronic Teacher and Principal Evaluation System (eTPES) in its first year of full implementation (2013-14), along with qualitative data collected from focus groups, surveys and interviews of teachers and administrators in 13 Ohio Local Education Agencies (LEAs) that piloted all or part of the new teacher and principal evaluation systems in the 2012-13 academic year. During 2012-13 Ohio provided funds and technical training to 13 LEAs to pilot the development of Student Growth Measures (see Table 1 for a brief statistical profile of these LEAs)[3].

Table 1. Descriptive Statistics of the 13 LEAs Piloting Student Growth Measures in 2012-13

Indicator	Mean	Standard Deviation	Minimum	Maximum
Enrollment	3178.00	3226.92	679.00	10639.00
Limited English Proficient (%)	0.83	1.58	0.00	5.54
Economic Disadvantage (%)	43.67	20.67	7.00	71.93
Gifted (%)	14.00	5.01	5.01	21.84
Disabled (%)	14.53	3.92	7.57	24.01
Minority (%)	9.31	8.05	1.62	28.26

Qualitative data from the 13 pilot LEAs were gathered via 25 focus groups with teachers, 12 individual or group interviews with administrators (mostly principals) and a teacher survey. In most LEAs, teacher focus groups were split by the type of SGM used in the teacher's evaluation, as defined by Ohio law. Teachers who received teacher-level Value-Added data ("Category A" teachers) were interviewed together as a group and teachers who use ODE-approved Vendor Assessments or use locally-developed measures such as SLOs ("Category B & C" teachers) were interviewed separately as a group. In the few cases where teachers were not separated based on their SGM category, teacher category was identified for individual respondents, allowing for proper comparison of responses across categories. The rationale for separating Category A or value-added teachers from Category B and C teachers was based on preliminary observations and discussions with teachers and principals, which indicated that teachers with Value-Added data perceived themselves as different or at a disadvantage in terms of SGM performance on OTES relative to teachers in the other SGM categories.

The interviews and focus group discussions were conducted during the 2013-14 school year. Transcripts of these interviews and focus groups were analyzed using NVivo qualitative analysis software. Additionally, each interview or focus group was coded by school district typology (as

[3] These 13 LEAs represent both types of rural, and small town districts, as well as Urban high student poverty with average student population size and Suburban very low student poverty and low student population size. Nine LEAs were in Race to the Top. All SGM categories.

defined by the Ohio Department of Education), thus allowing for comparison of responses by district typology. The teacher survey was administered online and was deployed to all teachers in the 13 pilot LEAs in late February through early April of 2014. In all, 603 teacher surveys were completed and the resulting data are included in the analysis we report in this study. The survey included both closed and open-response questions. The overall response rate from the 13 districts was 22 percent. The response rate varied by district however, from a low of 10 percent to a high of 51 percent. The response rate from nine of the thirteen districts was greater than 20 percent.
For the quantitative portion of the study we relied upon the 2013-14 eTPES dataset, which included all OTES data for 86,660 teachers and all OPES data for 5,284 principals and assistant principals. OTES and OPES data for each teacher and principal included the SGM ratings, the Performance on Standards ratings, and the weights assigned to each SGM category. The research team then added demographic variables such as district typology. Data were available for the majority of Ohio's school districts, as well as most community schools participating in Race to the Top, joint vocational school districts and career-technical planning districts, education service centers, and two STEM schools. Some Ohio LEAs did not participate in the first full year of OTES/OPES implementation for reasons such as local master contract prohibitions. Community schools not participating in Race to the Top were not required to participate.

III. THEMES EMERGING FROM FOCUS GROUPS AND THE SURVEY

Given the novelty of Student Growth Measures as a significant component of the teacher evaluation system, in the focus groups and teacher survey we probed teachers' perceptions and opinions about Student Growth Measures being used for teacher evaluation purposes.

The Value of Using Student Growth Measures for Evaluating Teachers

For the most part teachers acknowledged that there was value in being held accountable for the academic growth of their students. They also saw value in assessing student progress, in using student test data formatively, and felt that measuring student growth via a pretest-posttest assessment was better than just focusing on a proficiency measure.

- "I think the expectations are clear, it's a lot more feedback for the teacher than our typical checklist of, you know, "you did this, and you did this, and boom, boom, boom, and everything is wonderful." It provides the teacher with more thorough feedback." Category A Teacher

- "It shows the effectiveness of a teacher and useful data to adjust your teaching." Category A Teacher

While some teachers talked about the value in moving to an evaluation system that includes student growth as one of the measures of effectiveness, many more teachers in this first year of implementation expressed concerns about basing a teacher's evaluation on her/his students' test performance, and this was most evident from the survey responses; when asked to respond to the following statement: "It is valuable to use Student Growth Measures in teacher evaluation" some 54 percent of survey respondents either strongly disagreed or disagreed while only 24 percent of respondents either agreed or strongly agreed or agreed with this statement (see Table 2).

Table 2. Teachers' Perceptions Regarding Value of Student Growth Measures in Teacher Evaluation

Question: "It is valuable to use Student Growth Measures in teacher evaluation."

Response	Number	Percent
Strongly Agree	15	2.5
Agree	129	21.5
Neutral	133	22.2
Disagree	158	26.4
Strongly Disagree	164	27.4
Total	599	100.0

These response patterns held steady regardless of how the data were disaggregated, be the disaggregation by district typology, district or building rating on the Ohio School Report Card,

OTES category (i.e., Category A versus Category B/C teachers), etc. Some of the opinions expressed in the focus groups provide more insight on these teachers' perspectives:

- "It's a good idea to do, I think there is a lot of debate over how to do that effectively for the teacher and then obviously for the benefit of the student." - Category B/C Teacher

- "You know, it is good to track student progress. To link that to a teacher's success may not be the best strategy or the best method to evaluate a teacher because there are so many outside factors that we cannot control. But, those are not accounted for, at least we don't feel like they're accounted for in our evaluation, when it comes to like students actually achieving success." - Category A Teacher

Fairness of Multiple Measures Comprising SGM

Teachers and administrators expressed concerns regarding the fairness of a system such as OTES wherein a variety of assessments were being treated equally as components of SGM. Administrators in over half of the pilot districts raised equity concerns that often focused on the variety and perceived inequality of the different types of measures.

- "About the only thing I've got to say about Value-Added is the fact that Value-Added versus SLOs is not an equal measure." - Administrator

- "The equity of this whole thing is making us a little mad too because the Value-Added is a lot different than the teachers doing their own and it's going to create a rift". - Administrator

The majority of teachers (76%) surveyed indicated that they did not feel using a variety of assessments for teachers' evaluations made for a fair evaluation system. What was noteworthy here was that once again the response pattern remained unchanged even when the data were disaggregated by classroom teacher and intervention specialist, teacher SGM category, (category A, B, or C), district Value-Added letter grade, district poverty status (high vs. non-high poverty), and number of years a teacher reports they received Value-Added data.

Teachers also noted that they feel substantial pressure regarding OTES, including the student test results or showing student growth and that the pressure is negatively affecting teacher morale, willingness to accept student teachers, and some teachers noted an overall negative impact on school culture/community.

Perceptions of Value-Added

Although there were general similarities in perceptions of Value-Added across the teacher discussion groups, teachers who did not receive Value-Added data (i.e., Category B & C teachers) were somewhat less negative about Value-Added than their Category A teachers (who also reported more pressure). Many teachers also expressed concern about not understanding how Value-Added is calculated.

- "I guess my point is too, I just feel like with me, your Value-Added is more like high-stakes. We are carrying the load on the report card especially and…and especially when you are in a small district like at 5th grade science doesn't pass and it is like boom, 'We know it's you.'"- Category A Teacher

Category A teachers also talked more about a perceived challenge of showing growth from one year to the next amongst the highest and lowest performing students.

- "I have the gifted kids on my team and we worry about showing a year's worth of growth and so far, at the initial data that we have looked at, it is turning out to be very challenging." - Category A Teacher

Assessing Non-Core Subjects

Teachers in non-core subjects such as art, music and computer science reported additional challenges of assessing student progress. Teachers in approximately a third of the discussions noted it was difficult to assess non-core subjects.

- "In my class in computers, I've never given a paper-pencil test. All of my tests, I teach Photoshop as one of my programs and I was instructed that it would be a better measure to give a paper-pencil multiple choice test versus giving a project on the computer and I disagree with that method. I can see them working on the computer. I can see them problem solving, finding different things on the computer versus just a piece of paper-pencil test. And so what I see on a test and what I see in my classroom is very different. And the fact that we don't have, being an elective and computers, I don't have a test to go to. I make my test. So then, I've spoken to my department about this that, 'How valid, really, is that test?'"- Category C Teacher

Concerns about Testing

Many teachers expressed concern in the focus groups about too much testing and testing children who are too young. Teachers in all categories expressed concerns about testing methods. Category B & C teachers talked more often about the challenges of testing very young students such as Kindergarten, first and second grades and using computers to do so. Whereas, category A teachers talked more often about the changes in the response patterns and formats, for example moving from a three-choice response pattern to a four-choice response pattern on the OAAs.

Locus of Control

Teachers' lack of control over numerous factors that they believe impact student performance was a consistent theme that emerged from the teacher focus groups and the survey data, and particularly in the context of linking teacher evaluations to students' test performance. Across the board, regardless of teacher SGM category (A, B or C), many teachers expressed substantial concerns about a host of external factors that can impact student academic performance. The factors discussed ranged from students' home lives, poverty, students' or their parents' health issues, students coming to school tired or hungry, unexpected events on or just before scheduled

testing, and students simply lacking motivation. Teachers indicated concern that these factors could negatively affect student test performance and were clearly beyond a teacher's control, yet would affect the teacher's evaluation.

Perceptions and Awareness of Shared Attribution

At the time of the survey, nearly a third of the responding teachers indicated they did not know what Shared Attribution is. Only 14 percent of teachers reported that Shared Attribution was used as a portion of their SGM in the 2013/2014 school year. Shared Attribution was also discussed in the focus groups. Oftentimes, similar to the survey responses, teachers would have little understanding of the phrase "Shared Attribution." Teachers expressed some concerns about the practice. A few teachers said that it could inspire teamwork in teachers while other teachers believe that it will spark animosity among teachers. Many teachers expressed that they did not like the idea of other teachers taking some of their scores for shared attribution and that some of their score can depend on other teachers.

Perceptions of the Performance on Standards Rubric

Discussions in the administrators' focus groups revealed no lack of understanding among administrators regarding the performance rubric. Administrators discussed pre-conferences, walk-throughs, number of required observations and teacher evidence. They often noted value in the performance standard rubric, indicating that it led to valuable conversations, and allowed for more clarity for teachers about performance expectations. Some found a few sections to be vague, while others reported redundancies. They also frequently indicated that it was difficult to allocate the amount of time required to evaluate every teacher. Indeed, this was a common theme not only in the administrator focus groups but also in the teacher focus groups.

Teachers recognized that the new rubric allowed for more in-depth discussion and feedback from principals, but they also expressed frustration with a perception that they could never receive an "Accomplished" rating and teachers generally did not see a correlation between the two sides of the rubric.

- "It provides the teacher with more thorough feedback." Category A Teacher

- "I think it poses some problems going forward because if, on the rubric you score well and you consistently score well on that side, but the student growth measures are based on a yearly thing and it's a class every year so it shows some consistency problems. If you're doing well in the classroom and your teaching is sound yet you're jumping all over on the other side because your student growth measures are changing." -Category A Teacher

Concerns Voiced About Special Education Students

A general theme that emerged from the focus groups was concern about using special education students' scores in teacher evaluation. Some teachers went as far as questioning whether teachers will want to continue in the special education field because of this challenge.
Teachers in approximately a quarter of the focus groups emphasized that students make gains such as improved organizational skills, increased confidence or becoming more willing to work with a group that do not get measured via a test. This led to frustration with the emphasis on testing versus the growth of the "whole child."

Too Much Change Too Often

A general theme that emerged from the focus groups is teacher frustration with the amount of change mostly regarding their evaluation system. They have concerns about putting effort into their evaluation if the format changes during the year and a general inability to keep up with requirements when they change often.

- "And then between last year and this year, even during this year's process, halfway through, things change. So is it going to change halfway through this year, is it going to change next year, we think we got it right. All these things, it gets really frustrating for everybody that you worked real hard to do something and then all of the sudden it is 'oh no we are doing it this way now.' And that really, I think it upsets teachers, they are less supportive, less open-minded to it because it's probably just going to change again." Category-B/C Teacher

- "A lot of people in our building have said it is a lot of change very quickly and they still don't really—I mean yes, they've turned in their SLO goals and have been approved, but they still have a lot of questions about it because they followed the directions we gave them but so much has changed so quickly, they can't keep up with everything." – Category B/C Teacher

Summary of Major Themes Emerging from the Qualitative Data

The most common themes that emerged from the teacher and administrator focus groups/interviews and the teacher survey were those of fairness and accuracy of OTES. Fairness issues included lack of teacher control over external factors influencing student academic performance, lack of similarity among the multiple SGMs that are used in OTES, and so much of a teacher's evaluation riding on one day's test performance by his/her students. Accuracy issues included a perceived lack of congruence between the two sides of OTES and the difference in quality/rigor of SLOs being developed across the state. Teachers and administrators were also frustrated with the multiple changes to OTES and OPES in the early implementation phases. The strongest positive theme that emerged from the data was the teachers' perceived value in the new performance rubric and the deeper, more focused conversations with their principals about their performance in the classroom.

IV. ANALYSIS OF THE ELECTRONIC TEACHER PRINCIPAL EVALUATION SYSTEM (ETPES) DATA

While on the preceding pages we have highlighted key perspectives of teachers and administrators regarding Ohio's educator evaluation system, we now turn to exploring the ratings received by educators for the 2013-14 school year. This exercise achieves two important goals. First, this first year of statewide evaluation data provides a unique opportunity to understand how not only the various components that make up OTES and OPES but also decisions and implementation at the LEA-level shape an individual's Final Summative Rating. Do, for example, the weights used for the SGM portion of the final rating matter? Are teachers in poorer\wealthier districts less\more likely to be rated Skilled or Accomplished? What about teachers in the Big 8 districts versus their colleagues in other districts in the state? Second, these data also enable us to empirically test some of the perspectives voiced by educators in our focus groups and in the survey. For example, are Value-Added teachers really at a disadvantage when compared to teachers rated on the basis of approved Vendor Assessments or SLOs? Does the Performance on Standards portion of the Final Summative Rating match up strongly or weakly with the Student Growth Measures side?

The Ratings Database Available for Analysis

Ohio utilizes the electronic Teacher and Principal Evaluation System (eTPES) -- for gathering and documenting the various components that go into an educator's evaluation to ultimately generate the overall rating. The 2013-14 eTPES data inform all analyses that follow, and we begin with a brief description of this database. In particular, the 2013-14 OTES (for 86,660 teachers) and OPES (for 5,284 principals and assistant principals) data available span close to 800 LEAs, including Ohio school districts, community schools, joint vocational school districts and career-technical planning districts, Educational Service Centers, and two STEM schools (see Table 3).

Table 3. Distribution of OTES and OPES records by LEA Type

A. OTES

LEA Type	No. of Records	Not in RttT	In RttT	No. of LEAs
Career Technical Planning District	24	24	0	1
Community School	4,748	1,040	3,078	230
Educational Service Center	1,532	1,532	0	46
Joint Vocational School District	1,981	1,981	0	29
School District	78,318	24,586	53,732	470
STEM	57	26	31	2
Total	86,660	29,189	57,471	778

RttT = Race to the Top

B. OPES

LEA Type	No. of Records	Not in RttT	In RttT	No. of LEAs
Community School	314	87	227	193
Educational Service Center	40	40	0	14
Joint Vocational School District	168	168	0	41
School District	4,760	2,088	2,672	599
STEM	2	0	2	1
Total	5,284	2,383	2,901	848

RttT = Race to the Top

Some 66 percent of the 86,660 OTES records and 55 percent of the 5,284 OPES records are from LEAs participating in Race to the Top (RttT). The records for the school districts (traditional) span all eight of the district types defined and utilized by the Ohio Department of Education.

The Distribution of Final Summative Ratings

Table 4 maps the Final Summative Rating for principals and assistant principals (OPES) and for teachers (OTES). As evident from both distributions, both teachers and principals had an almost 90 percent chance of being rated Skilled or Accomplished, and a 1 percent or lower chance of being rated Ineffective.

Table 4. Final Summative Ratings for OPES and OTES (2013-14)

Rating	OPES	Percent	OTES	Percent
Ineffective	29	0.55%	864	1.00%
Developing	592	11.20%	8,058	9.30%
Skilled	2,566	48.56%	38,145	44.02%
Accomplished	2,097	39.69%	39,593	45.69%
Total	5,284	100.00%	86,660	100.00%

Another way to examine the origin of the Final Summative Ratings for teachers is to populate the OTES lookup table (refer to Figure 5) with the percentage of teachers falling within each cell. In order to receive the lowest *Final Summative Rating* of Ineffective, a teacher had to be rated as either *Ineffective* or *Developing (ratings 1 and 2)* on the *Performance on Standards* side and *Below* on the *SGM* side of OTES. Summing the percentages in the red cells in Table 5 provides the percentage of teachers (approx. 1 percent) rated Ineffective in OTES. Summing the dark green cells provides the percentage of teachers rated Accomplished (note that percentages across Tables 4 and 5 vary slightly because, for a small percentage of teachers, Performance on Standards data were not available in eTPES).

Table 5. Percent of Teachers in each OTES Lookup Table Cell 2013-14

SGM Rating	Teacher Performance on Standards 4	3	2	1
Above	14.4%	32.4%	2.1%	0.1%
Expected	8.7%	32.3%	3.1%	0.2%
Below	0.8%	5.0%	0.8%	0.1%

Accomplished
Skilled
Developing
Ineffective

Does the Type of Student Growth Measure Matter for Final Summative Ratings?

For the Student Growth Measure (SGM) side of OTES, Ohio's teachers are categorized by the types of assessments or tests that are available for his or her subject(s) and grade(s) taught.

- Category A1 teachers' students take the state assessments (e.g. the OAA in 2013-14) and receive Value-Added data for all subjects and grades taught.
- Category A2 teachers teach some grades/subjects with available state assessments, so they get Value-Added data. They also teach other grades/subjects for which they do not get Value-Added data, so Value-Added data does not make up their entire SGM rating.
- Category B teachers do not get Value-Added data but have testing data from Approved Vendor Assessments that meet criteria established by the Ohio Department of Education.
- Category C teachers teach grades/subjects with no available Value-Added data or growth data derived from Approved Vendor Assessments. These teachers use locally-developed measures such as Student Learning Objectives (SLOs) or the Shared Attribution of the Value-Added data for their building or district.

In the 2013-14 academic year:

- Approximately 6 percent, or 5,144 teachers who received an OTES rating were Category A1 teachers
- Approximately 14 percent, or 11,865 teachers who received an OTES rating were Category A2 teachers
- Approximately 14 percent, or 12,120 teachers who received an OTES rating were Category B teachers
- Approximately 66 percent, or 57,531 teachers who received an OTES rating were Category C teachers

Of the SGM components, besides Value-Added data that are available for use in Ohio schools and districts, approximately 87 percent of schools and districts used at least some SLOs in 2013-14, approximately 38 percent used at least some Shared Attribution, and approximately 47 percent of

schools and districts used at least some Vendor Assessment data as part of OTES/OPES.

A persistent concern voiced by participants in focus groups and by survey respondents was the issue of inherent differences among the type of SGM used (Value-Added, approved Vendor Assessment, Shared Attribution, and Student Learning Objectives)--differences that (at least in the eyes of some participants) would be consequential for the ratings received by Category A, B, and C teachers. We test this perception by disaggregating the Final Summative Rating by an educator's SGM Category (see Table 6 below). What this disaggregation illustrates is that, while only about 31 percent of A1 teachers (i.e., teachers who taught subjects with Value-Added data exclusively) received an Accomplished rating, as many as 50 percent of C teachers (i.e., teachers with no Value-Added or Vendor Assessment data) were rated Accomplished.

Table 6. OTES and OPES Final Summative Rating by Student Growth Category

A. OTES

Final Summative Rating	A1	A2	B	C	Total
Ineffective	37	140	107	580	864
Column percent	0.72	1.18	0.88	1.01	1
Developing	610	1,532	1,145	4,771	8,058
Column percent	11.86	12.91	9.45	8.29	9.3
Skilled	2,881	5,393	6,398	23,473	38,145
Column percent	56.01	45.45	52.79	40.8	44.02
Accomplished	1,616	4,800	4,470	28,707	39,593
Column percent	31.42	40.46	36.88	49.9	45.69
Total	5,144	11,865	12,120	57,531	86,660
Column percent	100	100	100	100	100

B. OPES

Final Summative Rating	A	B	C	Total
Ineffective	27	0	2	29
Column percent	0.85	0	0.11	0.55
Developing	422	56	114	592
Column percent	13.24	18.18	6.38	11.2
Skilled	1,333	212	1,021	2,566
Column percent	41.81	68.83	57.1	48.56
Accomplished	1,406	40	651	2,097
Column percent	44.1	12.99	36.41	39.69
Total	3,188	308	1,788	5,284
Column percent	100	100	100	100

Interestingly, the picture is partially reversed in the case of administrators; only 36 percent of Category C administrators were rated Accomplished versus 44 percent of Category A administrators.

This pattern of teacher ratings by SGM Category is even starker when broken apart for the Big 8 districts versus other school districts in the state (see Table 7 below). Note, for example, that within each SGM Category, there are markedly smaller percentages of Big 8 teachers rated Accomplished (as compared to their colleagues in other Ohio school districts).

Table 7. OTES Final Summative Rating by Student Growth Measures Category and Big 8 Status

Rating	A1		A2		B		C		Total
	Not Big 8	Big 8	Not Big 8	Big 8	Not Big 8	Big 8	Not Big 8	Big 8	
Ineffective	32	5	66	74	71	36	341	239	864
	0.66%	1.75%	0.67%	3.74%	0.67%	2.28%	0.70%	2.82%	1.00%
Developing	545	65	1,102	430	890	255	3,540	1,231	8,058
	11.22%	22.73%	11.15%	21.72%	8.44%	16.17%	7.22%	14.53%	9.30%
Skilled	2,728	153	4,393	1,000	5,523	875	19,246	4,227	38,145
	56.15%	53.50%	44.44%	50.51%	52.39%	55.49%	39.23%	49.89%	44.02%
Accomplished	1,553	63	4,324	476	4,059	411	25,932	2,775	39,593
	31.97%	22.03%	43.74%	24.04%	38.50%	26.06%	52.86%	32.75%	45.69%
Total	4,858	286	9,885	1,980	10,543	1,577	49,059	8,472	86,660

Further, within each SGM category Big 8 teachers are in general (i) twice as likely to be rated Developing, and (ii) at least three times as likely to be rated Ineffective as their colleagues in other school districts.

Final Summative Ratings by LEA Type

When the Final Summative Ratings are disaggregated by LEA type (see Table 8 below) some differences are visible, albeit markedly only for Community Schools that had a lower percentage of teachers in the highest two categories than the other LEA types. Note that row percentages are reported for each row.

Table 8. Final Summative Ratings (by LEA Type)

A. OTES

LEA Type	Ineffective	Developing	Skilled	Accomplished	Total
CTPD	0 (0.00%)	4 (16.67%)	10 (41.67%)	10 (41.67%)	24 (100.00%)
Comm. School	144 (3.03%)	1,025 (21.59%)	2,617 (51.12%)	962 (20.26%)	4,748 (100.00%)
ESC	11 (0.72%)	145 (9.46%)	764 (49.87%)	612 (39.95%)	1,532 (100.00%)
JVSD	22 (1.11%)	177 (8.93%)	844 (42.60%)	938 (47.35%)	1,981 (100.00%)
School District	685 (0.87%)	6,700 (8.55%)	33,886 (43.27%)	37,047 (47.30%)	78,318 (100.00%)
STEM	2 (3.51%)	7 (12.28%)	24 (42.10%)	24 (42.10%)	57 (100.00%)
Total	864	8,058	38,145	39,593	86,660

B. OPES

LEA Type	Ineffective	Developing	Skilled	Accomplished	Total
Comm. School	3 (0.95%)	54 (17.19%)	191 (60.83%)	66 (21.02%)	314 (100.00%)
ESC	0 (0.00%)	3 (7.50%)	21 (52.50%)	16 (40.00%)	40 (100.00%)
JVSD	0 (0.00%)	13 (7.74%)	104 (61.90%)	51 (30.36%)	168 (100.00%)
School District	26 (0.54%)	520 (10.92%)	2,250 (47.27%)	1,964 (41.26%)	4,760 (100.00%)
STEM	0 (0.00%)	2 (100.00%)	0 (0.00%)	0 (0.00%)	2 (100.00%)
Total	29	592	2,586	2,097	5,284

Did teachers and principals in the Big 8 districts (Akron, Canton, Cincinnati, Cleveland, Columbus, Dayton, Toledo, Youngstown) receive similar ratings compared to their peers in all other school districts? Table 9 shows some marked differences between these two groups of districts. Specifically, almost 19 percent of teachers in the Big 8 districts were rated as Ineffective or Developing, compared to only 8 percent of their colleagues in other districts. What is also notable is that only 30 percent of the Big 8 district teachers were rated Accomplished compared to 50 percent of teachers in other districts. Further, these gaps are more pronounced for administrators; 14 percent of the Big 8 principals\assistant principals were rated Accomplished versus 46 percent of their peers in other districts.

Table 9. OTES and OPES Ratings for Big 8 versus All Other School Districts

A. OTES

	Ineffective	Developing	Skilled	Accomplished	Total
Not Big 8	331	4,719	27,631	33,322	66,003
Row percent	0.5	7.15	41.86	50.49	100
Big 8	354	1,981	6,255	3,725	12,315
Row percent	2.87	16.09	50.79	30.25	100
Total	685	6,700	33,886	37,047	78,318
Row percent	0.87	8.55	43.27	47.3	100

B. OPES

	Ineffective	Developing	Skilled	Accomplished	Total
Not Big 8	14	335	1,834	1,864	4,047
Row percent	0.35	8.28	45.32	46.06	100
Big 8	12	185	416	100	713
Row percent	1.68	25.95	58.35	14.03	100
Total	26	520	2,250	1,964	4,760
Row percent	0.55	10.92	47.27	41.26	100

District Typology and Final Summative Ratings

Since 1996 the Ohio Department of Education has used a classification system to stratify districts for research purposes. This typology has been revised over time, most recently in 2013, wherein cluster analysis was used to create an eight-fold classification. Whereas the preceding (i.e., Table 9) disaggregation of Final Summative ratings uses a broad disaggregation of the Big districts versus other school districts, the ODE typology allows for somewhat finer comparisons, both across district types but also within a broader grouping (for example, comparing rural districts with high versus average student poverty, etc.). In Table 10 we map the resulting distribution of Final

Summative ratings (note that the percentages are row percentages indicating, for example, that 0.45 percent of teachers in rural districts with high student poverty and small student populations are Ineffective, 44.34 percent of teachers in these districts are Accomplished, and so forth).

Table 10. OTES and OPES Final Summative Rating by District Type

A. OTES

District Type	Ineffective	Developing	Skilled	Accomplished	Total
Rural - High Student Poverty & Small Student Population	36	689	3,698	3,523	7,946
Row percent	0.45	8.67	46.54	44.34	100
Rural - Average Student Poverty & Very Small Student Population	33	430	2,476	2,704	5,643
Row percent	0.58	7.62	43.88	47.92	100
Small Town - Low Student Poverty & Small Student Population	29	521	3,603	3,547	7,700
Row percent	0.38	6.77	46.79	46.06	100
Small Town - High Student Poverty & Average Student Population	37	744	3,738	4,017	8,536
Row percent	0.43	8.72	43.79	47.06	100
Suburban - Low Student Poverty & Average Student Population	65	980	5,580	8,535	15,160
Row percent	0.43	6.46	36.81	56.3	100
Suburban - Very Low Student Poverty & Large Student Population	18	283	3,596	6,279	10,176
Row percent	0.18	2.78	35.34	61.7	100
Urban - High Student Poverty & Average Student Population	131	1,225	5,501	5,044	11,901
Row percent	1.1	10.29	46.22	42.38	100
Urban - Very High Student Poverty & Very Large Student Population	336	1,828	5,678	3,398	11,240
Row percent	2.99	16.26	50.52	30.23	100
Total	685	6,700	33,870	37,047	78,302
Row percent	0.87	8.56	43.26	47.31	100

B. OPES

District Type	Ineffective	Developing	Skilled	Accomplished	Total
Rural - High Student Poverty & Small Student Population	4	63	303	179	549
Row percent	0.73	11.48	55.19	32.6	100
Rural - Average Student Poverty & Very Small Student Population	3	29	176	142	350
Row percent	0.86	8.29	50.29	40.57	100
Small Town - Low Student Poverty & Small Student Population	2	23	253	244	522
Row percent	0.38	4.41	48.47	46.74	100
Small Town - High Student Poverty & Average Student Population	0	55	279	245	579
Row percent	0	9.5	48.19	42.31	100
Suburban - Low Student Poverty & Average Student Population	2	61	337	422	822
Row percent	0.24	7.42	41	51.34	100
Suburban - Very Low Student Poverty & Large Student Population	0	19	161	430	610
Row percent	0	3.11	26.39	70.49	100
Urban - High Student Poverty & Average Student Population	3	101	378	205	687
Row percent	0.44	14.7	55.02	29.84	100
Urban - Very High Student Poverty & Very Large Student Population	12	169	363	97	641
Row percent	1.87	26.37	56.63	15.13	100
Total	26	520	2,250	1,964	4,760
Row percent	0.55	10.92	47.27	41.26	100

As is evident from the preceding tables, the Urban districts tend to have larger percentages of their teachers rated Ineffective or Developing than all other district types. This pattern is even more pronounced for the administrators. Further, higher levels of student poverty seem to be associated with more teachers\administrators rated as Ineffective or Developing.

A Closer Look at Student Growth Measures

The 2013-14 statewide launch of OTES and OPES introduced new types of assessments to measure student growth at the classroom level. Teacher-level Value-Added Analysis was introduced statewide in 2011 for teachers in reading and math grades 4 through 8, so in-service teachers in those subjects and grades had some familiarity with the metric and their previous Value-Added ratings. However, teachers in subjects and grades without Value-Added data began implementing other types of assessments that would be used to calculate their SGM ratings for the first time in 2013-14. These included Approved Vendor Assessments, Student Learning Objectives, and the Shared Attribution of building- or district-level Value-Added data. While teachers may have had some experience with the Approved Vendor Assessments or assessments used as part of their SLOs, these assessments had not been used to calculate a SGM rating for a teacher according to the OTES guidelines. Here we take a closer look at the different types of SGMs used in OTES in 2013-14, although there is no way of knowing if similar patterns will hold true in future implementation years.

The percentage of teachers in each of the three rating categories for the *Student Growth Measures (SGM)* portion is indicated below for all teachers and for teachers by their SGM type. While the distribution of *Performance on Standards* ratings was similar for teachers across all SGM types (Cat. A1, A2, B, C), the distribution of *SGM* ratings varied by SGM type.

Table 11. Percent of Teachers in each of the SGM Rating Categories by Type of SGM

Category	% Below	% Expected	% Above
All Teachers	6.8%	44.2%	49%
Cat. A1	10.9%	57.2%	31.9%
Cat. A2	11.5%	46.9%	41.6%
Cat. B	5.7%	55.6%	38.7%
Cat. C	5.7%	39.9%	54.4%

It is clear from these data that differences in SGM ratings across SGM categories played a larger role in generating variation in teachers' Final Summative Ratings than did performance standards ratings. Table 12 again illustrates the differences in Final Summative Ratings by SGM category.

Table 12. Teacher (OTES) Final Summative Ratings by SGM Type (2013-14)

Rating	Cat. A1	Cat. A2	Cat. B	Cat. C
Accomplished	31%	41%	37%	50%
Skilled	56%	46%	53%	41%
Developing	12%	13%	9%	8%
Ineffective	1%	1%	1%	1%

Use of Shared Attribution

Slightly more than one-third of LEAs participating in OTES used Shared Attribution in 2013-14; Shared Attribution was used to determine all or part of the SGM rating for 31 percent (26,985) of all Ohio teachers participating in OTES. For the 2013-14 academic year, schools and districts knew what their building and district-level Value-Added ratings were before they made the final decisions on whether or not to include Shared Attribution as a portion of OTES and OPES. As a result, the decision to use Shared Attribution was somewhat related to the district overall Value-Added letter grade on the District Report Cards. For example, 48 percent of districts with an A in Value-Added and 61 percent of districts with a B in Value-Added for 2013-14 opted for shared attribution as one of the SGMs, while a lower percentage of districts receiving a D or F in Value-Added opted for shared attribution. The weight assigned to Shared Attribution ranged from 10 percent (for approx. 7,000 teachers) to 50 percent (for approx. 11,000 teachers).

Tables 13 and 14 below indicate the percentage of school districts (traditional) within each district type who opted for either all 50% of the SGM rating for Category C teachers be comprised of Shared Attribution (Table 13), or at least some portion of the SGM ratings be comprised of Shared Attribution (Table 14).

Table 13. Percent of Districts Using 50% Shared Attribution for OTES Cat. C by District Type

District Type	As % of District Type
Rural - High Student Poverty & Small Student Population	28.3%
Rural - Average Student Poverty & Very Small Student Population	24.4%
Small Town - Low Student Poverty & Small Student Population	21.0%
Small Town - High Student Poverty & Average Student Population Size	27.3%
Suburban - Low Student Poverty & Average Student Population Size	32.8%
Suburban - Very Low Student Poverty & Large Student Population	43.8%
Urban - High Student Poverty & Average Student Population	34.1%
Urban - Very High Student Poverty & Very Large Student Population	50.0%
Total	28.6%

Table 14. Percent of Districts Using any Percentage of Shared Attribution for OTES by District Type

District Type	As % of District Type
Rural - High Student Poverty & Small Student Population	43.5%
Rural - Average Student Poverty & Very Small Student Population	43.3%
Small Town - Low Student Poverty & Small Student Population	32.1%
Small Town - High Student Poverty & Average Student Population Size	36.4%
Suburban - Low Student Poverty & Average Student Population Size	47.5%
Suburban - Very Low Student Poverty & Large Student Population	56.3%
Urban - High Student Poverty & Average Student Population	43.9%
Urban - Very High Student Poverty & Very Large Student Population	50.0%
Total	42.0%

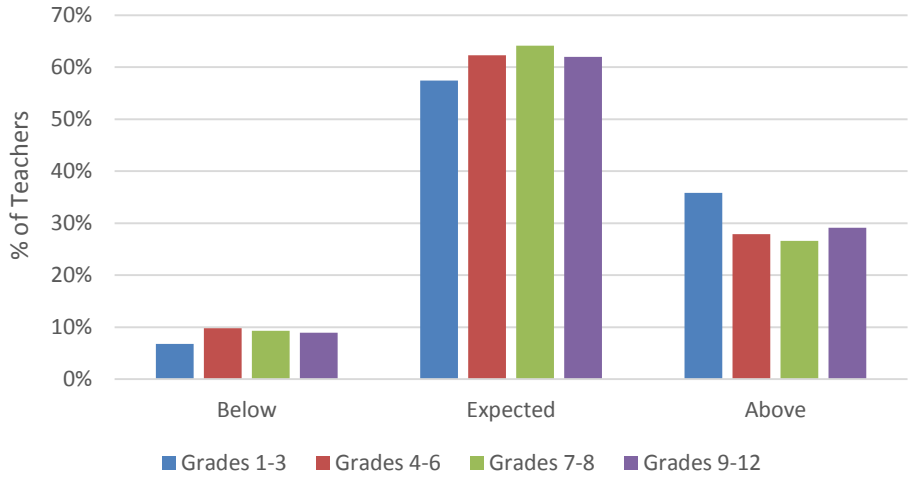
Student Growth Measures by Grade and Subject

The eTPES data made available to the research team did not include grade and subject(s) for teachers participating in OTES. However, the Ohio Department of Education was able to provide teacher credential data that could be merged with the eTPES dataset. The individual teacher grade(s) and subject(s) data are imprecise, as a teacher may be listed as credentialed to teach a grade or subject but may not be teaching that grade and/or subject in any given year. Therefore, the analysis of these data provide only a rough indication of final summative ratings by grade(s) and subject(s) taught. These data indicate that there is little difference in Final Summative Ratings by grade and a relatively small difference in Final Summative Ratings based on whether a teacher is listed as teaching a core subject (English, Math, Social Studies, Science) or a selected non-core or specialty subject (Arts, Physical Education, Foreign Language, Health).

Figures 6 and 7 examine Category B teachers (teachers with Approved Vendor Assessments) and Category C (teachers with locally-developed measures such as SLOs and/or Shared Attribution) ratings on the Student Growth Measures side of OTES disaggregated by the grade they teach. As the figures illustrate, there are relatively small differences in SGM ratings for both Category B and C teachers across grade bands. This analysis was limited to SGM Categories B and C as Category A grade bands were restricted to grades 4 through 8 only in 2013-14.

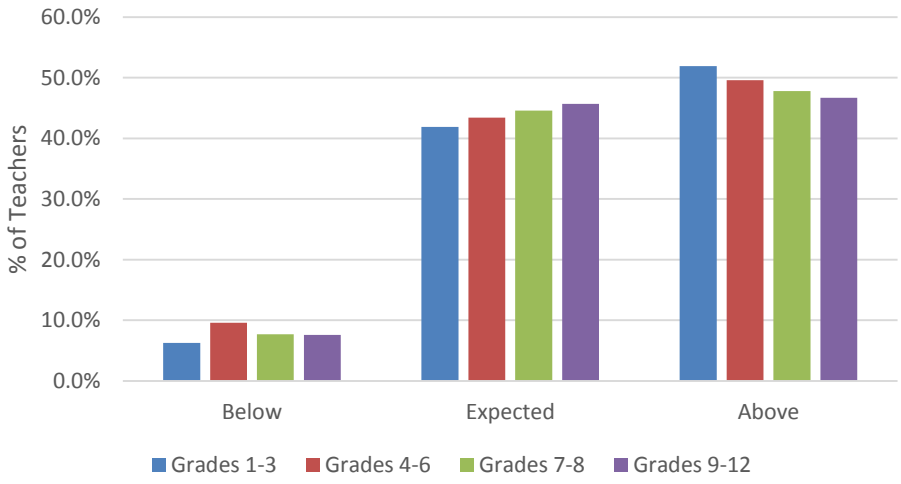
Category B teachers in the primary grades were somewhat more likely to be rated Above on the SGM side of OTES than teachers in higher grades. One explanation for this is that primary grade teachers are perhaps more likely to have experience with Approved Vendor Assessments for diagnostic purposes.

Figure 6. Category B SGM Rating by Grade Band 2013-14



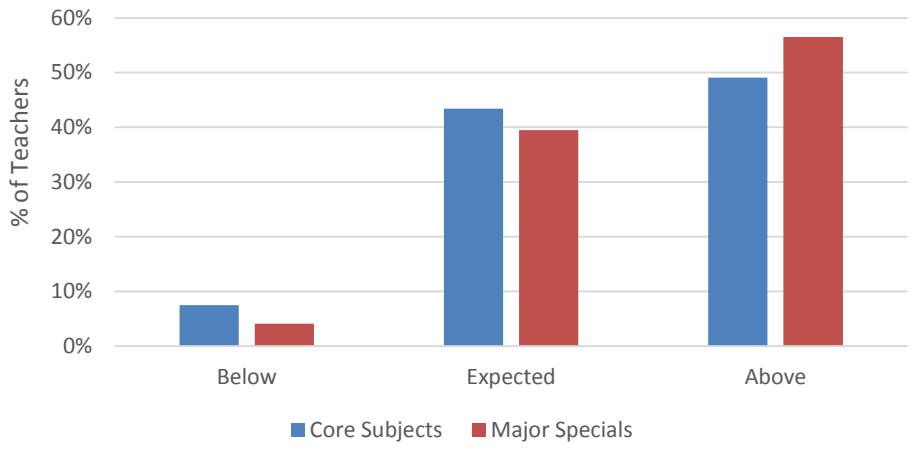
For Category C teachers (Figure 7), there is a slight pattern evident in which teachers in the lower grades are more likely to be rated Above on the SGM side of OTES than teachers in higher grades, but the differences in percentages are small.

Figure 7. Category C SGM Rating by Grade Band 2013-14



We also disaggregated Category C teachers' SGM ratings by whether the teacher taught a core subject (English, Math, Science, Social Studies) or one of the major specialty subjects (Art, Physical Education, Health, Foreign Languages). This analysis was limited to Category C teachers, as most teachers of specialty subjects fall into Category C. As Figure 8 indicates, teachers in the specialty subjects were somewhat more likely to receive an Above rating on the SGM side of OTES than were teachers of core subjects.

Figure 8. Category C SGM Rating by Core and Specials Subjects 2013-14



Do Weights Assigned to SGMs Matter for Final Summative Ratings?

In Table 15 we explore whether the weight assigned to a particular SGM component mattered for the Final Summative Rating. Specifically, were teachers with all of their SGM rating coming from Value-Added less or more likely to be rated Ineffective or Developing versus Skilled or Accomplished than their colleagues with Value-Added being used in conjunction with some other SGM measure? What about Approved Vendor Assessments, Student Learning Objectives and Shared Attribution? As the distributions mapped in Table 15 show, whether a single SGM component made up all of the SGM rating or not was consequential for the rating. Moreover, the consequences differed across the SGM components.

For example, note that only 21 percent of teachers with Value-Added making up all of their SGM rating are rated Accomplished as against 46 percent of their colleagues with Value-Added being used in conjunction with some other SGM component. A similar gap is evident for Approved Vendor Assessments; only 19 percent of teachers with Approved Vendor Assessments being the only SGM component are rated Accomplished versus 47 percent of teachers with Approved Vendor Assessments being used along with some other student growth measure.

Table 15. OTES Final Summative Rating by SGM Weight

Final Summative Rating	Value-Added			Vendor Assessments			Student Learning Objectives			Shared Attribution		
	< 50%	50%	Total	< 50%	50%	Total	< 50%	50%	Total	< 50%	50%	Total
Ineffective	116	61	177	68	39	107	240	432	672	120	68	188
Column percent	1.04	1.03	1.04	0.9	0.85	0.88	1.05	1.16	1.12	0.76	0.61	0.7
Developing	1,292	850	2,142	658	487	1,145	2,312	3,448	5,760	1,137	652	1,789
Column percent	11.63	14.4	12.59	8.76	10.57	9.45	10.13	9.22	9.57	7.18	5.84	6.63
Skilled	4,532	3,742	8,274	3,208	3,190	6,398	9,832	16,396	26,228	6,464	3,293	9,757
Column percent	40.8	63.41	48.64	42.69	69.26	52.79	43.07	43.85	43.55	40.84	29.51	36.16
Accomplished	5,168	1,248	6,416	3,580	890	4,470	10,445	17,114	27,559	8,105	7,146	15,251
Column percent	46.53	21.15	37.72	47.64	19.32	36.88	45.75	45.77	45.76	51.21	64.04	56.52
Total	11,108	5,901	17,009	7,514	4,606	12,120	22,829	37,390	60,219	15,826	11,159	26,985
Column percent	100	100	100	100	100	100	100	100	100	100	100	100

The situation is different in the cases of Student Learning Objectives, and Shared Attribution. Specifically, whether SLOs were the sole measure generating the SGM rating or used in conjunction with other measures appears to have no bearing on the Final Summative Rating. For Shared Attribution, 64 percent of teachers with Shared Attribution being the only SGM component ended up receiving Final Summative ratings of Accomplished, as compared to 51 percent of teachers with Shared Attribution being used along with some other student growth measure. As stated earlier, schools and districts were aware of their building- and district-level Value-Added ratings before they made the final decisions on including Shared Attribution. The distribution of Final Summative Ratings would be different if all schools and districts used Shared Attribution.

Student Growth Measures vs. Performance on Standards

We now turn to finer analyses that focus on the two Sub-ratings -- those derived for (i) Student Growth Measures, and for (ii) Performance on Standards – that feed into the Final Summative Rating. These analyses enable us to engage specific questions. For example, a theme that emerged consistently in our fieldwork was the perception that the two sides didn't really mesh well. If this is indeed a misperception then the distributions should be roughly similar for both components; about the same proportion of teachers should be rated high or low on one as on the other. Of course this exercise is somewhat limited by the fact that while the Performance on Standards rating follows a four category distribution (Ineffective, Developing, Skilled, Accomplished) Student Growth Measures ratings follow a three category system (Below, Expected, and Above). Further, expecting identical distributions for both of these components assumes that two educators who are equally effective on the seven dimensions that underlie the Performance on Standards rating are also equally effective at generating student growth. This assumption may or may not be true. Therefore we treat the analyses that follow as strictly exploratory.

We begin with simple cross-tabulations of the two sub-ratings (see Table 16). Note that in this table the percentages reported are row percentages, i.e., what percent of teachers rated Ineffective on the Performance on Standards have SGM ratings of Below (36.04 percent), Expected (38.44 percent), or Above (25.53 percent).

Table 16. OTES and OPES Performance on Standards Ratings by SGM Ratings

	Student Growth Measures Rating							
Performance on Standards Rating	**OTES**				**OPES**			
	Below	**Expected**	**Above**	**Total**	**Below**	**Expected**	**Above**	**Total**
Ineffective	120	128	85	333	4	14	5	23
	36.04	38.44	25.53	100	17.39	60.87	21.74	100
Developing	698	2,561	1,750	5,009	25	266	131	422
	13.93	51.13	34.94	100	5.92	63.03	31.04	100
Skilled	4,188	26,841	26,896	57,925	243	1,976	1,359	3,578
	7.23	46.34	46.43	100	6.79	55.23	37.98	100
Accomplished	683	7,202	11,984	19,869	64	459	738	1,261
	3.44	36.25	60.32	100	5.08	36.4	58.52	100
Total	5,689	36,732	40,715	83,136	336	2,715	2,233	5,284
	6.84	44.18	48.97	100	6.36	51.38	42.26	100

Note that for both teachers and administrators there is a notable pattern in the distribution of teachers rated Ineffective on the Performance on Standards as compared to those in any of the other three Performance on Standards ratings. That is, for teachers flagged as Ineffective on the Performance on Standards side of OTES, a significant percentage of these individuals were rated Expected (38.44 percent) or Above (25.53 percent) by the corresponding SGM rating. The same pattern is not evident for the other end of the Performance on Standards rating scale: Less than 3.5 percent of teachers rated Accomplished on the Performance on Standards are rated Below on SGMs. The pattern seen for teachers is also present for administrators. This pattern suggests that the two sides of OTES and OPES were more synchronized for Skilled or Accomplished (perhaps even for Developing) teachers than they were for Ineffective teachers.

The Big 8 Versus All Other Districts

Given the markedly different distributions of Final Summative ratings seen previously for the Big 8 districts, we disaggregate the preceding tables by Big 8 status as well (see Table 17).

Table 17. OTES & OPES Performance on Standards Ratings versus SGM Ratings by Big 8 Status

A. OTES

	All Other Districts				Big 8 Districts			
Performance on Standards Rating	**Below**	**Expected**	**Above**	**Total**	**Below**	**Expected**	**Above**	**Total**
Ineffective	33	71	61	165	65	38	18	121
	20.00%	43.03%	36.97%	100%	53.72%	31.40%	14.88%	100%
Developing	284	1,450	1,278	3012	263	618	248	1,129
	9.43%	48.14%	42.43%	100%	23.29%	54.74%	21.97%	100%
Skilled	2,588	19,663	22,714	44965	949	4,106	2,299	7,354
	5.76%	43.73%	50.51%	100%	12.90%	55.83%	31.26%	100%
Accomplished	435	5,554	10,222	16211	174	1,203	1,188	2,565
	2.68%	34.26%	63.06%	100%	6.78%	46.90%	46.32%	100%
Total	3,340	26,738	34,275	64,353	1,451	5,965	3,753	11,169

B. OPES

	All Other Districts				Big 8 Districts			
Performance on Standards Rating	**Below**	**Expected**	**Above**	**Total**	**Below**	**Expected**	**Above**	**Total**
Ineffective	1	6	3	10	2	7	2	11
	10.00%	60.00%	30.00%	100%	18.18%	63.64%	18.18%	100%
Developing	13	151	104	268	10	76	13	99
	4.85%	56.34%	38.81%	100%	10.10%	76.77%	13.13%	100%
Skilled	139	1,388	1,206	2733	72	332	64	468
	5.09%	50.79%	44.13%	100%	15.38%	70.94%	13.68%	100%
Accomplished	36	342	658	1036	28	71	36	135
	3.47%	33.01%	63.51%	100%	20.74%	52.59%	26.67%	100%
Total	189	1,887	1,971	4,047	112	486	115	713

Note that the percentages reported in Table 17 are row percentages calculated for each Performance on Standards rating, and separately for Big 8 versus all other districts. Thus, for example, almost 13 percent of Big 8 teachers with a Performance on Standards rating of Skilled had a Student Growth Measures rating of Below, almost 56 percent had an SGM rating of Expected, and 31 percent had an SGM rating of Above. What can we make of Table 17? First, it should be obvious that in all other school districts a larger percentage of teachers have SGM ratings of Above (regardless of their Performance on Standards rating) than do teachers in the Big 8 districts.

Second, if we focus on the question of symmetry between the Performance on Standards rating and the Student Growth Measures rating, the two sides of OTES appear to be more synchronized for the Big 8 districts than for all other districts. Note, for example, that the majority of Big 8 teachers rated Ineffective on the Performance on Standards (almost 54 percent) had an SGM rating of Below, but this concordance is true only for 20 percent of the teachers rated Ineffective in all other districts. As the Performance on Standards rating increases for Big 8 teachers, increasingly smaller percentages of these teachers are rated Below on the SGM portion of their evaluation.

Finally, it is worth noting that in the case of administrators, a larger percentage of the Big 8 principals and assistant principals have SGM ratings of Below for every Performance on Standards rating than is the case for their counterparts in all other districts.

Ratings Disaggregated by SGM Category

In this section we disaggregate the cross-tabulations of Performance on Standards ratings versus Student Growth Measures ratings by SGM type (see Table 18). Note that the percentages reported in these tables are calculated on the basis of all A1 teachers, all A2 teachers, all B teachers, and all C teachers, respectively. Thus, for example, 8.5 percent of A1 teachers had a Performance on Standards rating of Skilled and an SGM rating of Below, almost 11 percent of A1 teachers had SGM ratings of Below, about 30 percent of A1 teachers were rated Skilled, and so on.

Table 18. OTES and OPES Performance on Standards Ratings versus Student Growth Measures Ratings by SGM Category

A. OTES

	A1				A2				B				C			
Performance on Standards Rating	**Below**	**Expected**	**Above**	**Total**	**Below**	**Expected**	**Above**	**Total**	**Below**	**Expected**	**Above**	**Total**	**Below**	**Expected**	**Above**	**Total**
Ineffective	3	6	0	9	21	16	6	43	14	23	10	47	82	83	69	234
Percent of SGM Category	0.06	0.12	0	0.17	0.18	0.13	0.05	0.36	0.12	0.19	0.08	0.4	0.15	0.15	0.13	0.43
Developing	34	82	25	141	119	285	126	530	90	522	154	766	455	1,672	1,445	3,572
Percent of SGM Category	0.66	1.59	0.49	2.74	1	2.4	1.06	4.47	0.76	4.42	1.3	6.48	0.84	3.08	2.66	6.58
Skilled	437	2,076	931	3,444	1,032	4,010	3,124	8,166	461	4,594	2,757	7,812	2,258	16,161	20,084	38,503
Percent of SGM Category	8.5	40.36	18.1	66.95	8.7	33.8	26.33	68.82	3.9	38.89	23.34	66.13	4.16	29.75	36.98	70.89
Accomplished	85	780	685	1,550	193	1,257	1,676	3,126	106	1,433	1,649	3,188	299	3,732	7,974	12,005
Percent of SGM Category	1.65	15.16	13.32	30.13	1.63	10.59	14.13	26.35	0.9	12.13	13.96	26.99	0.55	6.87	14.68	22.1
Total	559	2,944	1,641	5,144	1,365	5,568	4,932	11,865	671	6,572	4,570	11,813	3,094	21,648	29,572	54,314
Percent of SGM Category	10.87	57.23	31.9	100	11.5	46.93	41.57	100	5.68	55.63	38.69	100	5.7	39.86	54.45	100

B. OPES

	A				B				C			
Performance on Standards Rating	**Below**	**Expected**	**Above**	**Total**	**Below**	**Expected**	**Above**	**Total**	**Below**	**Expected**	**Above**	**Total**
Ineffective	2	7	3	12	0	2	0	2	2	5	2	9
	0.06	0.22	0.09	0.38	0	0.65	0	0.65	0.11	0.28	0.11	0.5
Developing	25	135	78	238	0	49	7	56	0	82	46	128
	0.78	4.23	2.45	7.47	0	15.91	2.27	18.18	0	4.59	2.57	7.16
Skilled	214	992	888	2,094	5	175	25	205	24	809	446	1,279
	6.71	31.12	27.85	65.68	1.62	56.82	8.12	66.56	1.34	45.25	24.94	71.53
Accomplished	63	263	518	844	0	30	15	45	1	166	205	372
	1.98	8.25	16.25	26.47	0	9.74	4.87	14.61	0.06	9.28	11.47	20.81
Total	304	1,397	1,487	3,188	5	256	47	308	27	1,062	699	1,788
	9.54	43.82	46.64	100	1.62	83.12	15.26	100	1.51	59.4	39.09	100

The preceding tables reflect some interesting patterns. First, the distribution of Performance on Standards ratings are roughly similar regardless of SGM category. For example, less than 1 percent of teachers in each SGM category are rated Ineffective, less than 7 percent are rated Developing, 66 percent or more are rated Skilled, and 22 percent or more are rated Accomplished on the Performance on Standards. What is also noteworthy is that the highest percentage of teachers (30 percent) to be rated Accomplished on the Performance on Standards side are the A1 teachers, as compared to 22 percent of C teachers; some 26 percent and 27 percent of SGM Category A2 and B teachers received Accomplished ratings. The picture is different if you focus on the Student Growth Measures ratings. In particular, almost 11 percent of A1 teachers and 11.5 percent of A2 teachers received an SGM rating of Below, as compared to less than 6 percent of B and C teachers. A1 teachers also had the lowest percentage (almost 32 percent) rated Above on Student Growth Measures, in contrast to A2 teachers (almost 42 percent), B teachers (almost 39 percent), and most conspicuously, C teachers (54 percent). This suggests that A1 and A2 teachers had the highest probability of being rated Below and the lowest probability of being rated Above on the SGM side of OTES. Correspondingly, C teachers had the highest probability of receiving an SGM rating of Above and the lowest probability of receiving an SGM rating of Below.

The overall picture is roughly similar for administrators. That is, a larger percentage of Category A administrators were rated Below on Student Growth Measures than their B (almost 2 percent) and C (almost 2 percent) counterparts. The largest percentage (26 percent) of the Performance on Standards-rated Accomplished administrators also happened to be the Category A administrators, followed by Category C (almost 21 percent) and Category B (almost 15 percent) administrators. An overwhelming share (83 percent) of Category B administrators also received an SGM rating of Expected, as compared to about 44 percent of Category A and 59 percent of Category C administrators.

Performance on Standards and SGMs by District Typology

In Table 19 we map teachers' Performance on Standards ratings versus SGM ratings disaggregated by district type, and several interesting patterns are visible in the data. First, while the Urban districts have the smallest percentages of teachers with SGM ratings of Above – Urban Very High Poverty = 33.51 percent; Urban High Poverty = 45.71 percent – Suburban districts have the highest percentage of similarly rated teachers – Suburban Very Low Poverty = 64.29 percent, Suburban Low Poverty = 58.92 percent. Further, while the Urban Very High Poverty districts have the highest percentage of teachers rated Below (13.20 percent), Suburban Very Low Poverty districts have the lowest percentage of teachers rated Below (1.95 percent).

Second, if we focus on the distributions of Performance on Standards ratings within each block of SGM ratings it is quickly apparent that the distributions of the four Performance on Standards ratings vary across district types. Thus, for example, Urban Very High Poverty districts have the largest percentages of teachers rated Below – Ineffective (0.58 percent), Below – Developing (2.49 percent), Below – Skilled (8.50 percent), and Below – Accomplished (1.63 percent) than any other district type. This is also true for SGM ratings of Expected but the differences are less pronounced across district type.

Table 19. OTES Performance on Standards Ratings versus SGM Ratings by District Type

	SGM Rating = Below					SGM Rating = Expected					SGM Rating = Above					Total
District Typology	**I**	**D**	**S**	**A**	**Total**	**I**	**D**	**S**	**A**	**Total**	**I**	**D**	**S**	**A**	**Total**	
Rural - High Student Poverty & Small Student Pop	1	34	404	46	485	12	204	2714	691	3621	8	156	2504	962	3630	7736
	0.01%	0.44%	5.22%	0.59%	6.27%	0.16%	2.64%	35.08%	8.93%	46.81%	0.10%	2.02%	32.37%	12.44%	46.92%	100.00%
Rural - Avg Student Pov &Very Small Student Pop	2	28	260	31	321	5	123	1859	430	2417	6	134	1890	800	2830	5568
	0.04%	0.50%	4.67%	0.56%	5.77%	0.09%	2.21%	33.39%	7.72%	43.41%	0.11%	2.41%	33.94%	14.37%	50.83%	100.00%
Small Town - Low Student Poverty & Small Student Pop	3	26	279	45	353	6	168	2530	798	3502	9	166	2491	1024	3690	7545
	0.04%	0.34%	3.70%	0.60%	4.68%	0.08%	2.23%	33.53%	10.58%	46.41%	0.12%	2.20%	33.02%	13.57%	48.91%	100.00%
Small Town - High Student Poverty & Avg Student Pop	5	31	422	69	527	6	223	2793	610	3632	7	189	3065	905	4166	8325
	0.06%	0.37%	5.07%	0.83%	6.33%	0.07%	2.68%	33.55%	7.33%	43.63%	0.08%	2.27%	36.82%	10.87%	50.04%	100.00%
Suburban - Low Student Poverty & Avg Student Pop	5	58	524	125	712	18	282	3891	1223	5414	12	288	5754	2731	8785	14911
	0.03%	0.39%	3.51%	0.84%	4.77%	0.12%	1.89%	26.09%	8.20%	36.31%	0.08%	1.93%	38.59%	18.32%	58.92%	100.00%
Suburban - Very Low Student Poverty & Large Student Pop	3	12	116	58	189	3	85	2170	1019	3277	5	97	3274	2864	6240	9706
	0.03%	0.12%	1.20%	0.60%	1.95%	0.03%	0.88%	22.36%	10.50%	33.76%	0.05%	1.00%	33.73%	29.51%	64.29%	100.00%
Urban - High Student Poverty & Avg Student Pop	20	105	667	69	861	24	409	4135	839	5407	21	270	3987	999	5277	11545
	0.17%	0.91%	5.78%	0.60%	7.46%	0.21%	3.54%	35.82%	7.27%	46.83%	0.18%	2.34%	34.53%	8.65%	45.71%	100.00%
Urban - Very High Student Poverty & Very Large Student Pop.	59	253	865	166	1343	35	574	3666	1147	5422	11	226	2048	1125	3410	10175
	0.58%	2.49%	8.50%	1.63%	13.20%	0.34%	5.64%	36.03%	11.27%	53.29%	0.11%	2.22%	20.13%	11.06%	33.51%	100.00%
Total	98	547	3537	609	4791	109	2068	23758	6757	32692	79	1526	25013	11410	38028	75511

What is especially noteworthy is the distribution within the SGM rating block of Above. In particular, Urban High Poverty districts and Urban Very High Poverty districts have some of the smallest percentages of teachers rated Above on the SGM side and Accomplished on the Performance on Standards side (8.65 percent and 11.06 percent, respectively).

Calculating the percentage of teachers rated Ineffective, Developing, Skilled, or Accomplished in each district type (see Table 20 below) makes it clear that Urban districts also have the highest percentage of teachers with Performance on Standards ratings of Ineffective (Urban High Poverty = 1.03 percent, Urban Very High Poverty = 0.56 percent) and the smallest percentages of teachers rated Above (Urban High Poverty = 16.52 percent, Urban Very High Poverty = 23.96 percent). Suburban Very Low Poverty districts have the lowest (0.11 percent) and highest (40.60 percent) percentages of teachers rated Ineffective and Accomplished, respectively.

Table 20. OTES Performance on Standards Ratings by District Type

District Type	Total Ineffective (I)	Total Developing (D)	Total Skilled (S)	Total Accomplished (A)	Total
Rural - High Student Poverty & Small Student Population	21	394	5622	1699	7736
	0.27%	5.09%	72.67%	21.96%	100.00%
Rural - Average Student Poverty & Very Small Student Population	13	285	4009	1261	5568
	0.23%	5.12%	72.00%	22.65%	100.00%
Small Town - Low Student Poverty & Small Student Population	18	360	5300	1867	7545
	0.24%	4.77%	70.25%	24.74%	100.00%
Small Town - High Student Poverty & Average Student Population	18	443	6280	1584	8325
	0.22%	5.32%	75.44%	19.03%	100.00%
Suburban - Low Student Poverty & Average Student Population	35	628	10169	4079	14911
	0.23%	4.21%	68.20%	27.36%	100.00%
Suburban - Very Low Student Poverty & Large Student Population	11	194	5560	3941	9706
	0.11%	2.00%	57.28%	40.60%	100.00%
Urban - High Student Poverty & Average Student Population	65	784	8789	1907	11545
	0.56%	6.79%	76.13%	16.52%	100.00%
Urban - Very High Student Poverty & Very Large Student Population	105	1053	6579	2438	10175
	1.03%	10.35%	64.66%	23.96%	100.00%

Table 21: OPES Performance on Standards Ratings versus SGM Ratings by District Type

District Typology	SGM Rating = Below					SGM Rating = Expected					SGM Rating = Above					Total
	I	D	S	A	Total	I	D	S	A	Total	I	D	S	A	Total	
Rural – High Pov & Small Student Pop	0	4	28	10	42	0	25	227	67	319	0	9	133	46	188	549
	0.00%	0.73%	5.10%	1.82%	7.65%	0.00%	4.55%	41.35%	12.20%	58.11%	0.00%	1.64%	24.23%	8.38%	34.24%	100.00%
Rural - Avg Student Poverty & Very Small Student Pop	1	2	15	0	18	0	14	126	39	179	0	11	81	61	153	350
	0.29%	0.57%	4.29%	0.00%	5.14%	0.00%	4.00%	36.00%	11.14%	51.14%	0.00%	3.14%	23.14%	17.43%	43.71%	100.00%
Small Town - Low Poverty & Small Student Pop	0	2	7	6	15	1	9	194	50	254	0	9	145	99	253	522
	0.00%	0.38%	1.34%	1.15%	2.87%	0.19%	1.72%	37.16%	9.58%	48.66%	0.00%	1.72%	27.78%	18.97%	48.47%	100.00%
Small Town - High Student Poverty & Avg Student Pop	0	0	25	4	29	0	25	219	50	294	1	10	155	90	256	579
	0.00%	0.00%	4.32%	0.69%	5.01%	0.00%	4.32%	37.82%	8.64%	50.78%	0.17%	1.73%	26.77%	15.54%	44.21%	100.00%
Suburban - Low Student Poverty Avg Student Pop	0	2	23	8	33	1	28	251	62	342	1	24	288	134	447	822
	0.00%	0.24%	2.80%	0.97%	4.01%	0.12%	3.41%	30.54%	7.54%	41.61%	0.12%	2.92%	35.04%	16.30%	54.38%	100.00%
Suburban - Very Low Student Poverty & Large Student Pop	0	0	8	4	12	0	7	108	43	158	0	10	250	180	440	610
	0.00%	0.00%	1.31%	0.66%	1.97%	0.00%	1.15%	17.70%	7.05%	25.90%	0.00%	1.64%	40.98%	29.51%	72.13%	100.00%
Urban - High Student Poverty & Average Student Pop	0	3	33	4	40	6	57	311	35	409	1	32	156	49	238	687
	0.00%	0.44%	4.80%	0.58%	5.82%	0.87%	8.30%	45.27%	5.09%	59.53%	0.15%	4.66%	22.71%	7.13%	34.64%	100.00%
Urban - Very High Student Poverty & Very Large Student Pop	2	10	72	28	112	5	62	284	67	418	2	12	62	35	111	641
	0.31%	1.56%	11.23%	4.37%	17.47%	0.78%	9.67%	44.31%	10.45%	65.21%	0.31%	1.87%	9.67%	5.46%	17.32%	100.00%
Total	3	23	211	64	301	13	227	1720	413	2373	5	117	1270	694	2086	4760

Table 21 (see above) extends the preceding analysis to administrators' ratings, with generally similar results. For instance, focusing on the percentages of principals\assistant principals with particular SGM ratings, Urban Very High Poverty districts have the smallest percentage rated Above (17.32 percent) and the highest percentage rated Below (17.47 percent), in stark contrast to Suburban Very Low Poverty districts with the highest percentage (72 percent) rated Above and the smallest percentage (1.97 percent) rated Below. This pattern generally extends to Performance on Standards ratings as well (see Table 22 below).

Table 22. OPES Performance on Standards Ratings by District Type

District Typology	Total	Total	Total	Total	
	Ineffective (I)	Developing (D)	Skilled (S)	Accomplished (A)	Total
Rural - High Student Poverty & Small Student Population	0	38	388	123	549
	0.00%	6.92%	70.67%	22.40%	100.00%
Rural - Average Student Poverty & Very Small Student Population	1	27	222	100	350
	0.29%	7.71%	63.43%	28.57%	100.00%
Small Town - Low Student Poverty & Small Student Population	1	20	346	155	522
	0.19%	3.83%	66.28%	29.69%	100.00%
Small Town - High Student Poverty & Average Student Population	1	35	399	144	579
	0.17%	6.04%	68.91%	24.87%	100.00%
Suburban - Low Student Poverty & Average Student Population	2	54	562	204	822
	0.24%	6.57%	68.37%	24.82%	100.00%
Suburban - Very Low Student Poverty & Large Student Population	0	17	366	227	610
	0.00%	2.79%	60.00%	37.21%	100.00%
Urban - High Student Poverty & Average Student Population	7	92	500	88	687
	1.02%	13.39%	72.78%	12.81%	100.00%
Urban - Very High Student Poverty & Very Large Student Population	9	84	418	130	641
	1.40%	13.10%	65.21%	20.28%	100.00%

V. CONCLUSIONS AND NEXT STEPS

While the data presented in this study are detailed and address a number of granular research questions, both the qualitative and quantitative findings of this study allow for the following general conclusions and the need for continued research:

- Ohio, as in other states that are implementing similar new teacher evaluation systems, saw relatively few teachers and principals (approximately 1 percent of both groups) receive the lowest final summative rating of *Ineffective*. Tennessee (http://bit.ly/14pVmRR), New York (http://lohud.us/1y4qrXF), New Jersey (http://www.nj.com/education/2013/12/report_one-quarter_of_teachers.html), Georgia (http://bit.ly/1tZCS73) and Florida (see http://www.fldoe.org/core/fileparse.php/5423/urlt/1314EduEvalRatings.pdf) are among some of the other states that have experienced a similar pattern of a very small percentage of teachers receiving the lowest rating in the new evaluation systems. While some critics of the new evaluation systems may focus on this small proportion as evidence of systems that failed, we see this as a small part of the bigger picture. In particular, whereas in years past most teachers tended to receive the highest ratings, the new systems have started to differentiate between teachers in terms of their effectiveness in the classroom. This differentiation will only improve in precision with time as the evaluation framework is refined and improved. Indeed, some changes that have gone into effect for 2014-2015 will, most likely, generate different distributions of teacher effectiveness than we have seen with the 2013-14 data.

- Modifications should be made to state and local policy regarding SLOs so that they better align with the rigor of other Student Growth Measures used across grades and subjects. Teachers with SLOs were clearly more likely to receive higher final summative ratings than teachers with Value-Added data or with approved Vendor Assessment data. For example, 50 percent of teachers in Category C (with SLOs and/or Shared Attribution) received an *Accomplished* final summative rating, while only 31 percent of teachers with all or at least 26 percent Value-Added data for their SGM received an *Accomplished* Final Summative Rating. It is unlikely that this is a consequence of the relative effectiveness of teachers in tested grades and subjects (including State assessments or approved Vendor Assessments) compared to teachers in specialty subjects where there is no State or approved Vendor Assessment. The processes and guidance for developing and approving SLOs should be reviewed now that schools and districts have some experience with SLOs, as well as data from the SLO assessments. It is important to note that schools and districts were aware of their building- and district-level Value-Added ratings before they made the final decisions about using Shared Attribution of Value-Added data. The proportion of Category C teachers rated Accomplished would be lower if all schools and districts used Shared Attribution for Category C teachers. Ongoing research on OTES and OPES will look for evidence of future alignment of the various Student Growth Measures used in OTES and OPES.

- The Big 8 Urban districts in particular and districts with high levels of student poverty more generally, as well as the weight placed on the components of Student Growth Measures (i.e., whether for example Value-Added made up all or only a portion of a teacher's SGM rating), and the SGM Category appear to influence the Final Summative Rating. In general,

high poverty districts, both urban and rural, showed higher percentages of teachers rated as "Ineffective" or "Developing" than other districts. This pattern warrants further research.

- Teachers and some administrators in the 13 pilot districts that participated in the qualitative portion of this study see fairness issues with OTES---partly because of the difference in the types of Student Growth Measures (e.g. Value-Added based on a State assessment vs. SLOs where the teacher develops the assessment), partly because they feel none of the measures accurately assess a teacher's instructional impact, and also because of their perceived lack of control over so many external factors that influence a student's performance on the assessment. Some teachers also articulated an inaccurate or incomplete understanding of how Value-Added is calculated. Some teachers were positive about the concept of student's academic growth being included as a component of teacher evaluation. Additional professional development on topics such as assessment literacy and Value-Added analysis, coupled with a more rigorous process for SLO development and review may alleviate some of the fairness concerns held by teachers.

Since Ohio will implement additional changes to OTES and OPES starting in the 2014-15 academic year, a careful analysis of the 2014-15 OTES and OPES data, as well as additional field work to gauge teacher and principal experiences and perceptions as the system matures and changes, is critical to understanding whether the system is functioning as intended in order to develop a strong teacher and principal workforce to increase the academic achievement and success of Ohio's students. Perhaps the most critical question across all states implementing the new systems is how teacher and principal perceptions change over time and how these perceptions shape changes in practice. For example, do the new systems foster unintended consequences such as reduced collaboration among teachers or demoralizing the teaching workforce, or do they provide data that focuses teachers' efforts to improve student achievement. In a recent special issue of Educational Researcher, Doug Harris and Carolyn Herrington (2015) frame the essential questions for those implementing and examining the new teacher and principal evaluation systems in Ohio and other states:

> *The main underlying theory of these policies is that teacher accountability will motivate teachers to work harder and smarter and help attract and retain only those who are successful. Does this happen in practice? Does the increased scrutiny lead educators to work harder and smarter in helping their students? Does the recognition that comes with high performance ratings encourage a stronger focus on the student outcomes on which the educator performance measures are based? Are teachers more likely to demand and seek out instructional leadership from their principals, peers, coaches, and other sources?...Do these systems increase cohesion around common goals and expectations at the school level? (p. 72)*

This study begins the work of answering these questions by detailing the results of the first year of implementation of Ohio's new systems of teacher and principal evaluation. It is important to continue to look closely at this system in future years as it is refined in order to gauge whether schools and districts develop more expertise in implementing it effectively and if educators find value/use the data it provides for school improvement purposes.

VI. REFERENCES

Doherty, K. M. & Jacobs, S. (2013). *State of the States 2013. Connect the dots: Using evaluations of teacher effectiveness to inform policy and practice*. National Center on Teacher Quality. Retrieved January 12, 2015 from http://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report

Gill, B., Bruch, J. & Booker, K. (2013). *Using alternative student growth measures for evaluating teacher performance: what the literature says.* (REL 2013–002). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Retrieved from http://ies.ed.gov/ncee/edlabs.

Good, T. L. & Lavigne, A. L. (2015). Issues of teacher performance stability are not new: Limitations and possibilities. *Education Policy Analysis Archives*, [S.l.], v. 23, p. 2, ISSN 1068-2341. Retrieved January 12, 2015 from <http://epaa.asu.edu/ojs/article/view/1916>.

Harris, D. & Herrington, C. (2015). The use of teacher value-added measures in schools: New evidence, unanswered questions & future prospects. *Educational Researcher*. 44: 71-76.

Harris, D., Ingle, W. & Rutledge, S. (2014). How teacher evaluation methods matter for accountability: A comparative analysis of teacher ratings by principals and teacher value-added measures. *American Educational Research Journal*. 51: 73-112.

Kane, T. J. & Staiger, D. O. (2012). *Gathering feedback for teachers: Combining high-quality observations with student surveys and achievement gains*. Policy and practice brief prepared for the Bill and Melinda Gates Foundation.

Mihaly, K., McCaffrey, D.F., Staiger, D.O. & Lockwood, J.R. (2013). *A composite estimator of effective teaching* (MET Project research paper). Seattle, WA: Bill & Melinda Gates Foundation. Retrieved January 30, 2013, from <http://www.metproject.org/downloads/MET_Composite_Estimator_of_Effective_Teaching_Research_Paper.pdf.>

Nutt, P. (2004). Prompting the transformation of public organizations. *Public Performance & Management Review.* 27:4, 9-33.

Pennington, K. (2014). *ESEA Waivers and Teacher-Evaluation Plans: State Oversight of District-Designed Teacher Evaluation Systems*. The Center for American Progress. Available at: http://cdn.americanprogress.org/wp-content/uploads/2014/05/TeacherEvalWaivers-FINAL.pdf

Thorn, C. & and Harris, D. (2013). The accidental revolution: Teacher accountability, value-added, and the restructuring of the American school system. In Dorothea Anagnostopoulos, Rebecca

Jacobsen and Stacey Rutledge (Ed.). *The Infrastructure of Accountability: Data Use and the Transformation of American Education.* Harvard Education Press. Cambridge, MA.

Weisberg, D., Sexton, S., Mulhern, J. & Keeling, D. (2009). *THE WIDGET EFFECT: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness (*The New Teacher Project).

VII. APPENDIX- Focus Group and Interview Instruments and Protocols

A. Informed Consent

This document invites you to participate in an evaluation of the policy and practice surrounding Ohio's Student Growth Measures (SGM) component of the impending educator evaluation system (i.e. OTES/OPES). In brief, effective 2013-2014 student growth-as measured by value-added and/or Vendor Assessments, and/or other LEA Measures (including but not limited to SLOs) will comprise 50% of an educator's performance evaluation. Ani Ruhil and Marsha Lewis from Ohio University are therefore evaluating (a) what are LEAs doing with SGM? What sorts of challenges present themselves in implementing SGM as required by statute, and what can be done to smooth implementation? The findings will be shared with the Ohio Department of Education and school districts in Ohio so that both policy and practice surrounding the states SGM model can be improved.

Please note that your participation is entirely voluntary, and data collected during interviews will not include personal identifying information. Further, all findings will be reported in a manner that masks all district, building, educator, and administrators names/ identifiers. Given the voluntary nature of this evaluation you can choose to terminate participation at any point in time and without prejudice. Participation in the interview implies your consent, and we hope you participate since your insights will help other LEAs as they look to implement SGM.

The interview will last 60 to 90 minutes. There are no risks. Your feedback will provide valuable information to the Ohio Department of Education in the form of recommendations for LEAs moving forward with integrating these measures of student growth with instructional and evaluative processes.

If you have further questions regarding this study please contact the principal investigators, Ani Ruhil (740-597-1949; ruhil@ohio.edu) or Marsha Lewis (740-593-1435; lewism5@ohio.edu).

Ani and Marsha can also be reached at:

Building 21, The Ridges
The Voinovich School of Leadership and Public Affairs
1 Ohio University
Athens, Ohio 45701

For further questions regarding your rights as a research subject contact JO-Ellen Sherow, Director, Office of Research Compliance, Ohio University at 740-593-0664; sherow@ohio.edu.

B. SGMs Pilot Districts-Principal Interview Guide

Hello and thank you for allowing me to talk with you today. My name is _____ and I work for _____ _____.

First, it may be helpful for me to explain a little bit about why I am here to speak with you. The Ohio Department of Education has contracted with the Ohio Education Research Center, which we at Ohio University's Voinovich School are a part of, to conduct research regarding implementation of Ohio's Teacher Evaluation System as it pertains to Student Growth Measures. We are meeting with all of the districts who received pilot grant funding in 2012 to implement SGMs, which includes your district.

It is important to remember that this is information gathering from all of the pilot districts. We are not assessing your implementation of SGMs or your district's implementation of SGMs. There will be no district-level reports written. It is also important for you to understand that I do not work for the Ohio Department of Education and, as such, I am unable to answer any questions about Student Growth Measures.

The purpose of this interview today is to collect your thoughts and feelings about SGMs and we hope this discussion can help us gain insight into how SGMs are being implemented at the local level.

Another important thing to remember during our conversation is that everyone's ideas are important, and they should be allowed to freely express their thoughts and feelings. The ideas expressed here may be personal and should not be used against anyone inside or outside of this meeting.

(If meeting with multiple principals) - From time to time, I may interrupt to allow someone to speak who may not have said anything for a while. Also, I may have to interrupt someone to move on to another question because of the time limit under which we are working. I apologize in advance if this happens.

The discussion will be digitally recorded. The recording will be used for our reference only and will be erased once the research report is complete. Your feedback is important to us, but your identity is not. Our reports to the Ohio Department of Education will not include actual names of participants, so your individual comments will be strictly confidential. Should you feel uncomfortable at any time during the discussion, remember that your participation is voluntary.

Provide consent document so they can review it briefly.

Are there any questions about this procedure?

1. Can you briefly describe your district's work on Student Growth Measures so far?
    a. How are you using value-added data, vendor assessments, SLOs, etc this year for the SGM portion of OTES? Did the district make any changes this year regarding these measures such as adding new assessments because of OTES?

2. In the SGM Portion of a teacher's evaluation, how much weight are you giving to value-added for the teachers who get value added reports?

a. How is the weight determined?
b. Would you have liked more or less flexibility in the weight given to value-added and if so why?
c. Who is involved in the decision about the weight given to value-added?
d. What was the decision-making process like?
e. Were there any differences in opinion about the weight among teachers who have and don't have value-added data?

3. What about the use of vendor assessments?
    a. How is that going?
    b. Are there any challenges in using vendor assessments?
    c. Are there benefits to using vendor assessments?

4. What about LEA Measures such as SLOs?
    a. How is that going?
    b. How much are they used?
    c. How are they being developed, approved, and scored?
    d. Are there any challenges to using SLOs?
    e. Are any other LEA Measures being used and if so can you describe them?
    If not discussed, ask specifically about shared attribution measures, whether they are being used and the decision-process.

5. Are you working with other districts, your ESC, or anyone else on development of LEA Measures?

6. I have one final question. What do you think about the Teacher Performance Evaluation Rubric?
    a. How comfortable are you with the Rubric?
    b. Are there any challenges to using it?
    c. Are there any benefits to using this model?

C. Student Growth Measures Policy and Practice - Teacher Discussion Topic Guide

Category A Teachers (Have Value-Added data)

Hello everyone and thank you for allowing me to talk with you today. My name is _____ and I work for _____.

First, it may be helpful for me to explain a little bit about why I am here to speak with you. The Ohio Department of Education has contracted with the Ohio Education Research Center, which we at Ohio University's Voinovich School are a part of, to conduct research regarding implementation of Ohio's Teacher Evaluation System as it pertains to Student Growth Measures. We are meeting with all of the districts who received pilot grant funding in 2012 to implement SGMs, which includes your district.

It is important to remember that this is information gathering from all of the pilot districts. We are not assessing your implementation of SGMs or your district's implementation of SGMs. There will be no district-level reports written. It is also important for you to understand that I do not work for the Ohio Department of Education and, as such, I am unable to answer any questions about Student Growth Measures.

The purpose of this focus group today is to collect your thoughts and feelings about SGMs and we hope this discussion can help us gain insight into how SGMs are being implemented at the local level.

Another important thing to remember during our conversation is that everyone's ideas are important, and they should be allowed to freely express their thoughts and feelings. The ideas expressed here may be personal and should not be used against anyone inside or outside of this meeting. From time to time, I may interrupt to allow someone to speak who may not have said anything for a while. Also, I may have to interrupt someone to move on to another question because of the time limit under which we are working. I apologize in advance if this happens.

The discussion will be digitally recorded. The recording will be used for our reference only and will be erased once the research report is complete. Your feedback is important to us, but your identity is not. Our reports to the Ohio Department of Education will not include actual names of participants, so your individual comments will be strictly confidential. Should you feel uncomfortable at any time during the discussion, remember that your participation is voluntary.

Distribute Consent Document so they can review it briefly.

Does anyone have a concern about this procedure? (Wait for responses)

Start the recorder

Before we begin, I would like us to get to know one another a little better. You may know each other – but I am new to the group … so it would be very helpful if you all could introduce yourselves. Please share your first name, and a little bit about your teaching position: grade-level, subject area, how long you've been teaching at your current grade level and the building where you teach.

Now, for the first few times you respond I would like to ask you to please repeat your first name and the subject and grade you teach. This will allow us to tease out particular information from the discussion transcripts across the districts we are meeting with. For example, there may be different opinions or experiences among teachers of specific grade levels or subjects and we want to be able to report those differences.

1. What are your general thoughts about Student Growth Measures being a part of teacher evaluation?

2. Let's talk specifically about Value-Added data.
    a. What do you think about Value-Added data used for teacher evaluation?
    b. Are there any challenges to using Value Added data in teacher evaluations?
    c. What about any benefits?

3. Now let's focus on Student Learning Objectives and teacher evaluation.
    a. What do you think about Student Learning Objectives being used in teacher evaluation?
    b. Are there any challenges to using Student Learning Objectives for teacher evaluation?
    c. Any benefits?

4. Now let's talk about Vendor Assessments.
    a. What do you think about Vendor Assessments being used for teacher evaluation?
    b. Are there any challenges to using vendor assessments in teacher evaluation?
    c. Any benefits?

5. How is the percentage weight for value-added data, as a portion of your evaluation determined?
    a. Who makes the decision?
    b. Were teachers involved in the decision? In what capacity?
    c. What was the conversation like when the district was choosing weights on the student growth measures components?

6. What are your thoughts about shared attribution?
    a. Is shared attribution used in the district and why or why not?
    b. What did the district decide was going to count as shared attribution?
    c. Was this a simple decision or was there a lot of discussion leading up to the final decision? What was that discussion like?

7. What do you think about the new teacher performance rubric?
    a. How well (or not) does it mesh with the student growth measure side of your evaluation?

Wrap-Up
If not addressed previously ask: What are your thoughts about the fact that what makes up SGM portion of teachers evaluations can vary depending upon the teacher's subject or grade level?

8. Are there any other comments or concerns you would like to share with me? Was there a question that you thought I would ask that I didn't ask?

This concludes our discussion. Your contribution is extremely helpful as we gather data for ODE regarding implementation of Student Growth Measures. On behalf of the Ohio Education Research Center, I would like to thank you all for participating.

D. Student Growth Measures Policy and Practice - Teacher Discussion Topic Guide

Category B & C Teachers (Do not have Value-Added data)

Hello everyone and thank you for allowing me to talk with you today. My name is _____ and I work for _____.

First, it may be helpful for me to explain a little bit about why I am here to speak with you. The Ohio Department of Education has contracted with the Ohio Education Research Center, which we at Ohio University's Voinovich School are a part of, to conduct research regarding implementation of Ohio's Teacher Evaluation System as it pertains to Student Growth Measures. We are meeting with all of the districts who received pilot grant funding in 2012 to implement SGMs, which includes your district.

It is important to remember that this is information gathering from all of the pilot districts. We are not assessing your implementation of SGMs or your district's implementation of SGMs. There will be no district-level reports written. It is also important for you to understand that I do not work for the Ohio Department of Education and, as such, I am unable to answer any questions about Student Growth Measures.

The purpose of this focus group today is to collect your thoughts and feelings about SGMs and we hope this discussion can help us gain insight into how SGMs are being implemented at the local level.

Another important thing to remember during our conversation is that everyone's ideas are important, and they should be allowed to freely express their thoughts and feelings. The ideas expressed here may be personal and should not be used against anyone inside or outside of this meeting. From time to time, I may interrupt to allow someone to speak who may not have said anything for a while. Also, I may have to interrupt someone to move on to another question because of the time limit under which we are working. I apologize in advance if this happens.

The discussion will be digitally recorded. The recording will be used for our reference only and will be erased once the research report is complete. Your feedback is important to us, but your identity is not. Our reports to the Ohio Department of Education will not include actual names of participants, so your individual comments will be strictly confidential. Should you feel uncomfortable at any time during the discussion, remember that your participation is voluntary.

Distribute Consent Document so they can review it briefly.

Does anyone have a concern about this procedure? (Wait for responses)

Start the recorder

Before we begin, I would like us to get to know one another a little better. You may know each other – but I am new to the group … so it would be very helpful if you all could introduce yourselves. Please share your first name, and a little bit about your teaching position: grade-level, subject area, how long you've been teaching at your current grade level and the building where you teach.

Now, for the first few times you respond I would like to ask you to please repeat your first name and the subject and grade you teach. This will allow us to tease out particular information from the discussion transcripts across the districts we are meeting with. For example, there may be different opinions or experiences among teachers of specific grade levels or subjects and we want to be able to report those differences.

9. What are your general thoughts about Student Growth Measures being a part of teacher evaluation?

10. Let's talk specifically about Student Learning Objectives and teacher evaluation.
    a. What do you think about Student Learning Objectives being used in teacher evaluation?
    b. Are there any challenges to using Student Learning Objectives for teacher evaluation?
    c. Any benefits?

11. Now let's talk about Vendor Assessments.
    a. What do you think about Vendor Assessments being used for teacher evaluation?
    b. Are there any challenges to using vendor assessments in teacher evaluation?
    c. Any benefits?

12. Now let's focus on Value-Added data.
    a. What do you think about Value-Added data used for teacher evaluation?
    b. Are there any challenges to using Value Added data in teacher evaluations?
    c. What about any benefits?

13. How is the percentage weight for value-added data, as a portion of teacher evaluation determined?
    d. Who makes the decision?
    e. Were teachers involved in the decision? In what capacity?
    f. What was the conversation like when the district was choosing weights on the student growth measures components?

14. What are your thoughts about shared attribution?
    a. Is shared attribution used in the district and why or why not?
    b. What did the district decide was going to count as shared attribution?
    c. Was this a simple decision or was there a lot of discussion leading up to the final decision? What was that discussion like?

15. What do you think about the new teacher performance rubric?

16. and How well (or not) does it mesh with the student growth measure side of your evaluation?

Wrap-Up
If not addressed previously ask: What are your thoughts about the fact that what makes up SGM portion of teachers evaluations can vary depending upon the teacher's subject or grade level?

17. Are there any other comments or concerns you would like to share with me? Was there a question that you thought I would ask that I didn't ask?

This concludes our discussion. Your contribution is extremely helpful as we gather data for ODE regarding implementation of Student Growth Measures. On behalf of the Ohio Education Research Center, I would like to thank you all for participating.