The Impact of the Relationship Between OTES and OPES on the Implementation of Principal and Teacher Evaluations
How Are Ohio’s New Principal and Teacher Evaluations Impacting Schools?

Suzanne Franco, EdD, Wright State University
Jill Lindsey, PhD, Wright State University
Ted Zigler, EdD, Wright State University

Teachers and Leaders
#FR-2014-01
oerc.osu.edu
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

This case study investigated the implementation of the new Ohio educator evaluation systems and the impact of the relationship between the Ohio Teacher Education System (OTES) and the Ohio Principal Evaluation System (OPES) on implementation. In both evaluation systems, student growth measures make up 50% and performance on a standards-based rubric make up the other 50% of the final evaluation rating.

Purposeful sampling was used to recruit twelve LEAs, eleven representing the seven Race to the Top (RttT) regions and Ohio Department of Education defined typologies, and one LEA not participating in RttT. Superintendents of the LEAs were contacted by email to enlist participation. Follow-up phone interviews with the Superintendents or their designated representative provided details about the status of implementation of OTES and OPES during the 2012-2013 academic year. Follow-up site visits were conducted in the two LEAs piloting/implementing OTES with all teachers. Site visits included focus groups with teachers and administrators to provide detailed data collection regarding implementation and the relationship between OTES and OPES.

Of the twelve LEAs in the sample, eight were piloting and one was implementing OTES. Five LEAs were piloting and one was implementing OPES. Two LEAs, one RttT and the non-RttT, were not piloting or implementing either new evaluation system. Ten superintendents reported that OPES implementation was behind OTES implementation because there were too many changes happening all at the same time and too many demands on principals.

Content analysis of principal interview and teacher focus group transcripts revealed common themes related to: training, implementation decisions, responses to evaluation systems, impacts on school culture, issues/concerns, and misunderstandings/confusion about SGM. The sample of evaluation documents did not provide the desired information about which forms were used, completeness and types of information included in evaluations. At the time this study was completed, none of the LEAs had finalized the measures and percentages for the student growth measures component of OPES for 2013-14. Five of the twelve LEAs had measures and weights under consideration for the SGM measures component for OTES for 2013-14, and those are included in this report.

Key findings included, sample LEAs were not fully implementing OTES or OPES nor using student growth measures in evaluations for 2012-13; generally positive feedback characterized by a “we are in this together” approach; educators perceived the time demands for completing evaluations made the process unrealistic and unsustainable for larger buildings; educators offered evidence of competition and perceived unfairness due to variations in how student growth is measured across grade levels and subjects. Teachers expressed many misunderstandings and questions related to student growth measures, and feel they need data experts, resources, and examples to better understand the various student growth measures.
Funding for this report was provided by the Ohio Education Research Center.

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

**Project Team**
Suzanne Franco, PhD, Ohio Education Research Center/Wright State University
Jill Lindsey, PhD, Ohio Education Research Center/Wright State University
Ted Zigler, EdD, Ohio Council of Professors of Educational Administration/Ohio Dominican University

Questions regarding this report should be directed to Dr. Suzanne Franco at Suzanne.Franco@wright.edu or 513-317-9415.
CONTENTS

Contents ....................................................................................................................................................................................... 1
Executive Summary ................................................................................................................................................................ 2
I. The Problem ........................................................................................................................................................................... 4
II. Methods .................................................................................................................................................................................. 6
III. Findings ................................................................................................................................................................................ 7
   Implementation .................................................................................................................................................................. 7
   Preparation and Experiments ....................................................................................................................................... 8
   Student Growth Measurements ................................................................................................................................... 8
IV. Summary ................................................................................................................................................................ ............. 10
EXECUTIVE SUMMARY

This case study investigated the implementation of the new Ohio educator evaluation systems and the impact of the relationship between the Ohio Teacher Education System (OTES) and the Ohio Principal Evaluation System (OPES) on implementation. Both OTES and OPES were designed to be research based, transparent, fair and adaptable to the specific contexts of Ohio’s districts (rural, urban, suburban, large, and small). In both evaluation systems, student growth measures make up 50% and performance on a standards-based rubric make up the other 50% of the final evaluation rating.

Purposeful sampling was used to recruit twelve LEAs, eleven representing the seven Race to the Top (RttT) regions and Ohio Department of Education defined typologies and one LEA not participating in RttT. (An expanded typology system was adopted mid-study. New typology codes are presented but all new typologies are not represented in the sample). Superintendents of LEAs that met the purposeful sampling parameters were contacted by email to enlist participation. Follow-up phone interviews with the superintendents or their designated representative provided details about the status of implementation of OTES and OPES during the 2012-2013 academic year. Follow-up site visits were conducted in the two LEAs piloting/implementing OTES with all teachers. Site visits included focus groups with teachers and administrators to provide detailed data collection.

Of the twelve LEAs in the sample, eight were piloting and one was implementing OTES. Five LEAs were piloting and one was implementing OPES. Two LEAs, one RttT and the non-RttT, were not piloting or implementing either new evaluation system. The one LEA implementing OTES with all teachers did not include Student Growth Measures in the evaluation process. One LEA was piloting OTES with Student Growth Measures with all teachers but the final evaluation ratings were not used as official evaluations of record. The remaining seven LEAs were piloting OTES with a small sample of teachers. The six LEAs piloting or implementing OPES did not include Student Growth Measures. One LEA was implementing OPES (without SGM) as the official evaluation of record.

Ten superintendents reported that OPES implementation was behind OTES implementation because there were too many changes happening all at the same time and too many demands on principals. They felt it was important to get OTES accepted and processes in place before adding the demands of OPES. One LEA made the decision to implement OPES first and delayed implementation of OTES for similar reasons - too many changes all at once and too many state initiatives. The two LEAs not piloting or implementing either evaluation system reported they were waiting as long as possible before adopting the new systems due to too many changes happening all at the same time and legislative uncertainty related to the Ohio educator evaluation systems.

Content analysis of interview and focus group transcripts revealed common themes related to: training, implementation decisions, responses to evaluation systems, impacts on school culture, issues/concerns, and misunderstandings/confusion about SGM. The sample of evaluation documents did not provide the desired information about which forms were used, completeness and types of information included in evaluations. At the time this study was completed, none of the LEAs had finalized the measures and percentages for the student growth measures component of OPES for 2013-14. Five of the twelve LEAs had measures and weights under consideration for the SGM measures component for OTES for 2013-14. Three LEAs were planning to use Shared Attribution and all five LEAs were developing Student Learning Objectives (SLOs) for teachers in non-tested grades and subjects.
Key findings include sample LEAs not fully implementing OTES or OPES nor using student growth measures in evaluations for 2012-13; generally positive feedback characterized by a “we are in this together” approach; time demands for completing evaluations made them unrealistic and unsustainable for larger buildings; evidence of competition and perceived unfairness due to variation in how student growth is measured across grade levels and subjects. Teachers also expressed many misunderstandings and questions related to student growth measures, and are in need of data experts, resources, and examples to understand the various student growth measures.
I. THE PROBLEM

New educator evaluation systems, the Ohio Principal Evaluation System (OPES) and the Ohio Teacher Evaluation System (OTES), were developed and adopted in response to the 2009 Ohio House Bill 1. Anecdotal communications with area school districts suggested there may be a relationship between OTES and OPES that was influencing implementation, particularly in relation to the required use of student growth measures. The purpose of this case study was to investigate the following three topics:

1. Implementation of OTES and OPES (or locally equivalent approved teacher and principal evaluation systems) regarding sequencing, planning, feedback and student growth metrics
2. Preparation for evaluation and experiences of teachers and principals evaluated using the new evaluation systems
3. Processes and metrics adopted to measure student growth in teacher and principal evaluations

The Ohio Principal Evaluation System (OPES) and the Ohio Teacher Evaluation System (OTES) were developed by the Ohio Educator Standards Board in collaboration with Ohio superintendents, building administrators, higher education faculty, representatives from professional educator associations, and national experts in the field. Both systems were designed to be research-based, transparent, fair and adaptable to the varied (rural, urban, suburban, large, and small) contexts of Ohio's districts. The systems provide educators with a richer and more detailed view of their performance, and focus on specific strengths and opportunities for improvement.

The systems are comprised of two components: 50% is a Professional Performance rating determined by performance on specific goals, observations, and walk-throughs using a standards-based performance rubric; and 50% is a student academic growth rating. Measuring student growth is problematic in that Ohio does not have a unified, progressive assessment system. The state currently provides value-added data from the Ohio Achievement Assessments for students in some subjects in grades 4-8. Ohio value-added scores are calculated using the EVAAS Value-Added methodology, provided by SAS, Inc. The EVAAS methodology incorporates student test histories (across all state-tested subjects) in determining growth metrics. EVASS effectiveness reports are produced for individual students, teachers, grades by subject, buildings, district, aggregate composite reports, and three-year average reports.

For state tested and non-tested grades and subjects, schools may also use national vendor assessments that have been approved by the Ohio Department of Education. For subjects without state assessments or approved vendor assessments, districts must identify local measures using state designed guidance and criteria, such as developing Student Learning Objectives (SLOs) or Shared Attribution (student growth measures attributed to a district or group of buildings to encourage collaborative goals).

In 2010, Ohio’s Race to the Top application was funded. The RttT application included the phasing in of the new teacher and principal evaluation systems. Ohio Race to the Top (RttT) districts were required to indicate in their RttT proposal a timeline for implementing the new educator evaluation systems. All Ohio Districts must use OTES/OPES or a comparable evaluation system approved by the Ohio Department of Education. Ohio HB 153 (2011) required all districts to include student
growth measures as 50% of the evaluation of principals and teachers for the 2013-14 academic year. Six months into data collection for this case study, the Ohio Legislature passed HB 555, which required that the percentage of value-added measures (EVASS) included in the 50% Student Growth Measures (SGM) component of evaluation be proportionate to a teacher's schedule of tested subjects. It added the requirement that approved vendor assessment data must be included when available. Since, districts were not required to include SGMs in teacher or principal evaluation systems until the 2013-14 academic year, many districts delayed making SGM weights and measure decisions until the end of the 2012-13 school year. At the time this study was completed, none of the sample LEAs had finalized the measures and percentages for the SGM component of OPES for 2013-14. Five LEAs shared their proposed SGM measures and percentages for OTES and those are included in this report.
II. METHODS

A constructivist model (Guba & Lincoln, 1989) involving guided interviews with district personnel as well as interviews with teachers and principals were implemented in addition to document analyses. Purposeful sampling was used to recruit twelve LEAs, eleven representing the seven Race to the Top (RttT) regions and Ohio Department of Education defined typologies and one LEA not participating in RttT. (An expanded typology system was adopted mid-study. New typology codes are presented but all new typologies are not represented in the sample). Superintendents of LEAs that met the purposeful sampling parameters were contacted by email to enlist participation. Follow-up phone interviews with the superintendents or their designated representative provided details about the status of implementation of OTES and OPES during the 2012-2013 academic year. Follow-up site visits were conducted in the two LEAs piloting/implementing OTES with all teachers. Site visits included focus groups with teachers and administrators to provide detailed data collection. The research team requested redacted copies of completed OTES and OPES evaluation forms, but none of the LEAs released completed evaluation forms. One LEA provided a sample of redacted summary rating sheets used to compute final ratings. These will be discussed in the Findings section.

Guided interviews were conducted, recorded and transcribed. The interview protocol elicited educators’ views and experiences related to:

1. Training provided prior to using the systems
2. Components and forms used
3. Approach to using the systems; challenges in using the models
4. Types of documentation principals and teachers were asked to provide
5. Documentation principals and teachers provided
6. Student growth measures included in the 50% SGM component
7. How student growth measures were identified
8. Perceptions about whether the systems were helpful in developing leaders and teachers, advancing communication, improving schools and student learning
9. Relationship between the two models regarding student performance metrics, implementation sequencing and timeline.

Low-inference content analysis was conducted on focus group and interview transcripts. Common themes were identified related to training, implementation decisions, responses to evaluation systems, impacts on school culture, and misunderstandings/confusion related to SGM. These findings are presented below.
III. FINDINGS

Interviews and focus group responses provided rich data regarding training, implementation and perceptions of the models. Evaluator, principal and classroom teacher voices provided a breadth of feedback from multiple perspectives including perceptions of the evaluation systems' usefulness for improving schools and student learning. Common themes were identified related to: implementation decisions; training; responses to the evaluation systems; negative impacts on school culture; and misunderstandings or confusion about student growth measures (SGM).

Unfortunately, analysis of the sample of evaluation documents for principals and teachers did not provide the desired information about which forms were used, completeness, types of information included in evaluations, or how rubric ratings were determined.

The following are common findings across LEAs in this study:

1. Sample LEAs were not fully implementing OTES or OPES and none included student growth measures in evaluations for 2012-13.
2. Teachers and administrators offered generally positive feedback characterized by a “we are in this together” approach.
3. Teachers and principals felt the OTES/OPES systems were designed well and would encourage conversations about teaching and learning, but the time demands for completing evaluations made them unrealistic and unsustainable for larger buildings.
4. Teachers and principals expressed concerns about competition and unfairness in how student growth is measured across grade levels and subjects.
5. Teachers expressed misunderstandings and questions related to value-added and student growth measures.
6. Educators feel they need data experts, resources, and examples to help them better understand the various student growth measures being used in evaluation.

These findings are presented under the three broad areas investigated: Implementation, Preparation and Experiences, and Student Growth Measures.

Implementation

Of the twelve LEAs in the sample, eight were piloting OTES and one was implementing. Five of the twelve LEAs were piloting OPES and one was implementing. Two LEAs, one RttT and the non-RttT, were not piloting or implementing either new evaluation system. None of the sample LEAs were using student growth measure (SGM) as part of the final evaluation for educators in 2012-13. Seven LEAs were piloting OTES with a small sample of teachers and six LEAs were piloting or implementing OPES. One LEA was piloting OTES using SGM for all teachers but those evaluation ratings were not used as the official evaluations of record for the school year, and one LEA was implementing OPES without SGM as the official evaluation of record.

Ten superintendents reported that OPES implementation was behind OTES implementation because there were too many changes happening all at the same time and too many demands on
principals. They felt it was important to get OTES accepted and processes in place before adding the demands of OPES. One LEA made the decision to implement OPES first and delayed implementation of OTES for similar reasons - too many changes all at once and too many state initiatives. The two LEAs not piloting or implementing either evaluation system reported they were waiting as long as possible before adopting the new systems due to too many changes happening all at the same time and legislative uncertainty related to the Ohio educator evaluation systems.

**Preparation and Experiments**

Principals received formal training and exit testing in order to be credentialed to evaluate teachers using the OTES. However, few had been formally trained on the OPES. Similarly, few teachers had received formal training on OTES. In most cases, district administrators provided basic information about OTES at faculty meetings so teachers were generally aware of the components.

Educators provided generally positive feedback about both systems indicating that the Goal Setting process was useful, the evaluation process provided opportunities for discussion about teaching and student growth that would be helpful, and that the system seemed designed to help educators focus on specific areas for professional development and improvement. These comments were focused on the use of the standards-based rubric which educators felt accurately represented the work of teachers and administrators. Educators expressed acceptance of the idea that student growth was an important component of evaluation and that because both principals and teachers were going to be evaluated using student growth measures they felt “we are in this together.” However, educators also expressed great concern that the time demands required to use all components of the evaluation systems and to complete multiple classroom observations/walk-throughs were unrealistic and unsustainable, especially in larger buildings. Many teachers expressed concern that their principal was not as visible and available to them or students because of the rigorous schedule of required classroom observations and evaluation conferences. Educators also expressed concern that the OTES/OPES documentation and observation processes requires them to focus on and overemphasis on themselves rather than focusing on students and their learning. Many teachers described a feeling that the school culture was shifting away from collaboration toward fear of sharing good ideas and competition for high ratings. Teachers’ concerns about these changes in school culture were expressed with great emotion and anxiety.

Another issue of great concern centered on the public perceptions associated with the color-codes and labels of the rating categories. The highest possible rating was labeled Accomplished and color-coded green rather than the preferred blue as in a blue ribbon, and the second highest rating was labeled Proficient with yellow as the color code, suggesting warning rather than capable performance. As this report was being finalized, the OTES rating system labels and colors were changed. The Proficient label was changed to Skilled and color-coded pale green. The Developing label was maintained but the color was changed to a deep yellow.

**Student Growth Measurements**

Less easily resolved issues focused on the validity and fairness of using various student growth measures in the evaluation systems. Principals and teachers expressed concerns about the effect of the system using prior year value-added scores in evaluation; the use of the previous year’s student scores can be demoralizing, thus undermining current year efforts.

Teachers expressed misunderstandings and confusion about the calculation and validity of value-added and alternative growth measures. Validity was also a concern related to the use and administration of vendor assessments, particularly those identified as providing growth measures
that may or may not be calculated using EVAAS methodologies. Teachers and principals identified differences in required documentation, within their own building or districts, related to Student Learning Objectives and the apparent unfairness of some teachers designing their own growth measures while others were required to use state testing scores. Others raised the issue of a lack of equity related to the inability to correlate the various types of growth measures that were going to be used as interchangeable measures of teacher performance. Options like Shared Attribution evoked mixed responses with some teachers and principals feeling it offered a way to reward team effort where others felt shared attribution put too much pressure on the teachers in tested grades to have their scores be shared by others when salary increases and contract renewals hang in the balance. And another sector felt that team accountability would not always reflect a teacher’s contribution to student learning and would allow less capable or lazy teachers to receive higher than appropriate evaluations.

One particularly interesting finding was that principals and teachers in some LEAs were simply unaware of some of the student growth measure options, let alone able to conceptualize the SGM weighting options available for the three main categories of teachers being evaluated. The SGM types include: value-added scores calculated using student testing history and recent test performances; vendor assessments; Student Learning Objectives (SLOs); and Shared Attribution using building value-added, district value-added, building performance index, or district performance index.

None of the districts had formalized their SGM weights for the 2013-14 evaluation cycle. However, five of the twelve LEAs had SGM weights under consideration. Three of the five LEAs were planning to use Shared Attribution and six of the twelve LEAs were developing SLOs. The proposed plans documented wide variations in weights and measures supporting the concerns raised about whether such variation enabled comparisons of teacher performance.
IV. SUMMARY

This study investigated the implementation of the new Ohio educator evaluation systems and the impact of the relationship between the Ohio Teacher Education System (OTES) and the Ohio Principal Evaluation System (OPES) on implementation. Purposeful sampling was used to recruit twelve LEAs, eleven representing the seven Race to the Top (RttT) regions and Ohio Department of Education defined typologies and one LEA not participating in RttT. Superintendents of LEAs that met the purposeful sampling parameters were contacted by email to enlist participation. Follow-up phone interviews with the superintendents or their designated representative provided details about the status of implementation of OTES and OPES during the 2012-2013 academic year. Site visits and focus groups with teachers and administrators were conducted in the two LEAs in the sample that were piloting or implementing OTES with all teachers without using student growth measures.

Ten superintendents reported that OPES implementation was behind OTES implementation because there were too many changes happening all at the same time and too many demands on principals. One LEA made the decision to implement OPES first and delayed implementation of OTES for similar reasons. The two LEAs not piloting or implementing either evaluation system reported they were waiting as long as possible before adopting the new systems due to too many changes happening all at the same time and legislative uncertainty related to the Ohio educator evaluation systems.

Content analysis of interview and focus group transcripts revealed common themes. The sample of evaluation documents did not provide the desired information. At the time this study was completed, none of the LEAs had finalized the measures and percentages for the student growth measures component of OPES for 2013-14. Five of the twelve LEAs had measures and weights under consideration for the SGM measures component for OTES for 2013-14. Three LEAs were planning to use Shared Attribution and all five LEAs were developing Student Learning Objectives (SLOs) for teachers in non-tested grades and subjects.

Key findings included, sample LEAs were not fully implementing OTES or OPES nor using student growth measures in evaluations for 2012-13; generally positive feedback characterized by a “we are in this together” approach; educators perceived the time demands for completing evaluations made the process unrealistic and unsustainable for larger buildings; educators offered evidence of competition and perceived unfairness due to variations in how student growth is measured across grade levels and subjects. Teachers expressed many misunderstandings and questions related to student growth measures, and feel they need data experts, resources, and examples to better understand the various student growth measures.