OREGON EDUCATION ASSOCIATION
CENTER FOR GREAT PUBLIC SCHOOLS

Teacher Evaluation & Support System Guidebook

Tools for Building an Effective, Collaboratively-Designed Evaluation System
ACKNOWLEDGMENTS

The Center for Great Public Schools wishes to thank a number of individuals who have helped develop the foundation for this guidebook, which had its genesis in conversations among Oregon Education Association leaders, staff and other education stakeholders beginning in 2010.

Individuals we wish to acknowledge with gratitude are: OEA President Gail Rasmussen, OEA Vice President Hanna Vaandering, and (in alphabetical order), Sam Aley, Susan Anderson, Kristie Buckley, Daniel Burdis, Colin Cameron, Andrea Cooper, Tony Crawford, Linda Darling-Hammond, Dave Fiore, Nancy Golden, Anne Goff, Gary Humphries, Tom Husted, Henry Kim, Kathi Koenig, Kevin Mechlin, Colleen Mileham, Mark Molner, Jen Murray, Theresa Richards, Hilda Roselli, Richard Sanders, Eric Schutz, Monica Smith, Terrel Smith, Gwen Sullivan, Judy Svoboda, Joe Swinehart, Becca Uherbelau, Courtney Vanderstek, David Wilkinson, Colleen Works, and Robert Young.

Center of Great Public Schools staff Erin Whitlock, Teresa Ferrer, Center Director Lindsey Capps, and OEA’s Editor and Designer Meg Krugel are deserving of special acknowledgment for their research, writing, organizing and design of this Guidebook.
# Table of Contents

**Executive Summary**

**Getting Started**
- 7 Principles for Success
- Union Leadership & Teacher Evaluation
- Collaboration as Key for the Development Process
- ESEA Waiver: Influencing Oregon’s Framework for TESS
- New Oregon Evaluation & Support Requirements
- Timeline, Process Checklist & Key Questions

**Tools**
- Design Process Checklist

**Standards**
- Getting to Know the InTASC Standards
- The InTASC Teaching Standards at a Glance

**Multiple Measures**
- Measures of Professional Practice
- Measures of Professional Responsibilities

**Student Learning Goals**
- Student Learning and Growth as a Measure of Teaching Effectiveness
- Student Learning and Growth Goal-Setting Process

**Using Measures of Student Learning**
- OEA's Recommended Use of Student Data
- Measures of Student Learning
- Using Measures of Student Learning as a Measure of Teaching Effectiveness

**Value Added Models**
- OEA's Analysis of Usage of Value-Added Models
- Shared Principles for Use of Value-Added Models

**Evaluation Procedures**
- Evaluation and Professional Growth Cycle
- Peer Assistance
- Teacher observation approaches

**CD-ROM Resources**
- Alternative Compensation Resources
- Bargaining Resources
- Collaboration Resources
- Communication Resources
- Glossary
- InTASC Resources
- Multiple Standards Measures Resources
- NEA Resources
- OEA Resources
- Research
- Rubric Resources
- Sample Frameworks
- Sample TESS
- Teacher Evaluation OARs
- References
Oregon Education Association’s Teacher Evaluation and Support System Guidebook has been developed with the goal of providing local education association and school district leaders a comprehensive resource to utilize in working collaboratively together in designing a teacher evaluation and support system. Think of this Guidebook as an owner’s manual for assembling and operating your new system. Whether your local association is just beginning this new journey or is well on the path toward a newly aligned teacher evaluation and support system, this Guidebook can prove useful in many ways.

Of course, before getting to the assembly and operation of your system, it is important to consider the purpose of that system. OEA and our educator members believe the purpose of any educator evaluation system is to support ongoing professional learning, growth and collaboration with the goal of continuous improvement in teaching and learning. Therefore, this Guidebook is intended to empower educators to lead the way in creating evaluation and support systems that are holistic and transparent with clear focus on and alignment to high-quality, research-based standards of practice.

What you’ll find in this Guidebook:

- State and federal requirements for educator evaluations
- Model core teaching standards
- Measures and evidence of effective teaching
- Research and resources to ensure evaluation systems are valid, reliable and fair
- Tools to guide you in the design process, including models for collaboration and consensus-driven decision-making
- Critical questions to keep in mind as you design, pilot and implement your new teacher evaluation and support system.
Executive Summary

The Oregon Education Association believes that every student in our public schools deserves competent, caring and effective teachers. A sound teacher evaluation and support system is one tool to achieve this goal.

In the spring of 2010, the Oregon Education Association convened a special workgroup to discuss the role of evaluation in promoting teacher quality and student success (see enclosed CD-ROM). The group — representing local education association leaders and UniServ staff within OEA, together with school district leadership, higher education, and other stakeholders — developed a white paper to guide local and state conversations on teacher evaluation. A key finding of the workgroup was that teacher evaluation cannot exist in isolation and must be part of an aligned continuum of standards and supports from pre-service, hiring, assignment, induction, mentoring, professional growth and development, and teacher leadership.

The work group concluded that teacher evaluation must be tied to a strong, extended system of support for teachers providing impactful opportunities for ongoing professional learning and collaboration across the career continuum — from pre-service, induction, and mentoring, to ongoing professional growth and development, and teacher leadership. Additionally, the work group recognized the importance of an evaluation system being supported by the teaching and learning conditions at play in their classrooms, schools, and communities. Teacher evaluation is one part of a whole system that impacts student learning and all parts of the system must be improved for students to flourish.

For a teacher evaluation and support system to be effective it must be a holistic, valid and reliable tool. As will be demonstrated in more detail throughout this Guidebook, the proper design of an evaluation system is critical. If an evaluation and support system is to enhance student learning, then it must be focused on improving the knowledge, skills and classroom practice of professional educators based on established standards of teaching, customized collaboratively by the locally-based design teams. Focused standards alone will not elevate teaching and learning without an aligned system of professional learning and support that is robust, collaboratively designed and improved by teachers and districts working together toward that common goal.

An evaluation system must also be grounded within a professional culture where teachers — as colleagues, team members, mentors or master teachers — are continuously engaged in purposeful goal setting, professional learning and collaboration. As part of this professional culture, teachers and other educators should be engaged as collaborative partners in the design of all aspects of the evaluation system.

The evaluation process itself requires training, resources, and time for observation, analysis, and goal setting that is collaborative and individual to each education professional. This goes hand-in-hand with the evaluation system itself being strongly aligned to individual professional practice and ongoing, job-embedded professional development for all areas. Ultimately, this requires sustained leadership and commitment — financial, social and political — from the school district, community, the state and elected leaders at all levels.

As school district teams of teachers, principals, other district administrators and the exclusive bargaining representatives work together to collaboratively develop evaluation systems around these core propositions, the result will not just be better evaluation systems. These collaborative processes and resulting evaluation and support systems can create better professional learning environments where teaching and learning — and our students — thrive.

ReDesign of Teacher and Administrator Evaluation Systems Required by Law

In the summer of 2011, Oregon Governor John Kitzhaber signed into law Senate Bill 290, legislation supported by the Oregon Education Association and passed with bipartisan majorities in both chambers of the Oregon Legislature. Almost a full year later in July 2012, the US Department of Education approved Oregon’s request for a waiver to the federal Elementary and Secondary Education Act (ESEA) and No Child Left Behind law.

SB 290 and the provisions of Oregon’s ESEA waiver related to educator effectiveness are predicated upon long-established evidence demonstrating that a teacher is the most important in-school factor impacting a student’s success. SB 290 and the Oregon waiver together establish new state policies aimed at improving the quality of both teachers and school leaders through new requirements and standards for evaluation and support systems. Both SB 290 and Oregon’s waiver will guide the work of locally-based collaborative design teams as they design their evaluation and support systems.

Senate Bill 290: Collaborative Design of Standards-Based Evaluation Systems with Multiple Measures of Performance

Senate Bill 290 directed the Oregon State Board of Education to adopt standards for teacher and administrator evaluations. In December 2011, the State Board of Education adopted model core teaching standards developed by the Interstate Teacher Assessment and Support Consortium, known as the InTASC Standards. The State Board adopted standards established by the Interstate School Leaders Licensure Consortium, or the ISLLC Standards, for all administrator evaluations.

The standards adopted for both teachers and administrators are high-quality, research-based standards that reflect what an educator should know and be able to do.

In addition to setting forth statewide standards of practice for teachers and administrators, Oregon’s Senate Bill 290 has three key requirements for school districts and education service districts:
Teacher evaluation systems must be designed collaboratively with teachers and their exclusive bargaining representative (local education association representing teachers in the district). [See Getting Started section for resources to assist in this process].

By July 2013, every school district and education service district in Oregon must align their teacher evaluation systems with model core teaching standards adopted by the State Board of Education [See Standards section for in-depth, detailed information on the State adopted InTASC Standards].

Aligning their teacher evaluation systems to state standards, and the application of standards across the evaluation and support system, school districts must create evaluation systems that take into consideration multiple measures of teaching effectiveness, and establish a formative growth process for each teacher that supports professional learning and collaboration with other teachers. [See Multiple Measures section for a discussion of measures to consider including in your district’s evaluation system; and how to integrate measures within your evaluation system].

OREGON’S ESEA WAIVER:
Comprehensive Framework for Evaluation Systems Integrating Standards and Multiple Measures

In providing Oregon the opportunity to obtain a waiver from the onerous provisions and punitive sanctions imposed under the federal No Child Left Behind law, the US Department of Education required that the State and local school districts “… commit to develop, adopt, pilot, and implement, with the involvement of teachers and principals, teacher and principal evaluation and support systems.”

In receiving approval of its waiver application from the US Department of Education in July 2012, Oregon committed itself to a set of basic guidelines for teacher and principal evaluation and support systems. These guidelines are reflected in the Framework for Teacher and Administrator Evaluation and Support Systems adopted by the State Board of Education and provided to all Oregon school districts in June 2012.

Oregon’s ESEA waiver expands on Senate Bill 290 with four additional requirements:

- School districts must comply with state framework for teacher and administrator evaluation and support systems. [See Framework for Teacher & Administrator Evaluation & Support section].
- Establishes 3-specific categories of measures to be included in teacher evaluation and support systems: Professional Practice, Professional Responsibilities and Student Learning and Growth. [See Multiple Measures section].
- Incorporate student growth as a “significant factor” in individual teacher evaluations, requiring school districts to provide teachers the opportunity for individual goal setting around student learning. [See Measures of Student Learning section for strategies to meet this requirement in ways that are meaningful and educationally-relevant based on subject area and assignment.] ODE will be doing a pilot year in 2012-13 to determine what “significant” will mean in the course of teacher and administrator evaluations; more information about this factor will be forthcoming in the spring of 2013.

Four-levels of proficiency in assigning a summative rating of performance to each individual teacher.

Under the waiver, Oregon committed itself to a pilot year for 2012-2013 in which 50 sites will be participating. Schools involved in the federal School Improvement Grant (SIG) program, the Teacher Incentive Fund (TIF), or the state School District Collaboration Grant Program will be potential candidates, and other sites may be allowed to apply. These pilots will integrate the student learning goals [see the Using Measures of Student Learning section] into their evaluation systems and determine what “significant” means per the ESEA Waiver criteria. Furthermore, school districts not participating in the ODE pilot year, have until the 2013-2014 school year to pilot their evaluation system in compliance with SB 290 and the waiver. [See additional ESEA waiver requirements and timelines in the Getting Started section].

Teacher Evaluation = Three “ Buckets” of Evidence

**Executive Summary**

**PROFESSIONAL PRACTICE**
Observations
Artifacts

**PROFESSIONAL RESPONSIBILITIES**
Self Assessment
Leadership Roles
Prof. Development
Family Engagement
Growth Plans

**STUDENT GROWTH & LEARNING**
Student learning goals based on multiple measures

---

**Executive Summary**

Teacher Evaluation = Three “ Buckets” of Evidence

**PROFESSIONAL PRACTICE**
Observations
Artifacts

**PROFESSIONAL RESPONSIBILITIES**
Self Assessment
Leadership Roles
Prof. Development
Family Engagement
Growth Plans

**STUDENT GROWTH & LEARNING**
Student learning goals based on multiple measures

---
Inside Oregon’s Framework: Triangulation of Multiple Measures to Support the Whole Teacher

Oregon’s Framework for Teacher and Administrator Evaluation and Support Systems established under the ESEA waiver is grounded in the concept of triangulation. Triangulation is used in many complex “human” fields where you cannot easily or accurately assign a quantitative number or value. The logic behind triangulation is that no single measure reveals sufficient or reliable information. Multiple data points on the other hand provide richer and more accurate results. Conclusions based on one measure are vulnerable to erroneous conclusions, while multiple pieces of evidence using different types of data are more likely to result in valid conclusions.

So how does triangulation work in practice? Triangulation requires a method or measure of teaching effectiveness to be validated by others, leading to greater confidence in the overall feedback an educator receives. Different methods and measures are designed to validate one another and support the same conclusion. This process will lead to teachers and their evaluators having greater confidence in evaluation system while reducing subjectivity.

When agreement is reached between measures from the three different categories required under Oregon’s Framework, there is greater confidence that the original assessment was accurate. If the data sources differ, that triggers a need for the evaluator to assess the training system in place to train evaluators on the measure, the measure itself, and/or the educator again more deeply to figure out why results varied. Triangulation also addresses the limitations of each source of information. Judgments about practice are limited by subjectivity and bias, while judgments based on student performance are limited by technical considerations.

Adapted from Massachusetts Teacher Association
7 Principles for Success

OEA believes that seven key principles must serve as the foundation for developing or reforming any Teacher Evaluation and Support System (TESS):

1. **Safe and open collaboration** is necessary. When assessment of teacher practices is transparent and openly collaborative, teachers can build professional communities and learn from one another. This process can occur only in non-threatening environments of formative assessment and growth.

2. **Measures of teacher performance** are most helpful and meaningful when they are based on multiple valid ratings and clear teaching standards. Teachers need clear and actionable feedback based on standards for teaching and student learning that are comprehensive and transparent, and on criterion-referenced assessments of teacher practice. Feedback is most useful as part of a comprehensive teacher development system. Summative evaluations of teachers should use uniform criteria for effectiveness that are relevant for all teachers.

3. **Integrated systems must link** evaluation procedures with curricular standards, professional development activities, targeted support, and personnel decisions.

4. **Validated evaluation measures** are essential. Measures of teaching effectiveness need to be based on widely accepted standards of teaching that capture a range of teaching behaviors and use multiple valid evaluation methods.

5. **Teachers’ input in determining** performance and learning outcomes should be part of the system. Although standards for teaching practice and student learning are essential, each teacher should have an opportunity to help define a set of practices and student learning objectives to be assessed. Teacher input can provide vital learning goals for the unique, contextualized circumstances of each particular classroom.

6. **Teacher Evaluation and Support** Systems (TESS) need to be co-created or designed with teachers working through the local association. This may be the most important principle of all. Ideals and visions need to be balanced with local context, and political and financial reality. There is no one-size-fits-all solution at a state level. OEA will work with locals to craft local solutions based on the principles outlined here.

7. **Teacher Evaluation and Support** Systems are found to be more effective when:
   - they ensure that evaluators are well-trained
   - evaluation and feedback are frequent
   - mentoring and coaching are available
   - processes are in place to support due process and timely decision making by an appropriate body.

**UNION LEADERSHIP & TEACHER EVALUATION**

Student learning is at the center of everything a teacher does. OEA believes that effective teaching engages all students in the learning process; focuses on interactions and activities between teachers and students, and students with their peers; involves collaboration among teachers; centers on a continuous professional learning cycle where planning, practice, implementation, reflection, analysis, and modification of practice occur; and leads to growth in student knowledge, skills, and well-being. Evaluation of this process needs to be flexible, robust, and based on the value of continuous growth and improvement in the profession. OEA’s resolutions exemplify this value for our association, making advocacy for the active assessment of all teachers through regular and comprehensive evaluation procedures a goal to which we strive (*OEA Resolution V.14*).

Our members believe this needs to be implemented in a professional environment which respects the diversity of our students and workforce, and built from the perspective that evaluations should be fair and objective for all school employees, and should be developed by, and acceptable to, the association and the governing board in compliance with state law (*OEA Resolution V.22*). Additionally, OEA’s Core Values of lifelong learning, collaboration, respect for diversity, integrity, and professionalism insist that the evaluation systems we co-create exemplify these values in every aspect.

“OEA’s Core Values of lifelong learning, collaboration, respect for diversity, integrity, and professionalism insist that the evaluation systems we co-create exemplify these values in every aspect.”
Collaboration as Key for the Development Process

Collaboration is the standard for the design and implementation process. This means there needs to be some standardized methods of engaging in this work that create equal access, opportunity, and voice to all participants. OEA has some key recommendations in establishing a collaborative process to design and implement your TESS (see CD-ROM for activities and ideas suggested below):

1. **Establish a process for coming to a consensus-driven decision.**
2. **Develop group norms to guide and enhance** participants’ behaviors throughout the process.
3. **Establish a procedure for disagreements** (how to make a consensus-decision when there is NOT a consensus, such as bringing in an outside facilitator)
4. **Start with lower-stakes activities** when beginning with a consensus-based process; working collaboratively takes trust, and trust takes time to build over a continuous cycle where each “side” takes risks, is asked to follow-through on certain important tasks/commitments, then follows through with their responsibilities and is the recipient of the “other side’s” trust for having done so:
   a. Collaborate on a group definition of effective teaching.
   b. Collaborate on a group set of goals for the teacher evaluation and support systems.
   c. Come to consensus on a common vocabulary. There are many terms in developing a TESS that are misunderstood or misused (student achievement vs. student growth is one resounding example), so looking at the Glossary of Terms in the InTASC Standards is helpful. Making sure everyone is on the same page for what important terms mean is one way to build up trust and to prevent future misunderstandings.
   d. Come to consensus on a common understanding of the InTASC Standards, Oregon’s new Model Core Teaching Standards to which all teacher evaluation systems must be aligned by July 1, 2013.
**ESEA Waiver Frameworks**

The following is a summary of the ESEA Waiver requirements that must be a part of your Teacher Evaluation and Support System (TESS):

1. **Require the TESS be used** for continual improvement of instruction

2. **Require four performance level ratings** of effectiveness (see chart below)

3. **Require the TESS have multiple, valid measures** in determining performance levels
   a. This will include a rubric or performance indicator to clarify performance expectations for each of the ten InTASC Standards. The rubric must have the required four performance level ratings of effectiveness.
   i. Oregon Department of Education (ODE) will be conducting pilots in 2012-2013 to determine what “significant” means.
   b. If other frameworks are used, such as the Danielson frameworks, a “crosswalk” must be made to clarify which of the InTASC Standards are covered by each domain/criteria.
   b. This will include choosing at least one measure from each category of measures:
      i. Professional practice
      ii. Professional responsibilities and iii. Student learning
   c. This will also include a robust set of measures of student learning for all students as a “significant factor” in teachers and administrators’ evaluations. The use of student data will be discussed in more detail in this OEA Guidebook.

4. **Require teachers to be evaluated** on a regular basis
   a. Probationary teachers: annually
   b. Contract teachers: 2-year cycle

5. **Require clear, timely, and useful feedback** to guide professional development
   a. Establish a formative growth process for teachers that supports professional learning and collaboration with other teachers and administrators
   b. Use evaluation methods and professional development, support and other activities that are based on curricular standards and that are targeted to the needs of each teacher
   c. Align professional development opportunities with educator self-reflection, assessment, and goal-setting.
   d. The focus of local evaluation and support systems is to help educators improve their practice to improve student learning. Collaborative teams should determine what kind of support a teacher or administrator can expect if they are not proficient on all standards.
   e. Professional learning should be aligned to the TESS. The Oregon Framework suggests professional learning may be guided by the Learning Forward standards (see CD-ROM): job-embedded, collaborative, and customized to individual educator needs. The standards assert that professional learning that increases educator effectiveness and results for all students must include:
      1. **Learning Communities:** occurs within learning communities committed to continuous improvement, collective responsibility, and goal alignment.
      2. **Leadership:** requires skillful leaders who develop capacity, advocate, and create support systems for professional learning.
      3. **Resources:** requires prioritizing, monitoring, and coordinating resources for educator learning.
      4. **Data:** uses a variety of sources and types of student, educator, and system data to plan, assess, and evaluate professional learning.
      5. **Learning Designs:** integrates theories, research, and models of human learning to achieve its intended outcomes.
      6. **Implementation:** applies research on change and sustains support for implementation of professional learning for long term change.
      7. **Outcomes:** aligns its outcomes with educator performance and student curriculum standards.

6. **Will be used to inform personnel decisions.** School districts must describe in policy how their TESS is used to inform personnel decisions (e.g., contract status and renewal, plans of assistance, placement, assignment, career advancement, etc.).

---

### Performance Levels and Definitions of Performance as Applied to Standards of Professional Practice

<table>
<thead>
<tr>
<th>Performance Levels</th>
<th>Definitions of Performance as Applied to Standards of Professional Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Does not meet this standard</td>
</tr>
<tr>
<td>2</td>
<td>Making sufficient progress toward meeting this standard</td>
</tr>
<tr>
<td>3</td>
<td>Consistently meets expectations for good performance under this standard</td>
</tr>
<tr>
<td>4</td>
<td>Consistently exceeds expectations for good performance under this standard</td>
</tr>
</tbody>
</table>
New Oregon Evaluation & Support Requirements

The following chart can serve as a checklist for your Teacher Evaluation and Support System (TESS) design teams. The chart presents compliance pieces from new state-level policy (SB 290), the ESEA Waiver, and already established Oregon Statutes and Oregon Administrative Rules, which your team can use to ensure your TESS is a collaboratively designed system that will measure up to all the requirements.

<table>
<thead>
<tr>
<th>New State Law &amp; Policy</th>
<th>Collaboratively Designed System</th>
<th>Pre-Existing State Law Bargaining Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation systems used to determine effectiveness of educators &amp; in making human resources decisions</td>
<td>Include job descriptions and related performance standards</td>
<td>Based on written criteria (including performance goals) (ORS 342.850 (2))</td>
</tr>
<tr>
<td><em>School districts must develop policy to say how the evaluation &amp; support system is used to inform personnel decisions</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation systems used to improve professional development of educators</td>
<td>Include multiple measures of educator effectiveness</td>
<td>Based on at least two observations in combination with other information (ORS 342.850 (1))</td>
</tr>
<tr>
<td>Evaluation systems used for continual improvement of instruction</td>
<td>Include student progress based on multiple measures (including student, school &amp; performance measures)</td>
<td>Includes pre &amp; post evaluation interviews (ORS 342.850 (2))</td>
</tr>
<tr>
<td>Evaluation systems designed jointly by district, teachers and exclusive bargaining representative</td>
<td>Built on research-based practices</td>
<td>If needed: includes written plans of assistance (ORS 342.850 (2))</td>
</tr>
<tr>
<td>Evaluations systems aligned with InTASC Standards and customized application of standards throughout evaluation and support system.</td>
<td>Separately developed for teachers &amp; administrators</td>
<td>Evaluation reports maintained in district personnel files, but only after reasonable notice to teacher (ORS 342.850 (4), (5))</td>
</tr>
</tbody>
</table>
*Evaluation systems use four levels of performance* | Customized for each district | Educator may attach their own written statement to any evaluation document in file (ORS 342.850 (6)) |
*Evaluation systems use multiple & valid measures in determining performance levels of educators, including, as a significant factor, student performance (defined locally and through ODE pilot process)* | Allow for individual differences in assignment | |
*Student performance will be evaluated via multiple measures embedded within collaboratively-designed student learning goals.* | Used to refine support & professional growth system based on needs of individuals, schools & district | |
*Teacher evaluation & support systems must align to professional development* | Oregon Framework suggests Professional development must be guided by NSDC/ Learning Forward Standards. | |
| Establish formative growth process that supports professional growth & collaborative learning | Use methods based on curricular standards targeting the needs of individual educators | |
| School boards will adopt policy specifying which school officials have file access (ORS 342.850 (9)) | | |
| Check your Collective Bargaining Agreement for other requirements | | |

*Denotes compliance policies that are part of Oregon’s ESEA Waiver and are additional requirements above our newly revised Evaluation statutory requirements.*
Oregon Framework Timeline for Implementation

<table>
<thead>
<tr>
<th>School Year</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011-12</td>
<td>• State Board adopted state framework.</td>
</tr>
</tbody>
</table>
| 2012-13     | • ODE pilots framework at selected sites (SB 252, SIG, others) to determine what “significant” use of student learning data in teacher and administrator evaluation means.  
• All districts develop local evaluation and support systems consistent with state guidelines/framework. |
| By July 1, 2013 | • All districts submit revised evaluation and support systems, implementation plans, and training plans. |
| 2013-14     | • All districts pilot implementation of local evaluation and support systems. |
| 2014-15     | • All districts fully implement local evaluation and support systems. |
| By July 1, 2015 | • All districts present local evaluation and support systems to a Regional Peer Review Panel. |

Process Checklist & Key Questions for Collaboratively Developing a Teacher Evaluation and Support System (TESS)

SHORT OUTLINE (see next page for full process)

I. BEFORE DESIGN PROCESS
   a. Designate members of your bargaining unit to be on Teacher Evaluation Design team
      i. Have your local President and Bargaining Team meet with the Superintendent
      ii. Consult OEA’s Center for Great Public Schools for any supports/trainings you may need
   b. Design a collaborative process for the meetings
   c. Schedule meetings on a timeline to finish process by July 1, 2013
   d. Design a communications plan

II. DESIGN PROCESS
   1. The Teacher Evaluation and Support System (TESS) framework
   2. The rubric used for this framework
   3. Multiple valid measures of teaching effectiveness (which includes multiple valid measures of student learning, professional practice, and professional responsibilities)
   4. Professional Development and Growth System (PDGS) that is informed by teacher evaluation
   5. Reworking or revising of current job descriptions and performance standards as necessary
   6. Training of both administrators and teachers in the new system
   7. Pilot & Implementation or roll-out plan
   8. Evaluation and continuous improvement plan
Pre-Design Process Checklist & Key Questions

In working through each design section, the Teacher Evaluation Design (TED) Team can use the following list of process elements and key questions to guide and focus their discussions and work. The TED team may determine that more comprehensive external feedback from individual stakeholder groups is necessary to continue. Some of these questions might also serve as useful starters for a membership survey, so feel free to use them in any format you see fit (See also Laura Goe’s “Questions to Ask About Measures and Models” on the CD-ROM):

BEFORE DESIGN PROCESS

☐ Access OEA’s Teacher Evaluation Guidelines and supporting documents.

☐ Executive Committee designate one or more people to create a Teacher Evaluation Planning Committee (TEPC) to oversee this work and adds TEPC report to every Executive Committee agenda for the next two years at least (at least one member of the Executive Committee must be on TEPC)

☐ Add members to committee so that it is made up of representatives from:

1. Different grade level buildings
2. Different specialty/content areas
3. Different demographics (experience, race/ethnicity, age, etc.)
4. One person+ on TEPC must have bargaining experience or be a bargaining team member
5. Include Special Education and English Language Development personnel in designing, implementing, and monitoring evaluation models

☐ Teacher Evaluation Planning Committee assign at least three members in the committee to form a sub-committee in charge of internal communication (see OEA guidelines)

1. Clarify expectations in terms of stakeholder groups’ purposes and authority in decisions
2. Define stakeholders’ roles and responsibilities in a way that capitalizes on their expertise

3. Create a two-way collaborative communication plan for all external (outside of group) communication (to members, parents, students, community, etc.)
4. Create a two-way communication system to distribute and gather information from membership
5. Consider the content, the target audience, and the mode and timing of these communications, as well as what will be done with the responses.

☐ Teacher Evaluation Planning Committee looks over OEA materials, attends regional training or invites Center for Great Public Schools staff to give overview to committee

☐ Teacher Evaluation Planning Committee initiates contact with district to set up *collaborative design process for developing the new Teacher Evaluation and Support System (TESS)*

☐ Teacher Evaluation Planning Committee holds initial meeting with their Bargaining Team to give overview of collaborative design process and discuss placeholder language options and/or how this process will be represented (via an MOU, contract language, separate Teacher Evaluation and Support System handbook, etc.)

1. Decide how the bargaining team and TEPC will be communicating with one another throughout the process.

☐ Teacher Evaluation Planning Committee (or representatives of TEPC) and district team form the new Teacher Evaluation Design Team (TED Team) so that teachers are at least equally represented.

☐ Teacher Evaluation Design Team sets up regular meetings at least monthly that are contractually supported

☐ Teacher Evaluation Design Team shares resources and spends time clarifying the following:

1. Group norms
   ◦ How do you revisit/reinforce them
2. What the collaborative process will look like  
   - Decide how you reach agreements  
   - What to do when you agreements are not met  
   - Decide how you maintain balance between parties  
3. Sharing of responsibilities  
   - Determine what roles will need to be filled (facilitator, note-taker, public scribe, time-keeper, “temperature-reader,” group norms sentinel, etc.) and record each roles job description  
4. How minutes are taken and shared with entire district  
5. How entire district is allowed to contribute to the work of the team (increases buy-in)  
6. End game goals and outcomes  
   - Define the purpose of the Teacher Evaluation and Support System.  
   - Make sure goals are stated in measurable terms which are explicitly, well-defined, and easily understood by all stakeholder groups.  
   - Make sure the Teacher Evaluation and Support System goals are aligned to meet the requirements of state policy and research-based best practices.  
   - Determine resources necessary to complete Teacher Evaluation and Support System design process  
   - See “Evaluation and Continuous Improvement Plan” notes below  
7. *Timeline for getting system design completed by July 1, 2013*  
   - Build in communication/feedback loops with entire district into timeline  
8. Define stakeholder groups  
9. Beginning Collaborative Decisions  
   - Collaborate on a group definition of effective teaching  
   - Come to consensus on a common vocabulary  
   - Come to consensus on a common understanding of the InTASC Standards  

* Asterisks represent legal requirements
Teacher Evaluation Design Team divides the work of the design into the following eight sections and decides how to move through each section and keep the comprehensive design aligned and cohesive:

1. The Teacher Evaluation and Support System (TESS) framework

- *How it will be used (formative and summative)*
- *Determine the frequency of evaluation (align to state policy) and the frequency of each measure used*
  - How many observations using the framework per year (different for various career levels?)
  - How it is informed by written performance goals
  - How it is informed by pre- and post-observation interviews
- *How it will be aligned to and improve Professional Development*
- *How it is included within a Professional Growth System*
- *How it is aligned to InTASC Standards and the application of the Standards across the entire evaluation and support system*
- *How it will have multiple measures*
- How it will be revised to apply to different specialties/assignments
- Consider determining a base-line level of performance prior to making decisions regarding teacher proficiency levels
- Consider the technical defensibility of each measure and the Teacher Evaluation and Support System as a whole to make personnel decisions (or compensation decisions if you are in a school district with TIF/SIG funds)
- Compare and contrast your new Teacher Evaluation and Support System framework with your current TESS to see what (if anything) is worth keeping and what needs to be tweaked, modified, or thrown away altogether
- Identify a common framework for effectiveness and include differentiated criteria where applicable, especially when considering teachers’ various assignments and specialties (such as Special Education and English Language Development personnel)
- Integrate evidence-based practices for teachers working with students who are English language learners and students with disabilities
- Ensure that evaluation framework can identify and provide targeted Professional Development

KEY QUESTIONS TO ASK:

- Where is the money going to come to support:
  - These collaborative conversations?
  - Release time
  - The training of all teachers/certified staff and administrators/evaluators?
  - Time? Pay?
  - The sustainability of an enhanced system?
- If evaluation of teachers is to measure the ability of the teacher to teach, what is our shared definition of effective teaching?
  - This definition is important to use as a reflective tool to make sure your Teacher Evaluation and Support System aligns to this definition
- What are our shared agreements about how a Teacher Evaluation and Support System should work?
- Does the process allow for the cooperative development and assessment of an evaluation system?
- Is there a system of checks and balances for the process?
- What does our current Teacher Evaluation and Support System look like?
  - What are its strengths? What works?
  - What are its weaknesses? What doesn’t work?
  - Is it aligned to any professional development system?
  - How often is it assessed by an inclusive group?
  - Is it manageable? Clear to everyone? Effective in growth and improvement of practice?
  - How does it apply across levels of experience/content area/specialists?
  - Is it differentiated across levels of experience/content area/specialists?
  - *Are there multiple valid measures of teaching effectiveness?*

* Asterisks represent legal requirements

DESIGN PROCESS CHECKLIST

Tools II.
**Does that include multiple valid measures of student performance?** *Must have at least one measure of student performance*

Are there established teacher evaluation tools you can modify and adapt to these purposes?

Does each tool align to InTASC Standards, and does each customized application of the Standards across the entire evaluation system reflect the integrity of InTASC, or are there crosswalks available?

How will the district and/or Association evaluate the validity, reliability, comparability, relevancy, meaningfulness of the measures being used?

Some questions you might ask yourselves are:

Does each tool align to InTASC Standards, and does each customized application of the Standards across the entire evaluation system reflect the integrity of InTASC, or are there crosswalks available?

Where does each tool align to InTASC Standards, and does each customized application of the Standards across the entire evaluation system reflect the integrity of InTASC, or are there crosswalks available?

**How will the district and/or Association evaluate the validity, reliability, comparability, relevancy, meaningfulness of the measures being used?**

Some questions you might ask yourselves are:

- Does each tool align to standards? And does each customized application of the Standards across the entire evaluation system reflect the integrity of InTASC, or are there crosswalks available?

- Where does each tool align to InTASC Standards, and does each customized application of the Standards across the entire evaluation system reflect the integrity of InTASC, or are there crosswalks available?

- Does the evaluation system achieve the purposes for which it was designed?

- How well does the evaluation system support effective teaching for all students?

- Are the measures valid in terms of measuring the agreed-upon definition of effective teaching?

- Do the measures meet high standards of reliability in every school and for every teacher?

- What current bargaining language do you have regarding your evaluation system?

- Do the measures meet high standards of reliability in every school and for every teacher?

- What current bargaining language do you have regarding your evaluation system?

- What are its strengths?

- What are its weaknesses?

- How will disputes of findings be processed?

- See bargaining section of OEA’s Teacher Evaluation and Support Systems Guidebook

---

### 2. The rubric used for this framework

- “Required four levels of performance”

- Define what the triggers are for moving into and out of plans of assistance (POA)

- Define trigger points for action (professional growth, professional development, etc.) - when action will be triggered, what kind of action, for how long, etc.

- Determine how evaluation results will be shared with teachers and when they will be given notice of the next steps toward professional growth or termination

- Determine how much time and assistance will be provided for a teacher to demonstrate improvement before termination is considered (assure that POA language and procedures align to this)

- “Rubric must be clearly delineated for all ten InTASC Standards at all four performance levels, or a crosswalk must be made to outline where each of the InTASC Standards is addressed.”

### KEY QUESTIONS TO ASK:

- Does the process provide a system of support to assist members to move from one performance level to another? Have the expectations been clearly articulated for each performance level and the triggers for progressing along the continuum?

- Does the process allow for reasonable, attainable identification of ways to correct significant discrepancies/move out of plans of assistance that is also aligned with supports?

- What would it look like for your teacher evaluation system to differentiate across specialty areas and assignments?

  - What types of specialists are a part of your bargaining unit?

  - How do their day-to-day jobs differ from that of a classroom teacher?

  - How do they differ in terms of professional responsibilities?

  - What assurances need to be considered or in place to assure equity of the evaluation system across specialty areas and assignments?

  - What would it look like for your teacher evaluation system to differentiate across years of experience?

- What supports do newer teachers need?

- How could you graduate your system to meet the needs of all levels of experience?

---

### 3. Multiple valid measures of teaching effectiveness

(Which includes multiple valid measures of student learning, professional practice, and professional responsibilities)

- “At least one measure identified for each category of teaching effectiveness: professional practice, professional responsibilities, and student learning (which will be the Student Learning Goal as outlined on page 33).”

- Measures are research-based for the purpose of individual teacher evaluation

- Measures are relevant; they measure what we value and how we define teaching effectiveness and will give meaningful feedback for improvement of instruction
Measures are valid (they measure what they intend to measure over a period of time and under similar conditions)

Measures are reliable (they measure what they intend to measure over a period of time and under similar conditions)

Measures are comparable (they measure the same definition of teaching effectiveness across different grade levels, schools, content areas, etc.)

Weigh measures fairly:
- Consider the validity/reliability/comparability of a measure and weigh those more heavily that have the most relevancy and research to back them
- Consider level of experience and job description
- Consider evaluator capacity (human and resource capacity strengths and limitations)
- Consider that after a pilot, or as a Teacher Evaluation and Support System progresses, weights of measures can be changed as some are found to be more reliable, valid, comparable, and relevant than others
- Consider aligning weights to district priorities, like collaboration between teachers
- Consider that if the ultimate goal is to increase teacher capacity to implement evidence-based practices which will improve student learning, growth and achievement, the observation instrument may need to carry more weight
- Consider phasing in reliance on new observation instruments

Measures support your Teacher Evaluation and Support System goals and purposes

Measures have a demonstrated impact on teacher practice

* For ODE Pilots, clearly describe how student growth will be used as a “significant factor.” Include, as a significant factor, at least two student learning goals* (see Student Learning Goals section)
- Best practice — student performance is based on multiple valid measures
- Determine what are meaningful, valid, reliable and comparable measures
- Determine how different student measures apply across assignments
- Student data measures should:
  - ★ Have the ability to accurately measure student progress between two points in time (growth)
  - ★ Be rigorous
  - ★ Be comparable across classrooms
- Measures have a demonstrated impact on student achievement

Determine what is required to sustain and support the collection, interpretation and application of these multiple valid measures of both teacher and student performance

Initiate a conversation with your bargaining team (consult UniServ Consultant and Center for Great Public Schools) if the district intends to link individual teacher data with their students’ data including unique identifiers for both students and teachers

**KEY QUESTIONS TO ASK:**

- How do you envision multiple valid measures of teaching effectiveness?
- How do you ensure that the measures meet the following criteria:
  - Validity
  - Reliability
  - Comparability
  - Relevancy/meaningfulness
- What tradeoffs will the Teacher Evaluation Design Team consider between the comprehensiveness/fidelity of the measures and the practicality (doability/tenability) of the system?
  - What will be the consequences of these tradeoffs?
- What does including multiple valid measures of student growth/learning look like in your school district?
- When considering tools to assess teachers’ attainment of standards of effective teaching (measures of practice) and tools to assess teachers’ impact on student outcomes (measures of effects), ask yourself these questions:
  - What measures are in place already and how challenging would a change in evaluation measures be for both evaluators and teachers?
  - What resources are needed to monitor and sustain the effectiveness of the measure?
- Additional questions to consider when developing or selecting measures of student learning:
  - Does the approach allow for the assessment of student learning over time?
  - How can we establish the following:
    - ★ The proposed measure assesses the expected knowledge and skill appropriately, in terms of the content of questions or tasks included and the coverage of the subject area.
    - ★ Some students are not disadvantaged by the specific questions or tasks included.
<table>
<thead>
<tr>
<th>DESIGN PROCESS CHECKLIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>The measure appropriately distinguishes among students.</td>
</tr>
<tr>
<td>Scores based on the measure accurately reflect meaningful changes in student learning in the subject area, either in strictly comparative terms (e.g., some students learned more than other students) or growth toward a standard (e.g., some student made more progress than others toward a goal that will help them be successful).</td>
</tr>
<tr>
<td>How can we ensure that student learning is being measured consistently across classrooms/building/district?</td>
</tr>
<tr>
<td>What steps in development and administration are needed to ensure that scores will have the same meaning within the subject area and that student growth will have a similar interpretation across subjects? (Example, does growth of 15% in mathematics mean the same thing as in English Language arts, and is it achievable across subjects?)</td>
</tr>
<tr>
<td>Is the approach and measure transparent and understandable to stakeholders?</td>
</tr>
<tr>
<td>What capacity/resources are needed to develop and implement the measures now and over time?</td>
</tr>
<tr>
<td>How easily can data from the model be used along with other data to assess teaching effectiveness?</td>
</tr>
<tr>
<td>What provisions are in place to ensure ongoing review, calibration, and adjustments, when necessary?</td>
</tr>
<tr>
<td>Does the district have a plan to ensure data accuracy?</td>
</tr>
<tr>
<td>Does the district intend to link individual teacher data with their students’ data, including unique identifiers for both students and teachers?</td>
</tr>
<tr>
<td>What teacher evaluation tools/rubrics will meet both the adopted state standards and the evaluation system you are developing for your school district?</td>
</tr>
<tr>
<td>☐ Who will be evaluated using the proposed system and will different measures be used for different staff?</td>
</tr>
<tr>
<td>☐ Given time and resource constraints, are the tools practical? What tradeoffs between practicality and comprehensiveness must be made?</td>
</tr>
<tr>
<td>☐ Will the ratings from each of these measures be weighted to tally a final evaluation rating?</td>
</tr>
<tr>
<td>☐ Will classroom observation tools employ checklists, rubrics, or narratives?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPECIFIC TO OBSERVATIONS:</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Selecting a Rubric:</td>
</tr>
<tr>
<td>Will the same rubric be used for all teachers?</td>
</tr>
<tr>
<td>☐ Selecting and Training Evaluators:</td>
</tr>
<tr>
<td>How will evaluators be selected?</td>
</tr>
<tr>
<td>What training will evaluators receive?</td>
</tr>
<tr>
<td>Will evaluators be required to demonstrate competency and inter-rater reliability before administering evaluations?</td>
</tr>
<tr>
<td>☐ Conducting Observations:</td>
</tr>
<tr>
<td>How many observations will be required?</td>
</tr>
<tr>
<td>Will this number differentiate between levels of experience/assignment?</td>
</tr>
<tr>
<td>When will the observations take place so that they are effective (e.g., not the last two weeks of school)?</td>
</tr>
<tr>
<td>What types of observations (formal and informal) and how many of each will take place and how long will each observation be?</td>
</tr>
<tr>
<td>☐ What will be the pre- and post-observation meeting protocols for formal observations?</td>
</tr>
<tr>
<td>☐ Collecting Information:</td>
</tr>
<tr>
<td>What information will be collected to support observation findings?</td>
</tr>
<tr>
<td>How will the results be shared with the teachers?</td>
</tr>
<tr>
<td>How will the district ensure that the results are valid and reliable?</td>
</tr>
<tr>
<td>How will disputes of findings be processed?</td>
</tr>
<tr>
<td>☐ Refining the Process:</td>
</tr>
<tr>
<td>How will the evaluator training be monitored?</td>
</tr>
<tr>
<td>How will the teachers be trained on the system and how will this training be monitored?</td>
</tr>
<tr>
<td>How will inter-rater reliability be monitored and by whom?</td>
</tr>
<tr>
<td>☐ Does the tool provide multiple valid measures or sources of evidence of teaching practice and effectiveness?</td>
</tr>
<tr>
<td>☐ Has the tool been pilot-tested in the field and been shown to have potential to be valid and reliable (aka, it measures what it’s supposed to measure time and time again)?</td>
</tr>
<tr>
<td>☐ Does the tool have professional credibility (aka, developed with teacher input, measures important aspects of teaching practice and effectiveness)?</td>
</tr>
<tr>
<td>☐ Does the tool provide feedback that teachers can use to improve their practice?</td>
</tr>
<tr>
<td>☐ Does the tool effectively differentiate among four levels of teaching practice and effectiveness that meet the state criteria?</td>
</tr>
</tbody>
</table>
4. "Professional Development and Growth System (PDGS) that is informed by teacher evaluation" (see page 49 for more)

- Define how mentoring or peer assistance is used (if at all)
- Determine how our Professional Development and Growth System is elevating teaching across the district
- Define how the Professional Development and Growth System will be aligned to meet the needs of all teachers at all performance levels and provided in a manner that is supported in research
- Develop opportunities for teachers to improve that are embedded in the evaluation cycle (see CD-ROM for Learning Forward’s Professional Development Standards)
- Develop supports to be provided to assist teachers with lower-levels/unacceptable performance
- Evaluate outcomes from Teacher Evaluation and Support System to determine whether the Professional Development and Growth System is elevating practice:
  - Each PD activity
  - Longitudinal analysis of teacher participation, support, and outcomes related to student learning/growth
  - Ensure that evaluation framework can identify and provide targeted PD for Special Education and English Language Development personnel
  - Develop policy on how Teacher Evaluation and Support System will inform personnel decisions consistent with the collective bargaining agreement and bargaining requirements

KEY QUESTIONS TO ASK:

- Does the process provide opportunities for teachers to seek and receive professional development that aligns with their evaluation?
- Does the process provide an inspiring roadmap for teacher growth and development?
- Does the process support relevant, robust, and timely professional assistance when needed?
- Envision an evaluation system in your district that includes a formative growth process:
  - What would these formative assessments look like?
  - How would these assessments inform teacher growth and development?
  - Who would conduct these assessments?
  - How would the “assessors” and “assessees” be trained on the formative growth process?
  - What criteria should be included?
  - How would these assessments align to/inform the PDGS?
  - What would this kind of system cost in time and money?
  - How would it be applied at different venues in the district?
  - If you could pilot it where should you begin?
- Envision how your school district could use the evaluation system to align to/inform a professional growth system in the district:
  - What would that look like?
  - How could this be aligned with curricular standards, district goals, school goals, and identified needs of students and teachers?
  - How could this be an application of the InTASC Standards?
  - What tools/systems would need to be in place in order to do this?
  - What resources would need to be in place?
  - What targeted support would need to be in place?
  - How could you assess the outcomes of the professional development (PD)?
  - How could the evaluation system and PD systems support teachers’ new knowledge and skills?
  - If the results of formative assessments are positive, how would that impact the PD opportunities?
  - If the results of formative assessments identify significant shortcomings, how would that impact the PD opportunities?
  - How can these PD opportunities tie into licensure requirements of the state?
  - How can we individualize PD opportunities and still have a fluency in our overall PGDS?

- How will the measures assist in the development of specific performance goals and targeted PD?
- Can teacher application and reflection be built into the professional learning activity?
- Are professional learning activities “job-embedded” or a one-time-only session? What are the human and fiscal resources that can be used to provide both types of PD?
Do teachers have common planning times to reflect upon new practices?

Can opportunities for teachers to observe effective teachers be provided?

Will professional learning communities be established?

5. Reworking or revising of current job descriptions and performance standards as necessary

**KEY QUESTIONS TO ASK:**

- Does your new Teacher Evaluation and Support System require reworking or revising current job descriptions and performance standards?

6. Training of both administrators and teachers in the new system

- Consider the training needs for each of the multiple valid measures chosen
- Establish evaluator training that includes explicit training on measures designed to assess Special Education and English Language Development personnel
- Consider that implementation fidelity is most important when the selected measures are dependent on human scoring with observation instruments or rubrics.
- Create explicit decision rules and requirements for quantity and quality of examples of evidence for making a justification of one performance rating over another which may differentiate for different assignments, experience levels, etc.
- *Design a plan for training, due by July 1, 2013 to ODE*

**Training Evaluators**

- Create a training system that is co-led by association leaders and members/teachers designated by the association, along with administrators
- Decide how fidelity will be ensured prior to the pilot, during the pilot, and throughout the life of implementation
- Decide how inter-rater reliability and calibration will be ensured prior to the pilot, during the pilot, and throughout the life of implementation
- Determine what specialized training for the evaluation of or review of specific content or specialty area teachers is needed
- Decide how and to what extent the training provided will create opportunities for guided practice
- Decide how and to what extent the training will provide specific feedback to improve reliability
- Determine what mechanisms will be in place to retrain evaluators/reviewers who are not implementing the system with fidelity
- Determine a schedule of regular monitoring of evaluators for inter-rater reliability and overall system fidelity
- Determine criteria by which an evaluator will be terminated from their evaluation duties
- Decide how evaluation responsibilities are a component of administrative evaluations and that administrators are evaluated on their ability to evaluate teachers with fidelity and reliability

**Additional Considerations for Training:**

- Teacher preparation for measuring student learning growth is limited or non-existent
- Most principals, support providers, instructional managers, and coaches are poorly prepared to make judgments about teachers’ contribution to student learning growth

**They need to know how to:**

- Evaluate the appropriateness of various measures of student learning for use in teacher evaluation
- Work closely with teachers to select appropriate student growth measures and ensure that they are using them correctly and consistently.
KEY QUESTIONS TO ASK:

- How can administrators be trained, supported and monitored to carry out an enhanced system?
- How is the association involved in the trainings? Co-trainers?
- How are all stakeholders trained in the system?

KEY QUESTIONS TO ASK / QUESTIONS THE PILOT SHOULD ADDRESS:

- What criteria will the Teacher Evaluation Design Team create for schools to qualify as a pilot site for the evaluation model?
- How long will the pilot last?
- Are there any conflicts between the proposed evaluation tools and the collective bargaining agreement? If so, what challenges are likely to arise in the negotiation process when labor and management attempt to resolve them?
- Who will train the teachers, administrators, and evaluators on the new evaluation tools? How will the union make sure that the training addresses reliability issues and equips evaluators to evaluate specific content areas and specialist effectively? Will evaluators be subject to frequent review or to ensure reliability?
- Who will serve as evaluators? What will the criteria be? How will their work be supported and funded? How is their role as an evaluator part of their evaluation system?
- When will the training occur? Will compensation be provided to teachers to complete training after the school date or will substitute teacher coverage be provided to complete it during the school day?
- How will the evaluation data collected during the pilot be used? Will the stakeholders be “held harmless” during the pilot?
- Which measures will be used for formative purposes and which will be used for summative purposes? How will the determination be made?
- Will feedback be collected systematically at each stage of the pilot so that changes to the Teacher Evaluation and Support System can be made accordingly? How will that happen? Will changes be made to the Teacher Evaluation and Support System part way through the pilot stage, or will all changes be made after the pilot but before the reforms are launched district-wide?

QUESTIONS TO ASK STAFF DURING THE PILOT:

Feedback from teachers, principal, and other staff members should be collected throughout the pilot to ensure that professional growth is targeted by and improved with the new Teacher Evaluation and Support System. Some questions to ask of staff during the pilot include:

7. Pilot & Implementation or roll-out plan

- Explore any potential conflicts between the proposed pilot and the collective bargaining agreement. If necessary, develop MOUs that will allow the pilot to move forward with a shared agreement.
- Consider ways to either pilot different building levels of the plan or scaffold the district phase-in from the less to more complex component parts.
- Hold teachers harmless during the pilot years of the Teacher Evaluation and Support System until a measure of validity and reliability can be guaranteed, as well as the degree of relevancy/meaningfulness established.
- Decide how long the Teacher Evaluation and Support System pilot will be and the resources necessary to progress.
- Decide how data will be taken during the Teacher Evaluation and Support System pilot, the frequency and who/how it will be analyzed.
- Decide the frequency with which modifications will be made and how they will be trained and communicated.
- Determine whether research will be conducted in conjunction with implementation to provide validation.
- Establish outcomes/goals to determine the overall effectiveness of the Teacher Evaluation and Support System.
- Decide if there are resources available to conduct an external or internal assessment of the evaluation.
- Rigorously define feedback loop during implementation so details can be assessed and reworked as it rolls out.
- *Design a plan/timeline for pilot implementation, due to ODE by July 1, 2013*
Do teachers believe that the new system leads to more targeted PD that can improve their performance?
Do individual teachers get to set goals in the new system?
Do teachers feel they have adequate information about the new Teacher Evaluation and Support System? Do they know where to go with questions?
Do teachers believe the new system is fair?
Do teachers understand what the expectations and criteria are for performance at each level, and what it is they need to do in order to progress along the continuum?
Is there anything about the new tools/ measures that is confusing to evaluators? To teachers? Do evaluators feel adequately trained to use these new tools? Do teachers?
How will communication be handled between the Association and members during the development and piloting of the system?

**QUESTIONS TO ASK AFTER THE PILOT:**
Review the feedback from the pilot and ask yourself (and those with whom you are bargaining or jointly developing the system) the following questions:

Which aspects of the Teacher Evaluation and Support System are not working as intended? Can they be modified or do they need to be replaced?
Are there aspects of our Teacher Evaluation and Support System that were not foreseen or planned for, which need to be discussed by the Teacher Evaluation Design Team and have a plan developed?
Is there any indication from the pilot that the evaluation results are influencing the district-wide Professional Development and Growth System planning?
In practice, does the Teacher Evaluation and Support System align with InTASC Teaching Standards? Are all tools/ measures an accurate application of the Standards?
Are instructional coaches or other resources available to support professional development needs identified in the evaluation results?
Will there be adequate resources/funding to support widespread improvements to professional development based on evaluation?
What are the transitions needed to progress to the next stage (additional piloting, initial implementation, etc.)?

**8. Evaluation and continuous improvement plan**

Make sure that a balanced team of teacher and district representatives are always a part of evaluating this system throughout its lifetime (contractually protect)
Design an evaluation of your Teacher Evaluation and Support System that can determine whether or not:
- Stakeholders value and understand the system
- Student learning/growth is improved
- Teacher practice is positively impacted
- Teacher retention is improved
- The system is implemented with fidelity
Establish a plan to evaluate measures to determine if they can effectively differentiate among teacher performance
Need to identify potential “widget effects” in measures
If measure is not differentiating among teacher practice, may be faulty training or poor implementation, not necessarily the measure itself or the teacher/practitioner
Examine correlations among results from measures
Establish a plan/schedule to evaluate processes and data each year and make needed adjustments

**KEY QUESTIONS TO ASK:**

To what extent does the Teacher Evaluation and Support System (TESS) assess what is under the direct control of teachers?
Does your TESS determine whether or not:
- Stakeholders value and understand the system?
- Student learning/growth is improved?
- Teacher practice is positively impacted?
- Teacher retention is improved?
- The system is implemented with fidelity?
What implications does the TESS have for induction?
What implications does the TESS have for recruitment?
What resources are needed to monitor and sustain the effectiveness of the Teacher Evaluation and Support System?
- How do we avoid having teacher evaluation being just another education “fad?”
- How do we continue to prioritize this work?
In December 2011, the Oregon State Board of Education adopted national Model Core Teaching Standards, known as the InTASC Standards. Created by the Council of Chief State School Officer’s (CCSSO) Interstate Teacher Assessment and Support Consortium (InTASC), these core teaching standards outline what teachers should know and be able to do. Teacher Evaluation and Support Systems (TESS) in Oregon should be collaboratively customized as an application of the standards by July 1, 2013. These standards serve as Oregon’s definition of effective teaching and also outline the common principles and foundations of teaching practice that cut across all subject areas and grade levels.

Additionally, these standards are meant to complement and elevate the tremendous work occurring in Oregon’s classrooms and schools on a daily basis, and will continue to foster the growth and professional development of educators. These standards also assume a newer role for teachers as the profession evolves, including responsibilities for facilitating a professional collaborative culture that has both implicit and explicit implications for leadership roles and responsibilities.

InTASC’s TEN STANDARDS

InTASC’s ten Standards are divided into four categories: The Learner and Learning, Content Knowledge, Instructional Practice, and Professional Responsibility. Each of the ten standards is further subdivided into three categories of indicators: performances, essential knowledge, and critical dispositions. “Performances” are the actions and teaching behaviors that can be observed and assessed, “essential knowledge” signals the role of declarative and procedural knowledge as necessary for effective practice, and “critical dispositions” indicates that habits of professional action and moral commitments that underlie the performances play a key role in how teachers do, in fact, act in practice.

Getting to Know the InTASC Standards

The three categories of indicators within the InTASC Standards (performances, essential knowledge, and critical dispositions) as a whole are not meant to be copied and pasted into a checklist or rubric. Essential knowledge and critical dispositions are crucial aspects of teaching effectiveness. They are not aspects that can be readily measured via observations, rubrics, or other measures of teaching. These standards should not be repurposed as a checklist or rubric. Rather, they are the starting point for the local customization in which, through the collaborative process discussed in this guidebook, a comprehensive evaluation system is developed that supports improved teaching effectiveness and student learning.

As you familiarize yourself with these standards in the creation of your TESS, keep in mind that while each standard emphasizes a discrete aspect of teaching, teaching and learning are infinitely dynamic, integrated and reciprocal processes. In order to fully capture this vibrant relationship, the standards overlap and must be taken as a whole in order to convey a complete picture of the acts of teaching and learning. Also, it is important to keep in mind that indicators are examples of how a teacher might demonstrate each standard.

The indicators used in the standard are simply examples of how a teacher might demonstrate proficiency for each standard; there may be other better indicators in a particular classroom. In other words, an evaluator should not expect a teacher to demonstrate each indicator during a performance assessment.

Check out the 10 standards on the next page!
THE LEARNER AND LEARNING

Standard #1: Learner Development
The teacher understands how students grow and develop, recognizing that patterns of learning and development vary individually within and across the cognitive, linguistic, social, emotional, and physical areas, and designs and implements developmentally appropriate and challenging learning experiences.

Standard #2: Learning Differences
The teacher uses understanding of individual differences and diverse cultures and communities to ensure inclusive learning environments that enable each student to meet high standards.

Standard #3: Learning Environments
The teacher works with other to create environments that support individual and collaborative learning, and that encourage positive social interaction, active engagement in learning, and self-motivation.

CONTENT KNOWLEDGE

Standard #4: Content Knowledge
The teacher understands the central concepts, tools of inquiry, and structures of the discipline(s) he or she teaches and creates learning experiences that make these aspects of the discipline accessible and meaningful for students to assure mastery of the content.

Standard #5: Application of Content
The teacher understands how to connect concepts and use differing perspectives to engage students in critical thinking, creativity, and collaborative problem solving related to authentic local and global issues.

INSTRUCTIONAL PRACTICE

Standard #6: Assessment
The teacher understands and uses multiple methods of assessment to engage students in their own growth, to monitor learner progress, and to guide the teacher’s and student’s decision making.

Standard #7: Planning for Instruction
The teacher plans instruction that supports every student in meeting rigorous learning goals by drawing upon knowledge of content areas, curriculum, cross disciplinary skills, and pedagogy, as well as knowledge of students and the community context.

Standard #8: Instructional Strategies
The teacher understands and uses a variety of instructional strategies to encourage students to develop deep understanding of content areas and their connections, and to build skills to apply knowledge in meaningful ways.

PROFESSIONAL RESPONSIBILITY

Standard #9: Professional Learning and Ethical Practice
The teacher engages in ongoing professional learning and uses evidence to continually evaluate his/her practice, particularly the effects of his/her choices and actions on others (students, families, other professionals, and the community), and adapts practice to meet the needs of each student.

Standard #10: Leadership and Collaboration
The teacher seeks appropriate leadership roles and opportunities to take responsibility for student learning, to collaborate with students, families, colleagues, other school professionals, and community members to ensure student growth, and to advance the profession.
## MEASURES OF PROFESSIONAL PRACTICE

<table>
<thead>
<tr>
<th>Measure / Description / Examples</th>
<th>Research</th>
<th>Strengths</th>
<th>Cautions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measure:</strong> Classroom Observation Instruments</td>
<td>Some highly researched protocols have been found to link to student achievement, though associations are sometimes modest. Research and validity findings are highly dependent on the instrument used, sampling procedures, and training of raters. There is a lack of research on observation protocols as used in context for teacher evaluation.</td>
<td>• Professionally credible as they often are developed by teachers and experts in teacher education&lt;br&gt;• Can be adapted for various subjects, grades, and contexts&lt;br&gt;• Can measure many aspects of practice that are associated with effective teaching&lt;br&gt;• Usually based on vetted standards of professional practice&lt;br&gt;• Some observation instruments have been demonstrated to be related to student achievement&lt;br&gt;• Provides rich information about classroom behaviors and activities&lt;br&gt;• Is generally considered a fair and direct measure by stakeholders</td>
<td>• Careful attention must be paid to choosing or creating a valid and reliable protocol and training and calibrating raters&lt;br&gt;• High quality, effective classroom observation that is valid and reliable requires extensive training of the evaluators and can be expensive and time-consuming; intensive training and calibrating of observers adds to expense but is necessary for validity&lt;br&gt;• To be successful: &lt;br&gt;• Needs to be frequent (every 2-3 weeks)&lt;br&gt;• Combination of formal and informal&lt;br&gt;• Candid, evidence based feedback is needed&lt;br&gt;• Professionally credible as they often are developed by teachers and experts in teacher education&lt;br&gt;• Can be adapted for various subjects, grades, and contexts&lt;br&gt;• Can measure many aspects of practice that are associated with effective teaching&lt;br&gt;• Usually based on vetted standards of professional practice&lt;br&gt;• Some observation instruments have been demonstrated to be related to student achievement&lt;br&gt;• Provides rich information about classroom behaviors and activities&lt;br&gt;• Is generally considered a fair and direct measure by stakeholders</td>
</tr>
</tbody>
</table>
| Measure: Non-classroom observations | Description: Observations of non-classroom settings, without students, but can include: staff meetings, parent meetings, PLCs/collaborative meetings, curriculum design meetings, preparation periods, etc | Not many explicit research studies on non-classroom observation only. | • Professionally credible as they often are developed by teachers and experts in teacher education
• Can be adapted for various contexts
• Can measure many aspects of practice that are associated with effective teaching
• Can be based on vetted standards of professional practice (InTASC)
• Can provide useful formative and summative information
• Is generally considered a fair and direct measure by stakeholders |
| Measure: Principal Evaluation | Description: Is generally based on classroom observation, may be structured or unstructured; uses and procedures vary widely by district. Is generally used for summative purposes, most commonly for tenure or dismissal decisions for beginning teachers. | Studies comparing subjective principal ratings to student achievement find mixed results. Little evidence exists on validity of evaluations as they occur in schools, but evidence exists that training for principals is limited and rare, which would impair validity of their evaluations. | • Can represent a useful perspective based on principals’ knowledge of school and context
• Is generally feasible and can be one useful component in a system used to make summative judgments and provide formative feedback. |
| Measure: Artifact Analysis | Description: These instruments and structured protocols used to analyze classroom artifacts in order to determine the quality of instruction in a classroom; may rate lesson plans, teacher assignments, teacher-created assessments, and scoring rubrics on particular criteria, such as rigor, authenticity, intellectual demand, and alignment to standards, clarity, and comprehensiveness. | Pilot research has linked artifact ratings to observed measures of practice, quality of student work, and student achievement gains. More work is needed to establish scoring reliability and determine the ideal amount of work to sample. Lack of research exists on use of structured artifact analysis in practice. | • Professionally credible
• Adaptable for different types of teachers
• Noninvasive, does not need to be done in real time
• Captures many aspects of teacher practice
• Can provide important formative and summative information on teacher practice
• Can be a useful measure of instructional quality if a validated protocol is used, if raters are well-trained for reliability, and if assignments show sufficient variation in quality
• Is practical and feasible because artifacts have already been created for the classroom |
| | Examples: • Instructional Quality Assessments (IQA) • Teacher Work Sample Methodology | • Few validated systems exist
• Comparability across different types of teachers has not been established
• Reliability across observers requires extensive training
• Does not include measures of student learning |
| | | • More validity and reliability research is needed |
| | | • Training knowledgeable scorers can be costly but is necessary to ensure validity |
| | | • This method may be a promising middle ground in terms of feasibility and validity between full observation and less direct measures such as self-report |
# Measures of Professional Responsibilities

<table>
<thead>
<tr>
<th>Measure / Description / Examples</th>
<th>Research</th>
<th>Strengths</th>
<th>Cautions</th>
</tr>
</thead>
</table>
| **Measure:** Teacher Self-Report/ Reflections | Studies on the validity of teacher self-report measures present mixed results. Highly detailed measures of practice may be better able to capture actual teaching practices but may be harder to establish reliability or may result in very narrowly focused measures. | • Can measure unobservable factors that may affect teaching, such as knowledge, intentions, expectations, and beliefs  
• Provides the unique perspective of the teacher  
• Is very feasible and cost-efficient; can collect large amounts of information at once  
• Can measure unobservable aspects of teacher quality  
• Can be easily administered  
• Can promote teacher self-reflection and analysis  
• Can promote a sense of self-efficacy  
• Do not include independent measures of impact on student learning  
• Can provide useful formative information | • Reliability and validity of self-report for summative purposes is not fully established and depends on instrument used  
• Using or creating a well-developed and validated instrument will decrease cost-efficiency but will increase accuracy of findings  
• This method should not be used as a sole or primary measure in teacher evaluation  
• Instructional logs have been used only for research purposes, so their validity for evaluation is questionable  
• Do not include independent measures of impact on student learning |
| **Description:** Teacher self-report measures may take the form of surveys, instructional logs, or interviews. Can vary widely in focus and level of detail. They ask teachers to report on what they are doing in the classroom, the extent to which they are meeting standards, and in some cases analyze the impact of their practice. They may consist of checklists, rating scales, rubrics, and may require teachers to indicate the frequency of particular practices. |  |  |  |
| **Examples:**  
• Study of Instructional Improvement instructional logs  
• Self-assessments  
• National Board for Professional Teaching Standards |  |  |  |
| **Measure:** Goal Setting | Reliability and validity of goal-setting for summative purposes is not fully established | • Is comprehensive and can measure aspects of teaching that are not readily observable in the classroom  
• Can be used with teachers of all fields  
• Provides a high level of credibility among stakeholders  
• Is a good tool for teacher reflection and improvement  
• Professionally credible  
• Can be done electronically  
• May include some assessment of student learning  
• Can provide important formative and summative information on teacher practice  
• Is very feasible and cost-efficient; can collect large amounts of information at once  
• Can be easily administered  
• Can promote a sense of self-efficacy | • Difficult to standardize (compare across teachers or schools)  
• May not include robust measures of student learning  
• Reliability and validity of goal-setting for summative purposes is not fully established  
• This method should not be used as a sole or primary measure in teacher evaluation |  |
| **Description:** These evaluation tools offer teachers the opportunity to set their own high but feasible objectives for their students’ growth in collaboration with their principal and/or other colleagues. The assessments teachers use may be common or standardized exams or teacher-developed assessments. Some tools require teachers to specify the professional development they will participate in to ensure their students achieve their growth objectives. |  |  |  |
| **Examples:**  
• Springfield School District (see CD-ROM)  
• Oregon Framework (See CD-ROM) |  |  |  |
Measure: Portfolios

Description:
Teacher portfolios are exhibits of evidence of teaching practice, school activities, and student progress. They usually are compiled by the teacher him- or herself. Portfolios may include teacher-created lesson or unit plans, descriptions of the classroom context, assignments, student work samples, videos of classroom instruction, notes from parents, and teachers’ analyses of their students learning in relation to their instruction. They can be used to document a large range of teaching behaviors and responsibilities.

Has been used widely in teacher education programs and in states for assessing the performance of teacher candidates and beginning teachers.

Examples:
• National Board for Professional Teaching Standards portfolio assessment
• Teacher Performance Assessment (TPA)
• Performance Assessment of California Teachers (PACT)

Measure: Peer Collaboration & Assistance

Description:
Peer assistance (PA) programs are joint, collaborative labor-management programs that focus on enhancing teacher quality by using expert teachers as mentors for probationary or career teachers. (See Evaluation Procedures Section for more).

Examples:
• Massachusetts (Reinventing Educator Evaluation)**
• New York (Teacher Evaluation and Development Handbook)**

Evidence (Humphrey, et. al, p. 22) suggests that CTs tend to conduct more comprehensive assessments than principals who are more typically using a checklist to identify issues that need attention. CTs typically spend more time with teachers during PA or PAR than principals or administrators—and their feedback is often more supportive, less threatening, and broader in scope.

• Works best when designed collaboratively with labor-management stakeholders
• Though PA may cost more initially, many school districts report that over time, PA programs reduce turnover costs, support teacher recruitment and retention, and decrease the time and expense associated with dismissing career teachers
• Can expand capacity for formative feedback
• Can work flexibly within a building or district’s unique schedules and needs
• May feel less threatening than receiving feedback from principal
• Could have more flexibility to have CT with similar experiences, thus feedback would be more valid and reliable

Diea: Scalers should have content knowledge of the portfolios
• The stability of scores may not be high enough to use for high-stakes assessment
• Portfolios are difficult to standardize (compare across teachers or schools)
• Time-consuming for teachers and to a lesser extent scorers
• May not represent day-to-day practice well
• Portfolio scores seldom have been shown to be consistently related to student achievement
• May not include robust measures of student learning

Peer Collaboration & Assistance also must be carried out per OAR 342.850 (2) (b)(E))

Description:
Peer assistance (PA) programs are joint, collaborative labor-management programs that focus on enhancing teacher quality by using expert teachers as mentors for probationary or career teachers. (See Evaluation Procedures Section for more).

Examples:
• Massachusetts (Reinventing Educator Evaluation)**
• New York (Teacher Evaluation and Development Handbook)**

** Please note that both these systems include Peer Assistance and Review components and this chart is only addressing the Peer Assistance part

Measure:

Research on validity and reliability is ongoing, and concerns have been raised about consistency/ stability in scoring. There is a lack of research linking portfolios to student achievement. Some studies have linked NBPTS certification (which includes a portfolio) to student achievement, but other studies have found no relationship.

Examples:
• Is comprehensive and can measure aspects of teaching that are not readily observable in the classroom
• Can be used with teachers of all fields
• Provides a high level of credibility among stakeholders
• Is a good tool for teacher reflection and improvement
• Professionally credible
• Can be done electronically
• May include some assessment of student learning
• Can provide important formative and summative information on teacher practice
• Portfolios represent teachers’ exemplary work

Measure:

Scalers should have content knowledge of the portfolios
• The stability of scores may not be high enough to use for high-stakes assessment
• Portfolios are difficult to standardize (compare across teachers or schools)
• Time-consuming for teachers and to a lesser extent scorers
• May not represent day-to-day practice well
• Portfolio scores seldom have been shown to be consistently related to student achievement
• May not include robust measures of student learning

Peer Collaboration & Assistance also must be carried out per OAR 342.850 (2) (b)(E))
### Measure:

Perception Surveys (student/parents) *(FORMATIVE ONLY)* as only licensed administrators can evaluate teachers per the Teacher Standards and Practices Commission (TSPC), who has been given the authority in statute to license and define the roles of Administrators (ORS 342.121 and ORS 342.140) and they have therefore defined licensed Administrators as the only “teachers” allowed to evaluate and supervise (OAR 584-005-0005) which is further defined and restricted by state bargaining law (Public Employee Collective Bargaining Act) ORS 243.650

### Description:

**Student & Parent Surveys:**

These questionnaires generally ask students & parents to rate teachers on an extent scale/licerht scale (e.g., from 1 to 5, where 1 = very effective, and 5 = not at all effective) regarding various aspects of teachers’ practice (e.g., course content, usefulness of feedback, starting classes on time), as well as how much students say they learned or the extent to which they were engaged. They very often are not used for teacher evaluation at the pre-collegiate level. Used to gather student opinions or judgments about teaching practice as part of teacher evaluation and to provide information about teaching as it is perceived by students.

**Examples:**

**Student Surveys:**

- Ronald Ferguson, Tripod Project Surveys at Harvard University
- Gallup Student Engagement Surveys
- Georgia plans to use teacher-focused student surveys as part of their teacher evaluation system starting at Grade 4 (according to the GA RTTT application)

**Parent Surveys:**

- Ronald Ferguson, Tripod Project Surveys at Harvard University
- Utah’s Pay for Performance pilot program uses parent satisfaction surveys
- Georgia plans to use parent surveys for teachers of students in Grades K-3 (according to the GA RTTT application)

**Student Surveys:**

- Several studies have shown that student ratings of teachers can be useful in providing information about teaching; may be as valid as judgments made by college students and other groups; and, in some cases, may correlate with measures of student achievement. Validity is dependent on the instrument used and its administration and is generally recommended for formative use only.


  Motivated stereotyping of women: “she’s fine if she praised me but incompetent if she criticized me” was done by social psychologists Sinclair & Kunda; found same effect for race in another study

**Student Surveys:**

- Provides perspective of students who have the most experience with teachers
- Can provide formative information to help teachers improve practice in a way that will connect with students
- Makes use of students, who may be as capable as adult raters at providing accurate ratings
- Provides perspective of schools’ primary clients
- Can provide information on how to improve relationships with students
- Have been validated in certain contexts
- Can provide important formative information on teacher practice

**Parent Surveys:**

- Can provide perspective of schools’ primary clients
- Can provide information on how to improve relationships with students
- May provide information that can be used for to inform teaching practice

**Student Surveys:**

- Student ratings have not been validated for use in summative assessment and should not be used as a sole or primary measure of teacher evaluation
- Students cannot provide information on aspects of teaching such as a teacher’s content knowledge, curriculum fulfillment, and professional activities
- Students are not able to observe much of what goes into a teachers’ practice and therefore might not capture important information
- Have not been validated for summative decisions in PK-12 classrooms
- How do very young children participate? Reading levels?

**Parent Surveys:**

- Parents not able to observe much of what goes into a teachers’ practice and therefore might not capture important information about a teachers’ classroom practice
- Have not been validated for summative decisions
- Do not include reliable information on student learning growth
- Participation is often low, therefore resulting information that may not be valid or reliable
- Language barriers, literacy rates, or other factors may prevent parents from participating in surveys.
Student Learning and Growth as a Measure of Teaching Effectiveness

**Student Learning and Growth: Evidence of teachers’ contribution to student learning and growth.**

The Oregon Framework requires at least two student learning goals and the identification of strategies and measures that will be used to determine goal attainment (see table to the right). Teachers also specify what evidence will be provided to document progress on each goal:

1. Teachers who are responsible for student learning in tested subjects and grades (i.e., ELA and mathematics in grades 3-8, 11) will use state assessments as one measure (category 1) and will also select one or more additional measures from category 2 or 3 that provide additional evidence of students’ growth and proficiency/mastery of the standards, and evidence of deeper learning and 21st century skills.

2. Teachers in non-tested (state test) subjects and grades will use measures that are valid representations of student learning standards from at least two of the following three categories, based on what is most appropriate for the curriculum and students they teach. [OEA recognizes this as a state requirement, but this may be a complex task as many subjects/content areas do not have measures outside of category 3.]

---

**Student Learning and Growth Goal-Setting Process**

- Teachers review baseline data and create goals that measure the learning of all students. Goals span a school year or complete course of study. OEA recommends you write growth goals on skill acquisition or content as demonstrated by evidence from multiple student learning measures, and do not write student learning goals about growth in the number of students “passing” particular assessments or student learning measures (i.e., do not write goals that read “10% more students will pass the OAKS reading test,” or “15% more students will receive all 4s or higher on their expository writing samples”). See the next page for examples of SMART goals.

- Teachers collaborate with supervisor/evaluator to establish student learning goals. In addition, teachers may collaborate to establish student learning goals for their grade levels, departments, or curricular teams.

- Teachers will establish at least two student learning goals and identify strategies and measures that will be used to determine goal attainment. They also specify what evidence will be provided to document progress on each goal. OEA recommends that each student learning goal contain multiple measures of student learning and to avoid a workload issue by using measures of student learning already in use in your classroom. For example, in a student learning goal created by a teacher in a “tested grade or subject,” a teacher would create a goal around improving or showing growth in a particular reading skill then would include a variety of student learning measures to show evidence of progress toward this goal, i.e., OAKS, DIBELS, a curriculum-based assessment, and a sample of students’ work/portfolios.

- Teachers complete goal setting in collaboration with their supervisor/evaluator. During the collaborative planning process, the teacher and supervisor/evaluator ensure that quality goal setting occurs through a discussion of the rigor and rationale of each goal, appropriate research-based strategies, quality of evidence and standards addressed. The SMART goal process is used in the development of student growth goals (SMART = Specific and Strategic; Measurable; Action oriented; Rigorous, Realistic, and Results-focused; Timed and Tracked).

- Teachers meet with supervisor/evaluator to discuss progress for each goal mid-year and at the end of the year. Goals remain the same throughout the year, but strategies for attaining goals may be revised.

- Teachers, along with their supervisor/evaluator, reflect on the results and determine implications for future professional growth planning.
**Student Learning Goals**

1. **SMART Goal Format**

   During the 20__-__ school year, __% of students will improve their ____________ skills by __% as measured by evidence from: __measure of student learning___, __measure of student learning___, and __measure of student learning___. **Note – can be more than three measures**

2. **SMART Goal = Writing**

   During the 2012-13 school year, 80% of students will improve their expository writing skills by 10% as measured by evidence from: assessment for chapter(s) 7-9 on expository writing skills, student writing samples graded on district rubric, and curriculum pre-/post-test.

3. **SMART Goal = Math**

   During the 2012-13 school year, 90% of students will improve their math problem-solving skills by 10% as measured by evidence from 2012-13 math problem solving strand data of OAKS, easy CBMs, curriculum assessment for chapter(s) 2-3, and student work samples.

---

**EXAMPLES OF SMART GOALS:**

1. **SMART Goal Format**

   During the 20__-__ school year, __% of students will improve their ____________ skills by __% as measured by evidence from: __measure of student learning___, __measure of student learning___, and __measure of student learning___. **Note – can be more than three measures**

2. **SMART Goal = Writing**

   During the 2012-13 school year, 80% of students will improve their expository writing skills by 10% as measured by evidence from: assessment for chapter(s) 7-9 on expository writing skills, student writing samples graded on district rubric, and curriculum pre-/post-test.

3. **SMART Goal = Math**

   During the 2012-13 school year, 90% of students will improve their math problem-solving skills by 10% as measured by evidence from 2012-13 math problem solving strand data of OAKS, easy CBMs, curriculum assessment for chapter(s) 2-3, and student work samples.

---

**Categories and Measures of Student Learning**

<table>
<thead>
<tr>
<th>Category</th>
<th>Types of Measures of Student Learning</th>
<th>Examples include, but are not limited to:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>State or national standardized tests</td>
<td>Oregon Assessment of Knowledge and Skills (OAKS SMARTER: Balanced when adopted), English Language Proficiency Assessment (ELPA), Extended Assessments</td>
</tr>
<tr>
<td>2</td>
<td>Common national, international, regional, district-developed measures</td>
<td>ACT, PLAN, EXPLORE, AP, IB, DIBELS, C-PAS, other national measures; or common assessments approved by the district or state as valid, reliable and able to be scored comparably across schools or classrooms</td>
</tr>
<tr>
<td>3</td>
<td>Classroom-based or school-wide measures</td>
<td>Student performances, portfolios, products, projects, work samples, tests</td>
</tr>
</tbody>
</table>

---

**Examples of SMART Goals:**

1. **SMART Goal Format**

   During the 20__-__ school year, __% of students will improve their ____________ skills by __% as measured by evidence from: __measure of student learning___, __measure of student learning___, and __measure of student learning___. **Note – can be more than three measures**

2. **SMART Goal = Writing**

   During the 2012-13 school year, 80% of students will improve their expository writing skills by 10% as measured by evidence from: assessment for chapter(s) 7-9 on expository writing skills, student writing samples graded on district rubric, and curriculum pre-/post-test.

3. **SMART Goal = Math**

   During the 2012-13 school year, 90% of students will improve their math problem-solving skills by 10% as measured by evidence from 2012-13 math problem solving strand data of OAKS, easy CBMs, curriculum assessment for chapter(s) 2-3, and student work samples.
OREGON EDUCATION ASSOCIATION’S RECOMMENDED USE OF STUDENT DATA IN TEACHER EVALUATION AND SUPPORT SYSTEMS

### I. SCHOOL LEVEL — INFORM INSTRUCTION AND PROGRAMS

<table>
<thead>
<tr>
<th>Use of Measure(s) of Student Learning</th>
<th>Value of Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorporating multiple valid measures of student learning to be used in a whole school Formative Evaluation process, designed to inform and improve practice</td>
<td>Low Stakes, High Efficacy</td>
</tr>
<tr>
<td>One measure of student learning (ex: just using OAKS/ statewide assessment scores) as sole measure of student learning to be used as a whole school Formative Evaluation process, designed to inform and improve practice</td>
<td>Low Stakes, Low Efficacy</td>
</tr>
<tr>
<td>One measure of student learning to be used as a “trigger” to examine a school’s performance more closely, but not to be used as part of a final or Summative Evaluation, or used for high-stakes decision making</td>
<td>Moderate Stakes, Low Efficacy</td>
</tr>
</tbody>
</table>

### II. INDIVIDUAL TEACHER LEVEL — INFORM INSTRUCTION

<table>
<thead>
<tr>
<th>Use of Measure(s) of Student Learning</th>
<th>Value of Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorporating multiple valid measures of student learning to be used in an individual teacher Formative Evaluation process, designed to inform and improve practice</td>
<td>Low Stakes, High Efficacy</td>
</tr>
<tr>
<td>One measure of student learning (ex: just using OAKS/statewide assessment scores) as sole measure of student learning to be used in an individual teacher Formative Evaluation process, designed to inform and improve practice</td>
<td>Moderate Stakes, Low Efficacy</td>
</tr>
<tr>
<td>One measure of student learning to be used as a “trigger” to examine a teacher’s performance more closely, but not to be used as part of a final or Summative Evaluation, or used for high-stakes decision making such as hiring, firing, or compensation, etc.</td>
<td>Moderate -High Stakes, Low Efficacy</td>
</tr>
</tbody>
</table>

### III. INDIVIDUAL TEACHER LEVEL — EVALUATE PERFORMANCE

*Careful consideration should be given to the “weight” of measures of student learning against other measures*

<table>
<thead>
<tr>
<th>Use of Measure(s) of Student Learning</th>
<th>Value of Use</th>
</tr>
</thead>
</table>
| Incorporating multiple valid measures of student learning and student learning then becoming one of multiple valid measures of teaching effectiveness – multiple measures within multiple measures. | CAUTION: Potential for Moderate to High Stakes, Moderate Efficacy  
  - Most measures of student learning do not have substantial field research validating them as measures of teaching effectiveness  
  - Moderate efficacy due to other (multiple) measures besides student learning measures |
| One measure of student learning to be used as a significant percentage of a Summative Evaluation or other high-stakes decision. | EXTREME CAUTION: High Stakes, Low Efficacy |
| One measure of student learning to be used as a sole measure for a Summative Evaluation system, or sole measure for high-stakes decisions such as hiring, firing, compensation, etc. | STOP: High Stakes, Low Efficacy |
### MEASURES OF STUDENT LEARNING

<table>
<thead>
<tr>
<th>Measure / Description / Examples</th>
<th>Research</th>
<th>Strengths</th>
<th>Cautions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classroom-level: Student growth and/or student learning targets developed collaboratively between teacher and principal for the academic year</strong></td>
<td>Research forthcoming – see “Examples”</td>
<td>• Can be used to gauge teachers’ contributions to outcomes in untested subjects (e.g., social studies, biology, music) • Can focus teachers’ practice on achieving particular outcomes based on their analyses of student learning needs • Can promote collaboration among teachers and instructional leaders • Can incentivize teachers to engage in professional learning opportunities that will help them achieve goals</td>
<td>• Safeguards must be in place to ensure that goals set are feasible yet also high and rigorous • Comparability among teachers may be problematic</td>
</tr>
<tr>
<td><strong>Measure:</strong> Student Growth Objectives or Goal-Driven Portfolio Plans</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Description:</strong> These evaluation tools offer teachers the opportunity to set their own high but feasible objectives for their students’ growth in collaboration with their principal and/or other colleagues. The assessments teachers use may be common or standardized exams or teacher-developed assessments. Some tools require teachers to specify the professional development they will participate in to ensure their students achieve their growth objectives.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Examples:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Student Learning Objectives - Austin Independent School District</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Student Growth Objectives - Denver Public Schools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• New Mexico Professional Development Plans</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Measure:</strong> Classroom Assessments</td>
<td>NWESD has done some research on these measures</td>
<td>• Multiple types • Can be quick or lengthy to administer • Alternative to large-scale tests</td>
<td>• May have reliability issues • Time consuming for individual teachers to create • Comparability among teachers may be problematic</td>
</tr>
<tr>
<td><strong>Description:</strong> Assessments integrated throughout district-wide curricula, or individually developed at the teacher, building, or district level to measure student progress through curricula and compared to standards</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Examples:</strong></td>
<td>NCLB Waiver examples</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## MEASURES OF STUDENT LEARNING

<table>
<thead>
<tr>
<th>Measure / Description / Examples</th>
<th>Research</th>
<th>Strengths</th>
<th>Cautions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School-wide: (building level)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Student growth and/or achievement as determined by approved statewide assessment system (i.e., OAKS, SMARTER)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Measure:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardized Tests:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Description:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Includes, but not limited to, comparable, norm-referenced summative or formative assessments scored and administered in a consistent manner.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Examples:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National, state or district assessments. (See Oregon Framework for Teacher &amp; Administrator Evaluation &amp; Support System on CD-ROM)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- AFT pg 31-34 interesting model from Georgia [http://www.isbe.net/PEAC/pdf/multiple_measures_intro_0411.pdf](http://www.isbe.net/PEAC/pdf/multiple_measures_intro_0411.pdf)

- Quantitative
- Consistent
- Less subjective
- Starts to measure impact of instruction on student performance

- Test score gains are more than an individual teachers effort
- Teachers ratings based on Value-Added models are unstable with large margins of error and dramatic swings
- May promote teaching to the test and leave less time for teaching higher-level thinking skills
- Can punish teachers who work with the most challenging or historically underserved populations of students, as well as those who work with high-level students (floor and ceiling issues)
- No country in the world uses annual test score gains to evaluate their teachers, including the highest-level performing countries
- Doesn’t link specific teacher practices/behavior to improved student performance
- The AYP model compares performance of different cohorts of students
- How do you maintain an evaluation system where only certain teachers have access to standardized tests?
- Most tests not scaled across years.
- Even sophisticated VAM models cannot account for team teaching and other effects.
### MEASURES OF STUDENT LEARNING

<table>
<thead>
<tr>
<th>Measure / Description / Examples</th>
<th>Research</th>
<th>Strengths</th>
<th>Cautions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School-wide: (building level)</strong></td>
<td><strong>Student growth and/or achievement as determined by approved statewide assessment system (i.e., OAKS, SMARTER)</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Measure:**
Measures of Student Engagement or Educational Attainment

**Description:**
These can include classroom observations, and/or self-reports of student engagement, as well as measuring Advanced Placement course participation rates, graduation rates, dropout rates, and student absenteeism.

**Examples:**

- Umatilla School District, Oregon (see CD-ROM)
- Georgia, where according to the Georgia Race to the Top application, “plans to invest in the development, testing, and evaluation of alternative quantitative measures to assess student engagement and student achievement”. The Maryland RTT application suggests that local school systems can “propose alternative priorities for annually measuring student growth and learning, such as—at the high-school level—gains in Advanced Placement participation and exam performance or decreases in the dropout rate”.

The Measures of Effective Teaching Study funded by the Gates Foundation is currently examining a measure of student engagement as a valid predictor of student growth. The Tripod Survey was used in this study to measure student engagement.

- Can assess and incentivize other important teaching outcomes
- Can provide some formative and summative information on the effects of teachers’ practice
- Difficult to attribute individual teacher contribution to such outcomes
- Have not been widely tested in the field to determine validity, reliability, etc.
<table>
<thead>
<tr>
<th>Measure: Measures of Student Work That Show Evidence of Growth</th>
<th>Research forthcoming – see “Examples”</th>
<th>• Can be more authentic assessments of student learning than standardized tests and therefore more valid assessments of teachers’ contributions to student learning</th>
<th>• Validity and reliability not well established</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong> These may take the form of alternative assessments of student learning, for example, writing samples, portfolios of student work, student oral presentations, capstone projects, and the like.</td>
<td></td>
<td>• Can provide important formative and summative information about a teacher’s practice</td>
<td>• Very difficult to standardize and therefore difficult to establish comparability and reliability</td>
</tr>
<tr>
<td><strong>Examples:</strong> Massachusetts will work to develop student performance assessments, and teachers and teacher teams will require training on how to gather student work that demonstrates individual student learning, as well as typical student learning in a class. The state also will develop district-based assessments that are comparable across subjects and grades. Finally, the state will include student work samples as evidence of student learning.</td>
<td></td>
<td></td>
<td>• Psychometric properties of such assessments are not well understood</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure: Gain Score Models or Pre-Test/Post-test Methods*</th>
<th>Research forthcoming – see “Examples”</th>
<th>• Are preferable to “status” models because they indicate change in student learning over time</th>
<th>• Cannot adequately control for students’ background characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong> Gain Score Models measure the difference between an earlier and a later test score, so unlike Value-Added Models (VAMs), they only require two test scores. This approach also relies on vertically equated assessments.</td>
<td></td>
<td>• Can provide information on some teachers’ contributions to student outcomes</td>
<td>• Can be prone to error if tests are too easy or too difficult</td>
</tr>
<tr>
<td><strong>Examples:</strong> Hillsborough County, FL; Eagle County, CO; and MA, MD, and NY among other Race to the Top winners will develop pre- and post-test measures of student learning for teachers in subjects not tested by the state assessment system</td>
<td></td>
<td>• Allows for comparisons between teachers</td>
<td>• Can lead to false attributions of value—that is, in team- or co-teaching situations, cannot separate out contributions of one teacher or another</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Can provide some formative and summative information on the effects of teachers’ practice</td>
<td>• Growth models can encourage non-educative test prep, cheating, etc.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Validity and reliability can only be achieved if the administration of the tests and interpretation of the scores are consistent across all classrooms</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Growth models can dis-incentivize collaboration depending on how they are used</td>
</tr>
</tbody>
</table>
**Measure:**

Value-Added Models

**Description:**

Value-Added Models (VAMs) are currently being used to determine teachers’ contributions to students’ test score gains and also are being used as a research tool (e.g., determining the distribution of “effective” teachers by student or school characteristics). VAMs measure the gains that students make and adjust those gains for student, teacher, or school characteristics. The gains are interpreted as the “value” that a teacher adds because the gains are presumed to be net of all other influences. This presumption is much debated. Stronger VAMs use well-designed vertically equated standardized achievement tests that measure relevant concepts and test students at least three times. Some tests are vertically equated, which means that a given score on the fourth-grade version of a test represents the same level of performance as that same score on the fifth-grade version of the test. OAKS is not a vertically equated assessment.

**Examples:**

- CLASS Project TIF Grant recipients, such as Salem-Keizer School District, Lebanon School District; also McMinville School District
- Tennessee Value-Added Assessment System (TVAAS)
- Ohio’s Education Value-Added Assessment System
- Dallas Value-Added Accountability System (DVAAS)
- Researchers have warned against using value-added estimates for high stakes purposes, including hiring, firing, compensation, and summative evaluations. (Baker, et al., 2010; Braun, 2005; Braun, Chudowsky & Koneigs, 2010; Corcoran, 2020; Kupermintz, 2003)
- Lack of consensus in the research community about whether VAMs can accurately isolate the effects of a single teacher, especially over an extended period. (Goldschmidt, et al., 2005; Koedel & Bett, 2009; Kane & Stanger, 2008)
- Teacher input may be an important influence on achievement, but it is not the only influence. (Braun, 2005)
- VAM numbers do not offer formative feedback about ways to improve instruction or classroom effectiveness. (Miller, 2009; Rowan, Correnti, & Miller, 2002)
- VAMs fail to accurately reflect the limitations of particular tests both for measuring the full range of desired knowledge and skills and for measuring learning gains. (Newton, Darling-Hammond, Haertel & Thomas, 2010)
- Students are not necessarily assigned to classrooms and teachers on a random basis. This has major ramifications on many VAMs. This will severely impact the validity and reliability of a value-added score. Accurate VAMs require comparability across demographics of an individual teacher’s classroom or school setting. (Harris, 2009)
- VAMs are expensive to adopt and maintain. Most school districts cannot afford to upkeep the rigorous data requirements or the degree of human resources and psychometric expertise required. (Goldschmidt, 2005; Harris, 2009)
- Little is known about the validity of value-added scores for identifying effective teaching, though research using value-added models does suggest that teachers differ markedly in their contributions to students’ test score gains. However, correlating value-added scores with teacher qualifications, characteristics, or practices has yielded unstable, invalid, and mixed results and few significant findings. Thus, it is obvious that teachers vary in effectiveness, but the reasons for this are not known.
- As all growth models, VAMs are preferable to “status” or attainment models that measure student proficiency at one point in time because they indicate change in student learning over time
- Can provide information on some teachers’ contributions to student outcomes (only those in tested subjects and grade levels)
- More likely to measure impact of teachers versus other student and school background factors than other kinds of growth models
- Requires no classroom visits because linked student/teacher data can be analyzed at a distance
- VAMs are expensive to adopt and maintain. Most school districts cannot afford to upkeep the rigorous data requirements or the degree of human resources and psychometric expertise required. (Goldschmidt, 2005; Harris, 2009)
- Little is known about the validity of value-added scores for identifying effective teaching, though research using value-added models does suggest that teachers differ markedly in their contributions to students’ test score gains. However, correlating value-added scores with teacher qualifications, characteristics, or practices has yielded unstable, invalid, and mixed results and few significant findings. Thus, it is obvious that teachers vary in effectiveness, but the reasons for this are not known.
- As all growth models, VAMs are preferable to “status” or attainment models that measure student proficiency at one point in time because they indicate change in student learning over time
- Can provide information on some teachers’ contributions to student outcomes (only those in tested subjects and grade levels)
- More likely to measure impact of teachers versus other student and school background factors than other kinds of growth models
- Requires no classroom visits because linked student/teacher data can be analyzed at a distance
- Estimates of teaching effectiveness can be unstable from year to year, which is why most VAMs require three or more years of data
- Difficult to verify the accuracy of the measures due to lack of transparency.
- Value added measures are not available for the majority of teachers
- Results do not provide sufficient information on how teachers can improve their effectiveness
- Can lead to false attributions of value—that is, in team- or co-teaching situations, cannot separate out contributions of one teacher or another
- Requires vertically scaled exams
- Growth models can encourage non-educative test prep, cheating, etc.
- Growth models can dis-incentivize collaboration depending on how they are used
- Models are not able to sort out teacher effects from classroom effects.
- Value-added scores are not useful for formative purposes because teachers learn nothing about how their practices contributed to (or impeded) student learning
- Value-added measures are controversial because they measure only teachers’ contributions to student achievement gains on standardized tests
- Requires no classroom visits because linked student/teacher data can be analyzed at a distance
<table>
<thead>
<tr>
<th>Measure:</th>
<th>Normative Growth Models*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>Normative Growth Models compare growth in student achievement to the amount of growth made by a representative population of students on the same test. A vertical scale is not necessary.</td>
</tr>
</tbody>
</table>
| Examples: | • Massachusetts Student Growth Percentile (SGP) (not currently used to measure teaching effectiveness, but rather to indicate what professional growth track a teacher will be on)  
• New Jersey Student Growth Percentile |

<table>
<thead>
<tr>
<th>Measure:</th>
<th>Categorical Growth Models*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>Categorical Growth Models calculate student growth based on changes in performance category placement (e.g., from “beginning” to “proficient”) from year to year. Changes in all possible category placements are judged subjectively and each is assigned a value that indexes its importance.</td>
</tr>
</tbody>
</table>
| Examples: | • Florida Value Tables  
• Minnesota Growth Model |

### Research forthcoming – see “Examples”

| • Are preferable to “status” models because they indicate change in student learning over time |
| • Can provide information on some teachers’ contributions to student outcomes |
| • Can allow for comparisons between teachers |
| • Unlike other growth models, these do not rely on vertically-scaled exams |

### Cannot be computed for teachers in untested grades and subjects

- Can lead to false attributions of value—that is, in team- or co-teaching situations, cannot separate out contributions of one teacher or another
- Growth models can encourage non-educative test prep, cheating, etc.
- Cannot adequately control for student background characteristics
- Growth models can dis-incentivize collaboration depending on how they are used

---

*Note that one option for using measures of student growth in teacher evaluation is to calculate school-wide student growth using any of the statistical models described in this table and use that as one of multiple measures of teachers’ contributions to school and student outcomes. Depending on how school-wide student growth data is used (e.g., inform teaching practice), it could have the potential benefit of motivating all teachers in the school to work with their students on particular knowledge and skills (for example, all teachers can promote student writing in their classes). However, if used in high-stake situations (e.g., continuation of employment), it may put undue pressure on teachers whose primary responsibility is the tested subject area.*
MEASURES OF STUDENT LEARNING

<table>
<thead>
<tr>
<th>Measure / Description / Examples</th>
<th>Research</th>
<th>Strengths</th>
<th>Cautions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure: Common Formative Assessments</td>
<td>District-determined measures of student learning, comparable across grade or subject district-wide (i.e., common formative and summative assessments)</td>
<td>Inter-rater reliability</td>
<td>- Can be time consuming to create</td>
</tr>
<tr>
<td>Description: Assessments created by school_districts (or potentially at the building level) that are comparable across subjects and grades.</td>
<td>Multiple options</td>
<td>Must plan ahead</td>
<td></td>
</tr>
<tr>
<td>Examples:</td>
<td>Easy to compare results among teachers</td>
<td>Need consensus with other teachers teaching class, if there are any, in creation of assessment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Works well with PLCs</td>
<td>Goal is student achievement, not intended for teacher evaluation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>More equitable for students</td>
<td>Validity and reliability not well established</td>
<td></td>
</tr>
<tr>
<td></td>
<td>May save teachers' time by creating efficiencies</td>
<td>Very difficult to standardize and therefore difficult to establish comparability and reliability</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Can be more authentic assessments of student learning than standardized tests and therefore more valid assessments of teachers' contributions to student learning</td>
<td>Psychometric properties of such assessments are not well understood</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Can provide important formative and summative information about a teacher’s practice</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ADDITIONAL CHALLENGES WITH USING STUDENT LEARNING AND GROWTH DATA

It is also important to consider the following challenges as you plan to use any measures of student learning and growth.

- Teacher preparation and ongoing professional development for measuring student learning growth is limited
- Most principals, support providers, instructional managers, and coaches are not adequately prepared to make judgments about teachers’ contributions to student learning and growth

They need to know how to:
- Evaluate the appropriateness of various measures of student learning for use in teacher evaluation
- Work closely with teachers to select appropriate student growth measures and ensure that they are using them correctly and consistently.

The use of student learning goals based on multiple measures of student learning is a more sound solution, rather than directly connecting individual teachers to their students’ statewide assessment (OAKS) scores.
Using Measures of Student Learning as a Measure of Teaching Effectiveness

Connecting Statewide Test Scores to Teacher Evaluation: An Educationally Sound Approach?

The primary goal for a teacher is to facilitate student learning. Toward that end, most teachers routinely utilize student learning data, including achievement data, to improve their instruction and practice. So it might seem to follow that basing teacher evaluations on student test scores would make sense. Not so fast. In reality, when it comes to measuring teacher performance, a battery of research has demonstrated that test scores are not very reliable at all. As such, test scores should never be used to make high-stakes personnel decisions, nor should student tests, such as statewide, summative, standardized assessments, be utilized as the sole measure of teaching effectiveness in a summative evaluation or the sole measure of student learning.

If you are integrating student assessments as one measure in your evaluation system, remember there are the good, the bad and the ugly. What are some things to watch out for when it comes to utilizing measures based on student learning and growth? First, be mindful of whether achievement data is derived from developmentally appropriate tests. If it is not developmentally appropriate, it is probably not a valid measure. Second, seek reliability. If the measure of student learning is unreliable, it is not valid for evaluating teaching effectiveness. Last, steer clear of student assessments where data is used only in a summative fashion for teacher evaluation and not for purposes of improvement and growth of teaching practice.

Just as it is vital to utilize multiple measures of teaching effectiveness to look holistically at the art and practice of teaching, when using student data one should never rely solely on test scores. It is imperative to use multiple measures of student learning including other forms of evidence of learning like student portfolios and work samples.

The Ultimate Goal is to Improve Teacher Practice and Student Learning and Growth

The primary goal of a quality Teacher Evaluation and Support System (TESS) is to promote professional growth and develop teacher capacity and to provide tools to improve the skills, knowledge and craft required of good teaching. Personnel decisions are only a very small fraction of the entire aim of said systems. Thus, if there are measures within a TESS that do not have a valid research foundation showing they work to improve teaching or learning, they should not be included in a system.

Additionally, there are many ways to incorporate student learning data into a Teacher Evaluation and Support System that will not inadvertently punish teachers for those flaws, but help teachers and educators to strengthen their practice and performance to improve student learning. A sound approach to incorporating student data is through the use of student learning goals. Rather than directly connecting individual teachers to their students’ statewide assessment scores, setting student learning growth goals provides an opportunity to use multiple measures of student learning to provide a more holistic picture of teacher practice and performance.

Using Statewide Test Scores: Why It Remains an Educationally Unsound Approach

Reliable and Valid Data: Oregon’s statewide assessment system (OAKS) has strand level data that is often so unreliable, it should not be used as the sole measure for even large-scale, building-level program decisions. Many school districts use additional formative assessments for the purpose of making programmatic changes. Even so, if statewide test data is not reliable or valid for program-level decisions, it should not be used for individual teacher evaluation.

Test Scores Show Patterns Over Time: While three to four more years of test scores might reliably show student growth or lack thereof, a single year of scores is not a reliable measure of teaching effectiveness.

Calibration for Growth: In order for three to four years of test scores to be viewed, the assessment itself needs to be vertically scaled so educators can compare one year to the next. Oregon’s assessment is not currently calibrated for this purpose. In other words, the assessment is not designed to show student growth from year to year. Until our state assessment is appropriately scaled, using state test scores for this purpose is invalid.

Context Matters: Many factors outside of school influence student achievement and therefore affect student test scores. These include home support, school attendance, family income level, and parents’ level of education. Moreover, there are other factors inside a school, which a teacher does not control, that can greatly impact a student’s academic and social readiness to pass a test. These include principals, peer groups, curricula, the school building/environment, supplies available, courses offered, etc. While the teacher certainly plays the largest in-school role for predicting student success, there are far too many other variables in play to directly link “student outcomes” solely to their “teacher inputs” (see CD-ROM, Research, Teacher Effectiveness on Student Learning).

Multiple Valid Points of Data: Test scores do not speak for themselves. Teacher Evaluation and Support Systems that incorporate student data need multiple valid points of data from both a quantitative and qualitative perspective to truly reflect the whole child and the process of learning. OAKS scores do not accurately reflect many things that teachers do to produce qualitative and other outcomes that cannot be captured on a multiple choice test.
Using Measures of Student Learning
as a Measure of Teaching Effectiveness

(such as higher-order thinking skills). **Using Statewide Test Scores: It is Not a Logistically Possible Solution**

- **Limited Tested Grades:** The state level assessment is given in grades 3-8 and 11. It is not equitable that some teachers may be evaluated on the less reliable statewide standardized assessment, while others may not.
- **Limited Tested Subjects:** The state assessment addresses selected subject areas and does not cover the majority of the content in the school curriculum. Reading, mathematics, science, and writing are the only subject areas assessed on the statewide tests, and they are not consistently assessed in every grade. Again, this creates a systemic inequity.
- **Over 60% of Teaching Staff Are Not Connected to Statewide Assessments:** Combining the grades where the state assessments are not tested and the non-tested subject areas, over 60% of certified teachers/specialists are not even connected with statewide test scores making a comprehensive system using state test scores for teacher evaluations logistically impossible.

**Possible Ramifications of Using Statewide Test Score Data**

- **Less Collaboration and More Competition:** If teachers are being evaluated based on state test scores, they are less likely to collaborate and help their colleagues. More and more research suggests that teacher collaboration is one of the best forms of professional development, and it is also indicative of student learning gains. Why would we want an educational system that does not fully promote collaboration?
- **Teaching to the Assessment:** If student test scores on a single assessment become the basis for teacher evaluation, then the test will become the major instructional focus, crowding out broader learning to an even greater extent.
- **The Campbell Effect:** Generally speaking, the Campbell Effect states that when test scores become the goal of the teaching process, they lose their value as indicators of educational status and distort the educational process in undesirable ways. In other words, the pressure to have students score well on a single test for teacher evaluation becomes so intense that it potentially leads to unscrupulous practices including:
  - Cheating on the test by both students and teachers
  - Data manipulation
  - Distorts education by narrowing the curriculum
  - Distorts education by teaching to the test

The Campbell Effect has been demonstrated in public and private sectors, and its most detrimental effect is demoralizing the workforce charged with carrying out the assessments.

**Designing a Teacher Evaluation and Support System That Supports Growth**

In designing a way to measure teaching effectiveness, it is important to keep a careful balance of things in mind. First, the goal of a new Teacher Evaluation and Support System is improvement. The main goal of evaluation is pointing out areas where teachers need improvement, then providing support so they can grow and develop. Creating a system that promotes growth is an educationally sound solution to increasing teacher quality across Oregon. This is supported by research and connected to good practice. Second, a well-rounded Teacher Evaluation and Support System includes multiple valid points of data to determine teaching effectiveness, collected via multiple valid measures linked to improving student performance including multiple valid measures of student learning. Multiple valid measures provide opportunities for triangulation of data and provide a much stronger and holistic representation of teaching effectiveness. Other considerations in an effective Teacher Evaluation and Support System include:

- Collaboration as a key principle in designing and implementing an evaluation system
- A clear and consistent administrator training and professional development component, including calibration and inter-rater reliability checks
- A comprehensive beginning teacher induction program
- A comprehensive mentoring system including mentor release time for collaboration with mentees
- Ongoing, focused and job-embedded professional development to aid improving instructional practice and student learning, following the Learning Forward standards (see CD-ROM)
- Teachers demonstrating use of student achievement to inform and improve instruction
- Teachers and administrators being trained in demonstrating the impact they have had on student learning using multiple pieces of evidence
- Transparent and timely access to data
- Multiple opportunities for classroom observations by highly-trained administrators
- Self-assessment and self-reflection embedded in teacher evaluation systems
- Collaboration time with colleagues and principal focused on instructional practice and student learning
- Release time for employees to observe mentor(s) and other accomplished colleagues
- Clear link to the InTASC Standards in the evaluation criteria and application of the standards across the TESS
- A differentiated evaluation rubric that measures growth of teaching practices against teaching standards
- Additional administrator support to help implement a growth oriented evaluation system.

When using student data, it is important to not solely use statewide test scores for formative or summative teacher evaluation. It is imperative to use multiple measures of student learning to gain a more holistic assessment of what students know, just as it is vital to utilize multiple measures of teaching effectiveness to gauge the entirety of the art and science of teaching.

*Adapted from Washington Education Association*
# OEA’s Analysis of Usage of a Value-Added Model for Different Purposes

## I. Value-Added Models Used to Inform Instruction and Programs at the School Level

<table>
<thead>
<tr>
<th>Use of VAM</th>
<th>Value of Use</th>
</tr>
</thead>
</table>
| VAM as one measure with other multiple valid measures of student learning to be used in a whole school Formative Evaluation process, designed to inform and improve practice | Low-Stakes, Moderate Efficacy  
  • Reliability & validity in increased because other multiple valid measures are in use (Goe, 2010) |
| VAM as sole measure of student learning to be used as a whole school Formative Evaluation process, designed to inform and improve practice | Low Stakes, Low Efficacy  
  • VAM is a limited diagnostic tool, in part because it cannot demonstrate in-school variables on performance |
| VAM to be used as a “trigger” to examine a school’s performance more closely, but not to be used as part of a final or Summative Evaluation, or used for high-stakes decision making | Moderate Stakes, Low Efficacy |

## II. Value-Added Models Used to Inform Instruction at the Individual Teacher Level

<table>
<thead>
<tr>
<th>Use of VAM</th>
<th>Value of Use</th>
</tr>
</thead>
</table>
| VAM as one measure with other multiple valid measures of student learning to be used in an individual teacher Formative Evaluation process, designed to inform and improve practice | Low-Stakes, Low Efficacy  
  • Although reliability & validity is increased because other multiple valid measures are in use (Goe, 2010), VAM is not designed to give specific formative feedback at the individual level (essentially, its use in this regard is moot – all useful feedback would be coming from the other multiple valid measures). |
| VAM as sole measure of student learning to be used in an individual teacher Formative Evaluation process, designed to inform and improve practice | Moderate Stakes, Low Efficacy |
| VAM to be used as a “trigger” to examine a teacher’s performance more closely, but not to be used as part of a final or Summative Evaluation, or used for high-stakes decision making such as hiring, firing, or compensation, etc. | Moderate -High Stakes, Low Efficacy |

## III. Value-Added Models Used to Evaluate Individual Performance

*Careful consideration should be given to the “weight” of VAM against other measures.*

<table>
<thead>
<tr>
<th>Use of VAM</th>
<th>Value of Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAM to be used as one measure with other multiple valid measures of student learning and student learning then becoming one of multiple measures of teaching effectiveness – a measure within multiple measures within multiple measures.</td>
<td><strong>CAUTION:</strong> High Stakes, Low Efficacy</td>
</tr>
<tr>
<td>VAM to be used as a significant percentage of a Summative Evaluation or other high-stakes decision.</td>
<td><strong>EXTREME CAUTION:</strong> High Stakes, Low Efficacy</td>
</tr>
<tr>
<td>VAM to be used as a sole measure for a Summative Evaluation system, or sole measure for high-stakes decisions such as hiring, firing, compensation, etc.</td>
<td><strong>STOP:</strong> High Stakes, Low Efficacy</td>
</tr>
</tbody>
</table>
A value-added model is a student growth model utilizing achievement data (namely, OAKS scores or other state-wide summative assessments) and additional student background data to attempt to isolate the specific effects of the teacher, school, or program on student academic achievement progress. This data is then used to determine whether or not a student’s average change in performance is meeting a pre-determined growth target. In many Oregon school districts utilizing value-added models (VAMs), a school-level model is being utilized, which attributes student outcomes to the school as a whole, and not to individual teachers.

Proponents of value-added models assert that objective data about whether students have learned must be taken into consideration. They contend that, despite its flaws, a VAM is the best model available to show whether a school’s test scores are improving over time. There are questions, however, about the reliability and validity of these models and the estimates they produce. Therefore, caution should be taken when using a VAM as a basis for high-stakes decisions related to individual teacher performance or pay.

Ultimately, if the purpose of an evaluation system is to improve teaching and learning, then it is important that all measures and models provide diagnostic insight that will help improve the teaching and learning processes. However, VAMs are not diagnostic tools designed to facilitate the improvement of teaching and learning, and thus conflict with this goal.

**VALUE-ADDED MODELS: WHAT TO CONSIDER**
- Researchers have warned against using value-added estimates for high stakes purposes, including hiring, firing, compensation, and summative evaluations. (Baker, et al., 2010; Braun, 2005; Braun, Chudowsky & Koneig, 2010; Corcoran, 2020; Kupermintz, 2003)
- There is a lack of consensus in the research community about whether VAMs can accurately isolate the effects of a single teacher, especially over an extended period and distinguished from other classroom and in-school effects. (Goldschmidt, et al., 2005; Koedel & Betts, 2009; Kane & Staiger, 2008)
- Teacher input may be an important influence on achievement, but it is not the only influence. (Braun, 2005)
- VAM numbers don’t say much about achievement in that they do not tell us why some classrooms are more effective, why some schools are more effective, or why some teachers are more effective. They do not offer formative feedback about ways to improve instruction. (Miller, 2009; Rowan, Correnti, & Miller, 2002)
- VAMs fail to accurately reflect the limitations of particular tests both for measuring the full range of desired knowledge and skills and for measuring learning gains. (Newton, Darling-Hammond, Haertel & Thomas, 2010)
- Students are not necessarily assigned to classrooms and teachers on a random basis. This has major ramifications on many VAMs. This will severely impact the validity and reliability of a value-added score. Accurate VAMs require comparability across demographics of an individual teacher’s classroom or school setting. (Harris, 2009)
- VAMs are expensive to adopt and maintain. Without continuing outside sources of funding, most school districts cannot afford to upkeep the rigorous data requirements or the degree of human resources and psychometric expertise required. (Goldschmidt, 2005; Harris, 2009)
- The potential of having teachers engage in a criteria-referenced measure (like VAM) which measures and ranks teachers against one another clashes with the priority of having a norm-referenced evaluation system (based on InTASC Standards).
- NEA’s Policy on Teacher Evaluations: Teacher evaluations must be comprehensive – based on multiple valid indicators to provide teachers with clear and actionable feedback to enhance their practice – and must include all three of the following components: 1) Indicators of Teacher Practice, 2) Indicators of Teacher Contribution and Growth, and 3) Indicators of Contribution to Student Learning and Growth. Such indicators must be authentic, reflect that there are multiple factors that impact a student’s learning beyond a teacher’s control, and may include the “various indicators” chosen by local or state affiliates. Unless such tests are shown to be developmentally appropriate, scientifically valid and reliable for the purpose of measuring both student learning and a teacher’s performance, such tests may not be used to support any employment action against a teacher and may be used only to provide non-evaluative formative feedback.

**OEA believes:** The strongest indicators of contribution to student learning and growth should come from multiple sources. And while using tests that are shown to be developmentally appropriate, scientifically valid and reliable for the purpose of measuring both student learning and a teacher’s performance is a better alternative, at this time no such instrument exists!
Shared Principles for Use of Value-Added Models Within Alternative Compensation Models

If your district and association leadership are engaged in a CLASS Project or have a Federal Teacher Incentive Fund grant that requires a linkage between evaluation and compensation, below are some guiding principles that can help frame the conversation for developing and implementing an incentive compensation plan. These principles were developed by the National Education Association (NEA), in conjunction with the American Association of School Administrators (AASA) and the National School Boards Association (NSBA):

1. School boards, administrators and unions/associations should review various models of incentive compensation plans, including research about their effectiveness, before developing a plan at the local level.

2. School boards, administrators and unions/associations should work together to build ongoing community and stakeholder support for both the incentive compensation plan as well as the necessary funding.

3. School boards, administrators and unions/associations should work together to develop and implement the plan utilizing collective bargaining where it exists. In locations where collective bargaining does not exist, teachers who would be using the new system should indicate their support for the program.

4. In the implementation of the incentive compensation plan, teachers should be provided assistance, including time, curriculum and professional development to increase student achievement.

5. The foundation of incentive compensation plans shall be professional-level base salaries.

6. Funding for the plan shall be adequate and sustainable.

7. The plan and its requirements should be transparent, easily understood and uniformly implemented.

8. A detailed implementation plan, with agreed-upon benchmarks and timelines, should be developed.

9. The incentive compensation plan should be based on a multi-factor approach (e.g., teacher evaluations, student performance growth, specific goals set by the teachers and management, increased responsibilities, assessments of student learning) that is research-based and improves student achievement.

10. All employees who meet the criteria for the incentive compensation plan should be compensated accordingly, and incentive compensation plans should foster collaboration not competition.

11. Evaluations, if a factor in incentive compensation plans, should be fair, of high quality and rigorous, and shall take into account multiple valid measures of student progress.

Please see OEA’s Professional Pay Toolkit on the CD-ROM.
Steps in an Evaluation and Professional Growth Cycle

**STEP 1: SELF-REFLECTION**
Based on the standards of professional practice, the first step of an evaluation system is self-reflection. The educator reflects on and assesses his/her professional practice and analyzes the learning and growth of his/her students in preparation for goal setting.

**STEP 2: GOAL SETTING**
(Student growth goals and professional goals)
Based on the self-assessment, the educator identifies goals aligned with the standards of professional practice that encompass both practice and impact on student learning. The educator sets both professional practice goals and student learning goals. SMART goals and/or learning targets are used as a tool for effective goal-setting.

**STEP 3: OBSERVATION AND COLLECTION OF EVIDENCE**
(Multiple measures)
The educator and evaluator collect evidence using multiple measures regarding professional practice, professional responsibilities, and student learning to inform progress throughout the process of evaluation.

**STEP 4: FORMATIVE ASSESSMENT/EVALUATION**
(Analysis of evidence, professional conversations and professional growth)
The evaluator and educator review the educator’s progress toward goals and/or performance against standards. This step includes three interdependent and critical parts: analysis of evidence, professional conversations, and professional growth. Both the educator and the observer analyze the evidence leading into a collaborative professional conversation. Feedback through professional conversations promotes awareness of growth that has occurred, and highlights professional growth needs. These conversations help the educator make adjustments in his/her practice and select relevant professional learning opportunities.

**STEP 5: SUMMATIVE EVALUATION**
This step is the culmination of multiple formative observations, reflections, professional conversations, etc. Evaluator assesses the educator’s performance against the standards of professional practice, attainment of student learning goals, and attainment of professional practice goals.

*from ODE Framework for Teacher Evaluation & Support Systems*
Peer Assistance programs can and should be part of a larger system of support and growth for all teachers in the district. The formative nature of peer assistance is a critical and often neglected piece of the full continuum of professional learning as defined by the Learning Forward Standards for Professional Learning (www.learningforward.org). Those standards are based on the profession’s best thinking and research about how to promote outcomes so that every educator engages in effective professional learning every day so every student achieves. Learning Forward Standards for Professional Learning includes developing systems of learning and growth that speak to the following critical elements:

- Learning communities
- Leadership
- Resources
- Data
- Learning Designs
- Implementation
- Outcomes

Peer Assistance programs can offer structured and nuanced formative feedback to teachers about their performance as well as create a community of learners that inform the professional development that is offered both individually and collectively in a school or district. A PA program must align to the larger evaluation and support system being developed by the district and the association in order to be effective and sustainable. The greatest potential benefit is that instructional support aligns with the evaluation system and is delivered by trained peers whose interest is improvement.

It is important to clearly articulate the roles and responsibilities of peer coaches so that the confidentiality firewall is protected. Otherwise, you not only risk jeopardizing the critical trust between the peer coach and their charge, but precariously venture into summative evaluation territory. In the state of Oregon, existing statutes and administrative rules separate a peer coach from a summative evaluator. (See PA (Peer Assistance) Versus PAR (Peer Assistance Review))

Peer assistants can be called many things: mentors, lead teachers, critical friends, peer coaches, instructional coaches

Their principal role is to guide and mentor teachers through constructive feedback tied directly to the standards being used in the evaluation system. Their primary goal of that feedback is to support the growth and development of their peers (thereby improving retention) but ultimately it can serve to elevate the professional learning of every teacher in the system. PAs can assist not only probationary teachers but also career teachers who may need assistance or are developing new skills. Besides observing teachers and providing feedback specific to the standards in the evaluation system, they might also be assisting their peers to do the following:

- Identify teaching goals
- Outline and seek professional growth activities
- Provide support and induction with lesson plans, parent communication, assessing student learning, differentiating instruction, finding necessary resources, etc.

Peer assistants should be carefully chosen and purposefully trained to perform these tasks within the guidelines of the PA program but also to serve as skillful mentors that stimulate reflective practice and continued growth. A Peer Assistance Committee, jointly designed by the local association and district, should provide the following governance duties:

- Develop program parameters (to include but not exclusive to)
  - Roles and responsibilities of peer assistants
  - Roles and responsibilities of teachers receiving coaching
  - Protocol for matching peer coaches with mentees (and creating a protocol for changing matches if necessary)
  - Clarifying and creating culturally competent equity standards for the program
  - Schedule of support
  - Articulated activities and tasks
  - Firewall protocols
  - Alignment to evaluation system and professional learning activities
  - Roles and responsibilities of administrators and other staff in supporting this program
  - Recruiting criteria and training for peer coaches
  - Identification of who receives peer coaching and when

- Recruit, train and evaluate peer assistants
- Oversee the program budget
- Facilitate ongoing evaluation and improvement of program
Safeguards should exist to do the following:

- Peer coaches do not replace the duties of a licensed administrator who is responsible for summative evaluation.
- Peer coaches do not provide information or testimony that is used in a summative evaluation.
- Peer coaches do not breech confidentiality guidelines with other staff.

The following guiding principles for any Peer Assistance Program are recommended in the NEA Teacher Evaluation and Accountability Toolkit:

- Peer Assistance Programs must be collaboratively developed and overseen by joint labor-management committees.
- Before any program is implemented the local association must educate members and build support for it. Any acceptance of this program as part of a collective bargaining agreement hinges on whether or not members believe the PA program is fair, their rights are protected and it has value for them as individuals and a collective.
- Peer coaches must be carefully selected for their skills and commitment to the program.
- Peer coaches who also serve other duties must have adequate time to devote to the preparation and implementation of their duties as peer coaches.
- Peer coaches remain bargaining unit members and do not cross over into summative evaluation.
- Confidentiality and firewall protocols are carefully and clearly defined in the joint committee rules as well as the collective bargaining agreement.
- Additional compensation for peer coaches needs to be significant enough, reflecting the additional responsibilities, learning, collaboration and time involved. This compensation should be in addition to continued compensation packages and steps owed to them.

The school district should provide a sustainable, stable funding source for the program.

A data tracking system should be developed so that the joint PA committee can analyze, modify and improve the program accordingly. Samples of data that team may be interested in are:

- Number of teachers served by the program (divided by probationary and career teachers)
- Financial data
- Caseload data
- Time served by peer assistants
- Retention over time
- # of Plans of Assistance over time
A variety of approaches to teacher observation support professional growth and student achievement. The following are several of those methods:

**LESSON STUDY**
In this three-pronged approach designed by Japanese educators, teachers collaboratively develop a lesson, observe it being taught to students, and then discuss and refine it.

**PEER COACHING**
In this non-evaluative professional development strategy, educators work together to discuss and share teaching practices, observe each other’s classrooms, provide mutual support, and, in the end, enhance teaching to enrich student learning.

**COGNITIVE COACHING**
Teachers are taught specific skills that involve asking questions so that the teacher observed is given the opportunity to process learning associated with teaching the lesson.

**CRITICAL FRIENDS GROUP (CFG)**
This program provides time and structure in a teacher’s schedule for professional growth linked to student learning. Each CFG is composed of eight to 12 teachers and administrators, under the guidance of at least one coach, who meet regularly to develop collaborative skills, reflect on their teaching practices, and look at student work. [See an Education World article, Critical Friends Groups: Catalysts for School Change.]

**LEARNING WALK**
The Learning Walk, created by the Institute for Learning at the University of Pittsburgh, is a process that invites participants to visit several classrooms to look at student work and classroom artifacts and to talk with students and teachers. Participants then review what they have learned in the classroom by making factual statements and posing questions about the observations. The end result is that teachers become more reflective about their teaching practices. Professional development is always linked to The Learning Walks.

*From “Teachers Observing Teachers: A Professional Development Tool for Every School”*
RELEVANT EXAMPLES OF BARGAINING LANGUAGE, RESEARCH AND OTHER RESOURCES:

- New York: Rochester Teachers Association, Section 53, Intervention, Remediation, and Professional Support
  o Bargaining language for working with career teachers: Page 142 of the NEA Teacher Evaluation and Accountability Toolkit

- New Jersey: Support on Site summarizes the New Jersey Education Association’s induction program
  o Page 108 & 109 of the NEA Teacher Evaluation and Accountability Toolkit

- Supporting Research on both Peer Assistance and Peer Assistance and Review Programs
  o Page 21 of the NEA Teacher Evaluation and Accountability Toolkit

- Boston Teachers Union (www.btu.org/member-resources)
  o Website advertises both the reasons why you might want a peer coach and the benefits you might realize in the following way:

PEER ASSISTANCE PROGRAM:
WE’RE HERE TO HELP

- Are you overwhelmed by the district’s mandates related to your content area or instructional implementation?
- Have you recently received administrative feedback that was a surprise to you?
- Are you looking to refine or polish your teaching craft? Do you need support?

PEER ASSISTANCE PROGRAM:
ABOUT OUR WORK

We realize that each teacher’s situation is different. While all our work is confidential and non-evaluative, together we might:

- Focus on the Eight Dimensions of Effective Teaching
- Reflect to increase student engagement and learning
- Find and create teaching resources and materials
- Collaborate with other teachers and school leaders.

On-site support is dictated by your individualized needs. Together we will decide the length and focus of our work.

PA (PEER ASSISTANCE) VERSUS PAR (PEER ASSISTANCE REVIEW)

Peer Assistance (PA) programs are typically joint, collaborative labor-management programs that focus on enhancing teacher quality by using a cadre of experienced and expert teachers as mentors or coaches for probationary and/or career teachers in their formative assessment. Peer Assistance Review (PAR) programs are similar except that cadre of coaches is also involved in the summative evaluation process of other teachers.

See the CD-ROM for a longer resource on how in Oregon, only PA is allowed per Oregon Statutes, Oregon Administrative Rules, and Oregon Public Employee Collective Bargaining Agreement.
Questions about this Guidebook?

CONTACT
Center for Great Public Schools
503.684.3300
oea-gps@oregoned.org

Oregon Education Association
6900 SW Atlanta St.
Portland, OR 97223

www.oregoned.org