Teacher and Principal Evaluation Pilot

Report to the Legislature

Randy I. Dorn
State Superintendent of
Public Instruction

July 2011

OSPI provides equal access to all programs and services without discrimination based on sex, race, creed, religion, color, national origin, age, honorably discharged veteran or military status, sexual orientation including gender expression or identity, the presence of any sensory, mental, or physical disability, or the use of a trained dog guide or service animal by a person with a disability. Questions and complaints of alleged discrimination should be directed to the Equity and Civil Rights Director at (360) 725-6162 or P.O. Box 47200 Olympia, WA 98504-7200.

TEACHER AND PRINCIPAL EVALUATION PILOT

Prepared by
Michaela W. Miller, Program Director

K-12 Education/Teacher and Principal Evaluation
Office of Superintendent of Public Instruction
Michaela W. Miller, Program Director

Randy I. Dorn
Superintendent of Public Instruction

Alan Burke, Ed.D.
Deputy Superintendent, K-12 Education

July 2011

Table of Contents

Executive Summary……...…v

I. Introduction..... 1

A. Legislative Background

1. Summary of E2SSB 6696

2. Timeline/System Development

3. Foundation of the Evaluation Systems

a. Revised Teacher and Principal Criteria

b. Four-Level System

i. Teachers

ii. Principals

c. Using Student Growth Data in Evaluations

4. TPEP Steering Committee

5. TPEP Pilot Districts

a. TPEP Pilot Site Overview

II. Process 8

A. Evaluation System Implementation Status

1. Evaluation System Overview

2. Major Components Common to all Pilots

a. Common Statewide Revised Criteria Definitions

b. Common Statewide Tier Labels

c. Common Statewide Tier Summative Statements

d. Common Statewide Satisfactory/Not Satisfactory Dilineation

3. District Determined Components

a. Instructional and Leadership Framework

b. Multiple Measures of Teacher and Principal Performance

III. Findings 16

A. Research and Reports

B. Blueprint for Changing and Implementing the Washington State Evaluation System

IV. Next Steps……..19

A. TPEP Pilot Sites

B. State-level Taskforce Committees

V. Recommendations 20

A. Superintendent Dorn's Recommendation

1. One or Multiple Models

2. Evaluation System Components

3. State-Approval Process

4. Challenges to a State Evaluation Approval Process

VI. Conclusion 22

VII. Bibliography..... 22

VIII. Appendices 24

A. Teacher Evaluation Model Visual Diagram
B. Principal Evaluation Model Visual Diagram
C. Washington Teacher/Principal Evaluation Pilot Cross-Case Analysis
- http://tpep.files.wordpress.com/2011/07/washington_tpep_cross_case_analysis.pdf

D. Washington Teacher/Principal Evaluation Pilot Case Studies and Evaluation Models
1. Anacortes
   - http://tpep.files.wordpress.com/2011/07/0390-anacortes_casestudy.pdf
   - http://tpep-wa.org/asd-eval-models/
2. Central Valley
   - http://tpep.files.wordpress.com/2011/07/0390-central-valley_casestudy.pdf
   - http://tpep-wa.org/cvsd-eval-models/
3. Kennewick
   - http://tpep.files.wordpress.com/2011/07/0390-kennewick_casestudy.pdf
   - http://tpep-wa.org/ksd-eval-models/
4. NEWESD 101 Consortium
   - http://tpep.files.wordpress.com/2011/07/0390-newesd-consortium_casestudy.pdf
   - http://tpep-wa.org/consortium-eval-models/
5. North Mason
   - http://tpep.files.wordpress.com/2011/07/0390-north-mason_casestudy.pdf
   - http://tpep-wa.org/nmsd-eval-models/
6. North Thurston
   - http://tpep.files.wordpress.com/2011/07/0390-norththurston_casestudy.pdf
   - http://tpep-wa.org/ntps-eval-models/
7. Othello
   - http://tpep.files.wordpress.com/2011/07/0390-othello_casestudy.pdf
   - http://tpep-wa.org/osd-eval-models/
8. Snohomish
   - http://tpep.files.wordpress.com/2011/07/0390-snohomish_casestudy.pdf
   - http://tpep-wa.org/ssd-eval-models/
9. Wenatchee
   - http://tpep.files.wordpress.com/2011/07/0390-wenatchee_casestudy.pdf
   - http://tpep-wa.org/wsd-eval-models/

E. Washington Teacher/Principal Evaluation Statewide Data Analysis Report
- http://tpep.files.wordpress.com/2011/07/washington-tpep-statewide-data-report.pdf

F. OSPI Practitioner Panel Review
- Practitioner Workgroups Proposal April 2011
  - http://tpep.files.wordpress.com/2011/07/practitioner-workgroups-proposal-april-2011.pdf
- TPEP Model Review Templates
  - http://tpep.files.wordpress.com/2011/07/tpep-model-review-templates-compiled.pdf
- TPEP Practitioner Panel Participants
  - http://tpep.files.wordpress.com/2011/07/tpep-practitioner-panel-participants.pdf
- TPEP Practitioner Panels System & State Feedback
  - http://tpep.files.wordpress.com/2011/07/tpep-practitioner-panels_system-state-feedback.pdf

G. OSPI School Employee Evaluation Survey Summary and Data
1. OSPI School Employee Evaluation Survey PDF Summary

- http://www.k12.wa.us/Communications/StimulusPackage/SEES_Summary_PDF_SY2009-10.pdf

2. **OSPI School Employee Evaluation Survey Data File**
   - http://www.k12.wa.us/Communications/StimulusPackage/SEES_Summary_Data_SY2009-10.xlsx

H. Washington Teacher/Principal Evaluation Pilot Site Interview Compilation (Video)

- http://tpep.files.wordpress.com/2011/05/tpep-site-interviews.m4v

List of Tables

Table 1	Comparison of Existing Evaluation Requirements and Revised E2SSB 6696 Components
Table 2	TPEP Pilot District and Statewide Implementation Timeline
Table 3	Revised Teacher and Principal Criteria
Table 4	TPEP Steering Committee Members and Organizations
Table 5	Participating TPEP Pilot Districts
Table 6	TPEP Pilot Site Overview
Table 7	Evaluation System Overview Glossary
Table 8	Common Statewide Revised Teacher Criteria Definitions
Table 9	Common Statewide Revised Principal Criteria Definitions
Table 10	Draft Common Statewide Tier Labels
Table 11	Draft Common Statewide Tier Summative Statements - Teacher
Table 12	Teacher Instructional Frameworks by District
Table 13	Measures Under Consideration by District - Teacher
Table 14	Measures Under Consideration by District - Principal
Table 15	State-level Taskforce Committees

Executive Summary

Background

The Teacher & Principal Evaluation Project (TPEP), which was created in Engrossed Second Substitute Senate Bill 6696 (E2SSB 6696) in the 2010 Legislative Session, offers Washington State the opportunity to identify the measures of effective teaching and leading. The new Washington State evaluation system must both hold educators accountable and be leverage for authentic professional growth. This emerging system, built on the foundation of the new teacher and principal criteria and developed by Washington State educators, provides a direction that will empower teachers, principals and district leaders to meet the needs of students in Washington State. The new evaluation system sets high expectations for what teachers and principals should know and be able to do, values diversity, and fosters a high commitment to teaching and leading as professional practice.

Setting the Context

According to the Joyce Foundation, by the end of 2010 twelve states had passed new state teacher/principal evaluation laws. Washington State is included in this bold group of states that embarked on a journey of creating a new system for measuring teacher and leadership performance. The research over the past 10 years establishing the critical importance of quality teachers and leaders (Barber & Mourshed, 2007; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Rivkin, Hanushek, & Kane, 2005; Rockoff, 2004) prompted policymakers to look to evaluation as a lynchpin to education reform. It is important to note that the other areas impacting teacher and leadership quality should not be overlooked and are inextricably linked to evaluation. “Such areas include these components of the educator career continuum: traditional certification, alternative certification, mentoring and induction, professional development, evaluation, compensation, equitable distribution, and tenure.” (Behrstock, Meyer, Wraight, & Bhatt, 2009).

Educators in Washington State overwhelming agree, the current evaluation system requires a much needed overhaul. During the 2010-11 school year, OSPI conducted a survey and ten forums with nearly 7,000 educators outside of the TPEP pilot sites and found that 80 percent indicated the primary purpose of the current evaluation system was compliance. Practitioners in and out of the TPEP pilot sites “want tools for improvement and growth.” (Fetters, J., & Behrstock-Sherratt, E., 2011). All indications are that Washington State took the right step to enact E2SSB 6696 and the strong belief that the evaluation changes will produce positive results for our students.

Summary of E2SSB 6696

The new law requires Office of Superintendent of Public Instruction (OSPI) to work in collaboration with organizations representing teachers, principals, district administrators, and parents to develop new evaluation models for both classroom teachers and principals. Representatives of the following organizations serve on the TPEP Steering Committee:

- Office of Superintendent of Public Instruction
- Washington Education Association
- Association of Washington School Principals
- Washington Association of School Administrators
- Washington State Parent-Teacher Association
- Washington State School Directors' Association (May 2011)

The Legislation also:

- Revised the evaluation criteria for both classroom teachers and principals;
- Created a four-level rating systems;
- Required OSPI to create a pilot with school districts in the 2010-11 (development) and 2011-12 (pilot implementation) school years;
- All districts to adopt new systems in the 2013-14 school year;
- Increased the length of the provisional status for new teachers; and
- Requires school districts to send OSPI information on the current evaluation system for all employee groups beginning in 2010-11 school year.

TPEP Funding

In the 2010 Supplemental Operating Budget, the Legislature provided $1.9 million in funding for the pilot project and statewide implementation activities for Fiscal Year 2011. More than half of the allocation was distributed in grants to the selected pilot districts and consortium. The allocation chart can be found on our OSPI website.
http://www.k12.wa.us/Communications/PressReleases2010/PrincipalTeacherEvaluationAllocations.pdf

TPEP Pilot Districts

Districts that are participating in the pilot program include:

- Anacortes
- Central Valley
- Kennewick
- North Mason
- North Thurston
- Othello
- Snohomish
- Wenatchee
- Almira
- Davenport
- Liberty
- Medical Lake
- Pullman
- Reardan-Edwall
- Ritzville
- Wilbur

Development Year

During the 2010-11 school year, the TPEP project studied the current evaluation research, worked with national experts, and relied heavily on the practitioners in the pilots to guide the design of their new evaluation systems. Over the course of the year, with the guidance of the

TPEP districts, common evaluation components were identified. These components will translate to a more consistent evaluation system of teacher and principal performance state-wide. At the same time, decision-making regarding specific local or regional factors was honored. The key to Superintendent Dorn's final recommendations will be finding the delicate balance that will maintain a rigorous state-wide evaluation system and value the local differences. Case studies were written that captured each of the pilot site development (Appendix D). In addition, a cross-case analysis was produced (Appendix C) that looked at the work of the TPEP districts as a whole.

The TPEP evaluation models will be piloted in the 2011-12 school year, and results analyzed in the summer of 2012.

TPEP Communication

OSPI took deliberate steps during the 2010-11 school year to create a transparent development process. In August 2010, OSPI launched the TPEP website (*http://tpep-wa.org/).* This website allowed the project to update and provide resources to both the pilot sites and the larger stakeholder community. The website currently houses all of the meeting agendas and materials the project used throughout the year for the development. Most importantly, the website contains the teacher and principal evaluation systems for each participating TPEP district and the consortium.

Superintendent's Recommendations

E2SSB 6696 requires the Superintendent of Public Instruction to make recommendations in this report regarding whether a single statewide evaluation model should be adopted, whether modified versions developed by school districts should be subject to state approval, and what the criteria would be for determining if a school district's evaluation model meets or exceeds a statewide model. The report also is to identify challenges posed by requiring a state approval process.

Summarized below are the Superintendent's recommendations, which are on the work of the TPEP Pilot sites and TPEP Steering Committee input. Please see the recommendation section for more details.

1. Districts should be encouraged to select from a limited number of state-approved teacher and principal evaluation models. However, it is recommended that a state approval process be developed for those districts not wanting to select from the state-approved models.

2. The Superintendent will require that all systems have specified components. These components will include, but not limited to:

    - Research-based instructional and leadership frameworks which utilize rubrics based in a four-tiered evaluation system.
    - State-adopted evaluation criteria, definitions, tier labels, summative statements, and a common satisfactory/not satisfactory delineation.
    - Multiple measures for determining effective teacher and principal performance.
    - Professional learning for principals and classroom teachers that will include training for all evaluators on the components of an evaluation system.

3. At the conclusion of the pilots, the Superintendent shall finalize the components and requirements that must be included in the evaluation systems. All school districts will be required to include all of the components as specified by the Superintendent.

   During the 2012-13 school year, school districts should be required to submit a description of their proposed evaluation systems that they intend to use beginning in the 2013-14 school year. The description of the system shall include how they will address each of the required components, which will be subject to a thorough, rigorous state review process conducted by OSPI with the assistance of principals, teachers, and administrators.

4. The challenges to a state review process rest in two areas: time and resources. The capacity over the next two years to approve all models will be time consuming and require state-level expertise and consultation to remain intact and be enhanced. If the system is to be functioning at a high level during the 2013-14 state-wide implementation year, serious consideration will need to be given to providing targeted resources to prepare all the districts in an intentional way for the new teacher and principal evaluation system.

I. Introduction

The acts of teaching and leading are incredibly complex tasks involving multiple and varied skills and knowledge. Research has established the critical importance that school leaders and classroom teachers have in impacting student learning (Barber & Mourshed, 2007; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Rivkin, Hanushek, & Kane, 2005; Rockoff, 2004). Research also has demonstrated that teacher evaluations are often sporadic and poorly designed and, as a result, provide little useful information on teacher effectiveness (Brandt, Thomas, & Burke, 2008; Weisberg, Sexton, Mulhern, & Keeling, 2009). To this end, the Washington State Legislature enacted E2SSB 6696 in the spring of 2010 to overhaul a generation-old evaluation system for both certificated classroom teachers and principals.

A. Legislative Background

1. Summary of E2SSB 6696

The significant shifts in requirements from the existing evaluation system to the new one envisioned in E2SSB 6696 cannot be understated. Nearly eight of ten administrators surveyed this year felt that the current evaluation system is used only for compliance. The table below identifies four of the most significant changes in the existing evaluation law.

Table 1

Comparison of Existing Evaluation Requirements and Revised E2SSB 6696 Components		
Component	**Existing Evaluation System**	**Revised E2SSB 6696 Evaluation System**
Tiers	Binary (Satisfactory/Not Satisfactory)	Four-tiered Professional Growth and Development System
Criteria	Criteria developed more than 25 years ago.	Criteria that describes effective teaching and leading developed by organizational stakeholder groups in the 2010 legislative session.
Provisional Status	2 years (prior to 2009/10 SY)	3 years
Educator Evaluation Data	No existing requirement	Requires evaluation data submitted to OSPI for all employee groups beginning 2010-11 SY

2. Timeline/System Development

Over the course of the two-year pilot, the district participants have three main tasks as outlined in E2SSB 6696. First, the pilots are to create models for implementing the evaluation criteria. Second, they are to create models for student growth tools. Finally, the pilots are to create professional development programs and evaluator training for teachers and principals.

Within those three tasks, there are seven specific responsibilities the districts are required to develop and pilot:

- Development of rubrics for evaluation criteria and ratings;
- Development of appropriate evaluation system forms;

- Identification of, or development of, appropriate multiple measures of student growth;
- Submission of data that is used in evaluations and all district-collected student achievement, aptitude and growth data (regardless of whether it is used in evaluations);
- Participation in professional development opportunities for principals and classroom teachers regarding the content of the new evaluation system;
- Participation in evaluator training; and
- Participation in activities to evaluate the effectiveness of the new system and support programs.

Timeline

Table 2 describes the timeline for both the TPEP pilot sites and the proposed timeline and activities generated for districts outside of the pilot districts. These proposed activities were developed after careful consideration and analysis of best practice generated from our TPEP pilots. Further information about the pilot development and state-wide implementation plan can be found on our OSPI external website (*http://tpep-wa.org/*), which contains updated information about the constantly evolving world of educator evaluation in Washington State. The site contains information about our TPEP 2010-11 meetings/agendas, district evaluation models and background information about the national and state landscape of educator evaluation. This resource is one OSPI intends to maintain and enhance as the project develops.

Table 2

TPEP Pilot District and Statewide Implementation Timeline
Pilot Districts **2010-11** • Develop Models/Tools/Rubrics *(OSPI Report due July 1, 2011)* **2011-12** • Implement Pilot Models/Tools/Rubrics *(OSPI Report due July 1, 2012)* • Pilot Districts engage in professional development, including inter-rater reliability training, instructional framework training for teachers/principals, and leadership training for teacher leaders, principals and district administration **2012-13** • Refine models, participate in evaluation professional development and evaluator training
All Districts **2010-11** • Observation of Pilot Development *Resource: TPEP website: http://tpep-wa.org/* • Stakeholder engagement and communication • Participation in Educator Evaluation Regional Forums (2010 -11) **2011-12** • Observation of Pilot Implementation *Resource: TPEP website: http://tpep-wa.org/* • Stakeholder engagement and communication • Participation in Educator Evaluation Regional Academies (2011-12) • TPEP Implementation Consortium Grants *(information provided in July 2011)*

2012-13

- Identification of Evaluation Models *(upon conclusion of TPEP pilot recommendations in June 2012)*
- Participate in Evaluation Professional Development and Evaluator Training

2013-14

- Full state-wide implementation of new teacher and principal evaluation systems
- Participate in Evaluation Professional Development and Evaluator Training

3. Foundation of the Evaluation Systems

The foundation of "what we know for sure" about Washington State's new teacher and principal evaluation systems rest in two important sections of the new evaluation law. First, Washington State has revised teacher and principal evaluation criteria that will apply to all school districts beginning in the 2013-14 school year. Second, Washington State is moving from a required binary summative rating of satisfactory/not satisfactory to a four-level professional growth and development model. These two revisions in the law form the underpinnings of the work of the TPEP pilots in 2010-11. In addition, much attention nationally has focused on the issue of student growth data in evaluations and the project took steps to ensure that all aspects of a complex new evaluation system were considered by providing research, resources and national experts regarding multiple measures.

a. Revised Teacher and Principal Criteria

Washington State has not had new evaluation criteria for certificated classroom teachers and principals in more than 25 years. Educators and stakeholders involved in the project agree that the change to the new criteria sets a clearer definition of effective teaching and leading. In some cases, the revised criteria lack the definitions needed to clearly distinguish one from another. Table 3 includes the current and revised teacher and principal evaluation criteria.

Table 3

Revised Teacher and Principal Criteria	
Current **Teacher Evaluation Criteria**	**Revised** **Teacher Evaluation Criteria**
1. Instructional skill 2. Classroom management 3. Professional preparation and scholarship 4. Effort toward improvement when needed 5. Handling of student discipline and attendant problems 6. Interest in teaching pupils 7. Knowledge of subject matter	1. Centering instruction on high expectations for student achievement 2. Demonstrating effective teaching practices 3. Recognizing individual student learning needs and developing strategies to address those needs 4. Providing clear and intentional focus on subject matter content and curriculum 5. Fostering and managing a safe, positive learning environment 6. Using multiple student data elements to modify instruction and improve student learning 7. Communicating and collaborating with parents and school community 8. Exhibiting collaborative and collegial practices focused on improving instructional practice and student learning

Current **Principal Evaluation Criteria**	**Revised** **Principal Evaluation Criteria**
1. Knowledge of, experience in, and training in recognizing good professional performance, capabilities and development 2. School administration and management 3. School finance 4. Professional preparation and scholarship 5. Effort toward improvement when needed 6. Interest in pupils, employees, patrons and subjects taught in school 7. Leadership 8. Ability and performance of evaluation of school personnel	1. Creating a school culture that promotes the ongoing improvement of learning and teaching for students and staff 2. Demonstrating commitment to closing the achievement gap 3. Providing for school safety 4. Leading the development, implementation, and evaluation of a data-driven plan for increasing student achievement, including the use of multiple student data elements 5. Assisting instructional staff with alignment of curriculum, instruction, and assessment with state and local district learning goals 6. Monitoring, assisting, and evaluating effective instruction and assessment practices 7. Managing both staff and fiscal resources to support student achievement and legal responsibilities 8. Partnering with the school community to promote student learning

b. Four-Level System

According to a School Employee Evaluation Survey coordinated by OSPI (required by section 4 (d) of E2SSB 6696), 209 of the 289 reporting districts in Washington State used a binary (satisfactory/not satisfactory) summative system for evaluating certificated classroom teachers in 2009-10. Although many districts appear to use a tiered approach during the evaluation cycle, none used it in 2009-10 as a final summative evaluation. According to the survey data collected for the Department of Education as a requirement of State Fiscal Stabilization Funding (SFSF), 99 percent of our certificated classroom teachers were deemed satisfactory. (See http://www.k12.wa.us/Communications/StimulusPackage/FiscalStabilization.aspx). This is consistent with findings from other national reports on this issue, specifically the 2009 report "The Widget Effect" from The New Teacher Project (Weisberg, Sexton, Mulhern, & Keeling, 2009).

According to our OSPI School Employee Evaluation Survey, 196 out of 295 school districts in Washington State used a binary (satisfactory/not satisfactory) summative system for evaluating principals in 2009-10. Although not as widely scrutinized at a national level, the same data appears to hold true when OSPI surveyed districts regarding the principal evaluation system. In the required SFSF reporting from the 2009-10 SY, 98 percent of all principals were deemed satisfactory on their summative evaluations in Washington State.

For both classroom teachers and principals, E2SSB 6696 requires a four-level evaluation rating system that must describe performance along a continuum that indicates the extent to which the criteria have been met or exceeded."

c. Using Student Growth Data in Evaluations

The Obama Administration, influential national organizations, and others have advocated that student growth data be an integral component in the evaluation of teachers and principals. The Washington Legislature, after considering the technical challenges of measuring student growth and other factors, chose to include an evaluation criterion regarding the teacher's use of student data to inform instruction and a principal's use of multiple student data points, but did not require that student growth data be included in the evaluation of teachers or principals to increase student achievement. However, E2SSB 6696 does state that "…when student growth data, if available and relevant to the teacher and subject matter, is referenced in the evaluation process, it must be based on multiple measures that can include classroom-based, school-based, district-based, and state-based tools. As used in this subsection, "student growth" means the change in student achievement between two points in time." Similar language also was included regarding using student growth data in the evaluation of principals.

The issue of student growth and the new teacher and principal evaluation system is multi-faceted and hinges on many other parts of a district or state instructional and human resource system. It also requires an extensive pre- and post assessment system in multiple disciplines and detailed information regarding which teacher or teachers actually taught the student the content area being assessed. Every district and state engaged in this work across the country has wrestled with these complex questions and there are very few definitive answers to such high stakes propositions.

TPEP districts have worked throughout the year, with the guidance of the E2SSB 6696 language, to decide what best fits their district that would be valid measures of student growth and how to attribute this growth accurately to individual teachers. The TPEP districts have and continue to tackle both formative and summative growth measures. All TPEP participants agree that student learning plays a significant role in the development of these new evaluation systems and that a focus on teacher and leader professional growth will only further impact student learning. According to national experts at the National Comprehensive Center for Teacher Quality, "Multiple measures of teacher outcomes allow for a more comprehensive view of a teacher's effectiveness based on a variety of evidence. Although summative student achievement data are relevant, data on teacher performance are most useful for targeting professional development and specifically addressing areas in which growth is needed."

Because of the complexity of the topic, during the next seven months OSPI will directly address the issue of using student growth data in evaluations by bringing together experts and practitioners from Washington State to identify specific recommendations for using growth data in evaluations. This Student Growth Taskforce will be one of three committees that TPEP will form by August 2011, which will make recommendations in February 2012. The other two committees will make recommendations regarding evaluator training, inter-rater reliability and the use of perception survey data.

4. TPEP Steering Committee

The legislation requires OSPI, in collaboration with state associations representing teachers, principals, administrators, and parents, to create models for implementing the evaluation system criteria, student growth tools, professional development programs, and evaluator training for certificated classroom teachers and principals. OSPI created the TPEP steering committee to oversee and monitor the policy direction and decisions of the TPEP Pilot work. Members of the Steering Committee are listed in Table 4.

Table 4

TPEP Steering Committee Members and Organizations	
Jonelle Adams	Washington School Directors Association (WSSDA) *added May 2011*
Alan Burke	Office of Superintendent of Public Instruction (OSPI)
Bob Butts	Office of Superintendent of Public Instruction (OSPI)
Judy Hartmann	Office of the Governor
Ramona Hattendorf	Washington State Parent Teacher Association (WSPTA)
Jim Koval	Office of Superintendent of Public Instruction (OSPI)
Gary Kipp	Association of Washington School Principals (AWSP)
Michaela Miller	Office of Superintendent of Public Instruction (OSPI)
Scott Poirier	Washington Education Association (WEA)
Paula Quinn	Association of Washington School Principals (AWSP)
Ann Randall	Washington Education Association (WEA)
Leslie Rose	Office of Superintendent of Public Instruction (OSPI)
Paul Rosier	Washington Association of School Administrators (WASA)
Bill Williams	Washington State Parent Teacher Association (WSPTA)
Lucinda Young	Washington Education Association (WEA)

(In alphabetical order)

One of the key elements of the success of the TPEP work thus far has been the intentional collaboration among the stakeholders outlined in the legislation. The collaboration at the state-level was a model for the expectation of the pilot districts to work as a team to ensure the work is done with stakeholder involvement. The TPEP Steering Committee met 15 times during the 2010-11 year to make joint policy decisions about the direction of the project. All members approved the final version of this report.

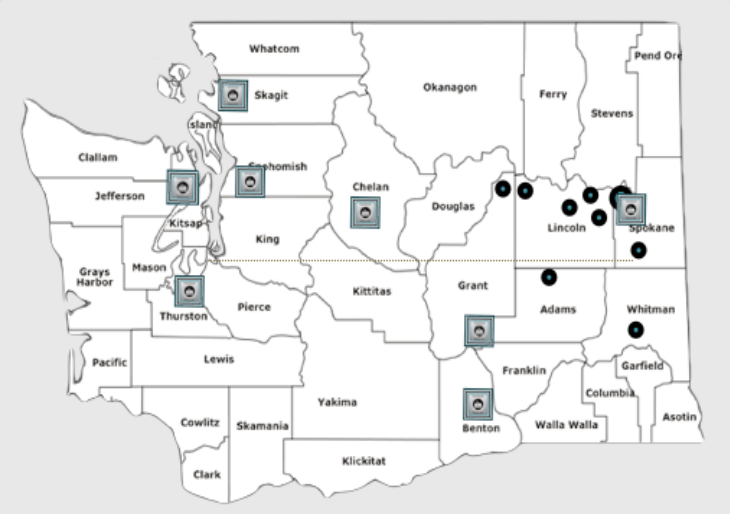
5. TPEP Pilot Districts

The pilot consists of eight districts and one consortium of smaller districts working with the TPEP steering committee organizations to develop nine new and innovative teacher and principal evaluation systems that comply with the legislation and lead to better teaching and learning.

The school districts participating in the pilot include:

Table 5

Participating TPEP Pilot Districts	
8 Districts	**1 Consortium**
Anacortes	Almira
Central Valley	Davenport
Kennewick	Liberty
North Mason	Medical Lake
North Thurston	Pullman
Othello	Reardan-Edwall
Snohomish	Ritzville
Wenatchee	Wilbur



a. TPEP Pilot Site Overview

The data table below shows information regarding the pilot districts, including general demographic student and employee information and the number of teachers, principals and district administrators participating in the pilot.

Table 6

TPEP Pilot Site Overview							
District Name	**Approximate Number of Students**	**Number of Schools**	**Number of Teachers**	**Number of Principals**	**Teachers in Pilot**	**Principals in Pilot**	**District Admin-istrators in Pilot**
Anacortes	2,700	7	147	8	140	8	2
Central Valley	12,400	22	722	36	54	36	2
ESD 101 Consortium	6,563	22	403	29	78	23	11
Kennewick	16,000	25	822	41	75	22	2
North Mason	2,200	6	132	7	30+	7	2
North Thurston	14,000	24	826	30	124	31	2
Othello	3,700	5	190	12	35	5	2
Snohomish	10,000	23	497	21	107	20	3
Wenatchee	7,700	14	456	19	56	18	3

II. Process

A. Evaluation System Implementation Status

1. Evaluation System Overview

This overview was developed with the TPEP pilot sites to create an understanding of the system changes required in order to fully implement the new evaluation law. The 2010-11 school year was considered a development year with the focus on learning, understanding system changes and detailing the operations of the new evaluation systems. The work was divided into two descriptors: “The What” and “The How.”

Table 7

Evaluation System Overview Glossary	
Teacher	**Principal**
The What Describes the development of the components of the new evaluation system. These pieces must be developed in order to implement the new evaluation, but are not the only pieces of the system. The “what” components are illustrated in Visual “A” (teachers) and Visual “B” (principals).	
Revised Teacher Evaluation Criteria	**Revised Principal Evaluation Criteria**
The Legislature passed E2SSB 6696 and Governor Gregoire signed the bill into law (RCW 28A.405.100) on March 29, 2010. The criteria formed the backbone of the new evaluation system. The TPEP districts have used the evaluation criteria language and existing or new instructional frameworks to develop the rubrics. According to the RCW, “the four-level rating system used to evaluate the certificated classroom teacher must describe performance along a continuum that indicates the extent to which the criteria have been met or exceeded.”	The Legislature passed E2SSB 6696 and Governor Gregoire signed the bill into law (RCW 28A.405.100) on March 29, 2010. The criteria formed the backbone of the new evaluation system. The TPEP districts have used the evaluation criteria language and resources such as AWSP’s document “Evaluating Principal Leadership in a Performance-Based School” to develop the rubrics. According to the RCW, “the four-level rating system used to evaluate the principal must describe performance along a continuum that indicates the extent to which the criteria have been met or exceeded.”
Criteria Definitions	**Criteria Definitions**
Based on feedback from experts and our TPEP districts, we have created definitions for each of the new teacher criterion. Each of the TPEP districts submitted definitions and we synthesized those into one brief sentence. This is intended to delineate the criteria in order to minimize the overlap between the criteria and create more consistency across the state in setting clear evaluation targets for teachers and principals as we move to statewide implementation.	The Criteria Definitions for the state’s principal criteria are being developed by AWSP and will be available soon. The research-based definitions are an extension of the AWSP document “Evaluating Principal Leadership in a Performance-Based School.”
Comprehensive Instructional Framework	**Leadership Framework**
The comprehensive instructional framework (common language/model of instruction) provides districts with a way to talk about instruction that is shared by everyone in the district/ESD. Dr. Robert Marzano states that teachers and principals use the instructional framework “to converse about effective teaching, give and receive feedback, collect and act upon data to	Districts in the pilot relied heavily on the work done by AWSP that culminated in the document “Evaluating Principals Leadership in a Performance-based School”. Districts also reported using the standards from the national organizations for principals, National Board for Professional Teaching Standards, Val-Ed and the work of Dr. Robert Marzano and Doug Reeves.

monitor growth regarding the reasoned use of the strategies, and align professional development needs against the framework." There are several instructional frameworks being utilized by the TPEP districts. Because the new teacher criteria are unique to Washington, the TPEP districts have aligned the instructional frameworks (and subsequently the rubrics) to the new state criteria.	Unlike an instructional framework for teachers, the leadership frameworks are not tied to a classroom experience and therefore have content beyond that of the classroom that reflect the varied work of the school principal.
Rubrics	**Rubrics**
The rubrics (based on the instructional frameworks) are the clearly defined continuum that describes unsatisfactory through effective teaching practice based on the 8 teacher criteria. The rubrics should be used to train principals to identify strengths and weaknesses in practice, based on clearly defined evidence and measures. These rubrics could take into account the variations of novice to expert teachers.	The rubrics are the clearly defined continuum that describes unsatisfactory through effective leading based on the 8 principal criteria. The rubrics should be used to train district administrators to identify strengths and weaknesses in practice, based on clearly defined evidence and measures. These rubrics could take into account the variations of novice to expert principals.
Measures and Evidence	**Measures and Evidence**
The measures and evidence are used to determine the "teacher's performance along a continuum that indicates the extent to which the criteria have been met or exceeded." The measures used in the evaluation system should have strong correlation to the criteria being evaluated. There are four areas under the "measures and evidence" section: classroom observation, teacher self-assessment, student growth data, other measures/evidence. This section should represent the district's system for determining final summative evaluation score.	The measures and evidence are used to determine the principal's performance along a continuum that indicates the extent to which the criteria have been met or exceeded." The measures used in the evaluation system should have strong correlation to the criteria being evaluated. There are four areas under the "measures and evidence" section: observation, perception data, student achievement growth data, and other measures/evidence. This section should represent the district's system for determining final summative evaluation score.
Final Summative Evaluation	**Final Summative Evaluation**
The final summative evaluation is a critical definition in order to increase consistency across the state as teachers are evaluated and data is submitted in aggregate. In the late fall 8 of the 9 TPEP sites and WASA submitted a summative evaluation statement for each of the 4 tiers. Similar to the standards-based system for students, clear targets for both the distinct criteria and the final summative evaluation will drive principals and teachers to an evaluation system that promotes growth and prevents stagnation.	The final summative evaluation is a critical definition that increases consistency across the state as principals are evaluated and data is submitted in aggregate. Similar to the standards-based system for students, clear targets for both the distinct criteria and the final summative evaluation will drive districts to an evaluation system that promotes growth and prevents stagnation.
The How	
Equally important to the new components of the system are the policies, procedures and culture surrounding the design and implementation of the new evaluation models.	
Stakeholder Engagement:	
The TPEP project has been a collaborative process from the beginning. Successful development and ultimately implementation has and will require looking at this process through multiple lenses. Please include any documents your district/consortium has used to incorporate authentic stakeholder engagement through the pilot development year. (This will include the norms and protocols you used in setting up your district's TPEP steering committee.)	

Communication:
Communication is a key component to successful development and implementation of the new evaluation system. The collaborative approach at both the state and district levels is critical. Include the plan and documents that would explain your communication process.
Professional Development:
Looking at this new evaluation system as a process in continuous improvement, professional development to train the staff involved in the pilot will be key. Please include your district's plan for ongoing professional development for your teachers, principals and district administrators involved in the 2011-12 TPEP pilot year.
Data:
Many aspects of the new teacher and principal evaluation system will depend heavily on the acquisition and use of data. Include a description of resources your district already uses relating to instructional data and any additional resources you will need to implement the new evaluation system. (Include any technology, databases related to teacher, student, and/or principal data.)
Forms & Tools:
Many parts of the new evaluation system will require changing the forms and tools used in the evaluation process. Please include and forms and tools developed for the new evaluation process. (Please note which ones are electronic and which are paper-based.) Examples: Principal observation tools (pre, during and post), MOUs, artifact collection and observation tools, parent or student surveys, etc.

2. Major Components Common to all Pilots

In addition to the state required common components (new criteria and four-level system), the Superintendent recommends the additional common state-level components. In order to create a performance-based evaluation system with meaningful aggregated state data, some common components will need to be established across all district evaluation systems.

a. Common Statewide Revised Criteria Definitions

Tables 8 and 9 list the draft definitions for the revised teacher and principal criteria. Nationally recognized evaluation experts have provided advice and guidance to the TPEP project during the 2010-11 development year. These advisors encouraged the state to more clearly articulate and distinguish the criteria. In order to establish a consistent performance-based evaluation system, a common set of agreed upon definitions are critical to the new evaluation system.

In collaboration with the TPEP pilot sites, OSPI and the TPEP Steering Committee organizations developed the following definitions, which are in draft form. The TPEP pilot sites may be using slightly different definitions connected to their rubrics. We will continue working over the course of the pilot year to refine the following definitions and establish the final version at the conclusion of the 2011-12 TPEP pilot.

Table 8

Common Statewide Revised Teacher Criteria Definitions	
Revised Teacher Evaluation Criteria	**DRAFT Criteria Definitions**
1. Centering instruction on high expectations for student achievement.	PLANNING: The teacher sets high expectations through instructional planning and reflection aligned to content knowledge and standards. Instructional planning is demonstrated in the classroom through student engagement that leads to an impact on student learning.

2. Demonstrating effective teaching practices.	INSTRUCTION: The teacher uses research-based instructional practices to meet the needs of ALL students and bases those practices on a commitment to high standards and meeting the developmental needs of students.
3. Recognizing individual student learning needs and developing strategies to address those needs.	REFLECTION: The teacher acquires and uses specific knowledge about students' individual intellectual and social development and uses that knowledge to advance student learning.
4. Providing clear and intentional focus on subject matter content and curriculum.	CONTENT KNOWLEDGE: The teacher uses content area knowledge and appropriate pedagogy to design and deliver curricula, instruction and assessment to impact student learning.
5. Fostering and managing a safe, positive learning environment.	CLASSROOM MANAGEMENT: The teacher fosters and manages a safe, culturally sensitive and inclusive learning environment that takes into account: physical, emotional and intellectual well-being.
6. Using multiple student data elements to modify instruction and improve student learning.	ASSESSMENT: The teacher uses multiple data elements (both formative and summative) for planning, instruction and assessment to foster student achievement.
7. Communicating and collaborating with parents and school community.	PARENTS AND COMMUNITY: The teacher communicates and collaborates with students, parents and all educational stakeholders in an ethical and professional manner to promote student learning.
8. Exhibiting collaborative and collegial practices focused on improving.	PROFESSIONAL PRACTICE: The teacher participates collaboratively in the educational community to improve instruction, advance the knowledge and practice of teaching as a profession, and ultimately impact student learning.

Table 9

Common Statewide Revised Principal Criteria Definitions	
Revised Principal Evaluation Criteria	**DRAFT Criteria Definitions**
1. Influence, establish and sustain a school culture conducive to continuous improvement for students and staff.	CULTURE: Simply put, culture is the way things get done. Principals influence the culture of a school in many ways. Exemplary principals assure that all classroom cultures maximize learning; they also impact all non-classroom areas and non-class time, with teacher and student leaders, to establish healthy norms which support learning.
2. Lead the development and annual update of a comprehensive safe-schools plan that includes prevention, intervention, crisis response and recovery.	SAFETY: The principal is ultimately responsible for the safe operations of the school. This includes both classroom and school-wide procedures. Principals in Washington are required to have and monitor a school plan that would provide for the safest operations possible.
3. Lead the development, implementation and evaluation of the data-driven plan for improvement of student achievement.	PLANNING: Today's principal leads using plans which are supported by evidence. Whether it is student achievement data, discipline data, school climate perception data, or other measures of school success, using data in planning is crucial. Data provides both the rationale and target for concerted action to move the school forward.

4. Assist instructional staff in aligning curriculum, instruction and assessment with state and local learning goals.	ALIGNMENT: Principals monitor and assist teachers, not just in the "how" of teaching, but also the "what." Aligning the curriculum, instruction and assessment within each class increases the likelihood that alignment from class to class happens, and students' learning experiences are connected.
5. Monitor, assist and evaluate staff implementation of the school improvement plan, effective instruction and assessment practices.	SUPERVISION: Principals assist and support teacher professional development through the evaluation process. They ensure that all students have teachers with strong instructional skills and dedication to the achievement of each student, by leading the hiring, evaluation and development of each teacher.
6. Manage human and fiscal resources to accomplish student achievement goals.	MANAGEMENT: Principals make resource decisions to achieve learning, safety, community engagement and achievement gap goals. These decisions include hiring and firing staff, maximizing financial resources, and organizing time, facilities and volunteers.
7. Communicate and partner with school community members to promote student learning.	COMMUNITY: Principals link the school to the community and visa versa. They assist teachers in connecting their students' learning to parent and community support.
8. Demonstrate a commitment to closing the achievement gap.	THE GAP: Principals monitor gaps between various populations in the school. They channel resources to reduce the gaps to ensure that all students have the maximum opportunity to achieve at high levels.

b. Common Statewide Tier Labels

Table 10 identifies the tier labels each TPEP district submitted during the development year to describe each level of the new four-tiered system.

Table 10

Draft Common Statewide Tier Levels				
Pilot site	**Level 1**	**Level 2**	**Level 3**	**Level 4**
Anacortes	Unsatisfactory	Emerging	Proficient	Exemplary
Central Valley	Not Demonstrated/ Unsatisfactory	Developing	Proficient	Accomplished
Kennewick	Unsatisfactory	Emerging	Proficient	Exemplary
North Mason	Unsatisfactory	Basic/Emerging	Proficient	Distinguished
North Thurston	Unsatisfactory	Basic	Proficient	Distinguished
Othello	Unsatisfactory	Basic	Proficient	Innovative
Snohomish	Unsatisfactory	Emerging	Proficient	Exemplary
Wenatchee	Unsatisfactory	Basic	Proficient	Distinguished
Consortium	Unsatisfactory	Basic	Proficient	Distinguished

c. Common Statewide Tier Summative Statements (Teacher)

The teacher draft summative statements for the new teacher evaluation system were developed in collaboration with the TPEP Steering Committee organizations and the TPEP

Pilot Sites. At the conclusion of the pilot, Superintendent Dorn will make the final summative statement recommendations in the report completed July 1, 2012. The summative statements for principals are still being finalized and will be added as an addendum to the report before the start of the 2011-12 TPEP pilot year.

Table 11

Draft Common Statewide Tier Summative Statements - Teacher	
1	Professional practice at Level 1 does not show evidence of understanding or demonstration of the concepts underlying individual components of the criteria. This level of practice is ineffective and may represent practice that does not contribute to student learning, professional learning environment, or effective teaching practice. This level requires immediate intervention and specific district support. Failure to show adequate growth is grounds for dismissal/nonrenewal.
2	Professional practice at Level 2 shows a developing understanding and demonstration of the concepts underlying individual components of the criteria but performance is inconsistent. This level may be considered minimally competent for teachers early in their careers or experienced teachers in a new assignment, but insufficient for more experienced teachers. This level requires specific and relevant support.
3	Professional practice at Level 3 shows evidence of thorough knowledge of all aspects of the profession. This is successful, accomplished, professional, and effective practice. Teachers at this level thoroughly know academic content, curriculum design/development, their students, and a wide range of professional resources. Teaching at this level utilizes a broad repertoire of strategies and activities to support student learning. At this level, teaching is strengthened and expanded through purposeful, collaborative sharing and learning with colleagues as well as ongoing self-reflection and professional improvement.
4	Professional practice at Level 4 is that of a master professional whose practices operate at a qualitatively different level from those of other professional peers. Teaching practice at this level shows evidence of learning that is student directed, where students assume responsibility for their learning by making substantial, developmentally appropriate contributions throughout the instructional process. Ongoing, reflective teaching is demonstrated through the highest level of expertise and commitment related to all students' learning, challenging professional growth, and collaborative leadership.

Adapted from "Framework for Teaching Levels of Performance Sample Operational Definitions" created by Pam Rosa, Danielson Group Associate

d. Common Statewide Satisfactory/Not Satisfactory Delineation (Teachers and Principals)

Eight of the nine TPEP pilot districts have agreed that the satisfactory/not satisfactory line should be between a Level 1 and Level 2 for both teachers and principals. The ninth district has brought forward concerns related to the policies for the rest of the tiers. Specifically, the policies related to teaching practice at Tier "2". The concerns have been taken into account by the steering committee.

Superintendent Dorn recommends that the satisfactory/not satisfactory delineation will be between a 1 and a 2 for the purposes of the pilot and recommendations regarding further changes to the RCW regarding Tier "2" will be included in the July 1, 2012 report.

3. District Determined Components

According to the May 2011 publication "*A Practical Guide to Designing Comprehensive Teacher Evaluation Systems"* from the National Comprehensive Center for Teacher Quality, "…states now must decide the extent to which the teacher evaluation model will make allowances for local flexibility and provide a balance between local and state control that encourages collective responsibility and accountability." Throughout the TPEP pilot work and based on evidence across the evaluation work across the country, there are components of any evaluation system that will have the most direct impact on student learning by ensuring stakeholder decision making at the district level.

a. Instructional and Leadership Framework (Teachers and Principals)

The Instructional Frameworks listed below (Table 12) served as the foundation for the rubrics designed by the TPEP districts. Districts listed under the Comprehensive Instructional Frameworks (teachers only) will continue to work with Danielson, Marzano and Center for Educational Leadership (CEL) frameworks to ensure there is alignment between the comprehensive instructional framework and the Washington State criteria. The comprehensive instructional framework is the research-based observation tool that covers all eight of the revised Washington State teacher criteria. The importance of the instructional and leadership frameworks to the development of the evaluation models cannot be underscored. Measuring teacher and principal performance will hinge on the clarity and usability of the observation instruments being developed by the TPEP districts through the use of instructional and leadership frameworks.

Table 12
Teacher Instructional Frameworks by District

District	Modified Version	Comprehensive Instructional Frameworks: Danielson	Comprehensive Instructional Frameworks: Marzano	Comprehensive Instructional Frameworks: 5-D (CEL)	AWSP	Marzano/Reeves	5-D (CEL)
Anacortes				X	X		X
Central Valley	X		X			X	
Kennewick	X				X		
North Mason		X			X	X	
North Thurston		X			X		
Othello	X				X		
Snohomish		X			X		
Wenatchee			X		X	X	
Consortium		X			X		

b. Multiple Measures of Teacher and Principal Performance

The TPEP districts were intentional throughout the year to investigate and select measures within their district TPEP teams that met both the new evaluation criteria and other elements of E2SSB 6696. Table 13 and 14 describe the measures that are currently under consideration in each of the pilot sites. The changes to this new paradigm of evaluation are vast, but perhaps the biggest change rests in the variety of measures used to capture information about teacher and principal performance. The TPEP districts will continue to refine and put these measures into practice next year in the pilot. Recommendations will be forthcoming at the conclusion of the 2011-12 pilot year.

Table 13
Measures Under Consideration by District - Teacher

District	**Observation**	**Student Data**	**Perception Surveys**	**Portfolio**	**Lesson Plans**	**Professional Growth Plans**	**Student Artifacts**	**Peer Evaluation**	**Feedback Meetings**	**Self-Assessment**
Anacortes	X	X	X	X	X		X	X		X
Central Valley	X	X			X	X	X		X	X
Kennewick	X	X			X	X	X	X	X	X
North Mason	X	X			X		X			X
North Thurston	X	X			X	X	X		X	X
Othello	X	X			X		X			X
Snohomish	X	X			X	X	X		X	X
Wenatchee	X	X		X	X	X	X	X	X	X
Consortium	X	X	X	X	X	X	X		X	X

Table 14
Measures Under Consideration by District - Principal

District	Observation	Building Level Student Achievement Data	Formative Assessment Results	Summative Assessment Results	Perception Survey	Portfolios	Artifacts	School Failure Rates	School Safety and Climate Plans	Other
Anacortes	Model still in development									
Central Valley	X				X		X		X	School improvement plans
Kennewick	X	X	X	X	X		X		X	
North Mason	Model still in development									
North Thurston	Model still in development									
Othello	Model still in development									

Snohomish	X	X	X		X		X		X	
Wenatchee	X	X	X	X	X	X	X		X	
Consortium	X	X	X	X	X		X	X	X	

III. Findings

A. Research and Reports

OSPI ensured a careful data collection system for the development year, focusing squarely on seeking feedback on the evaluation changes in the law and ongoing feedback from the TPEP pilot sites on the development of the new models. Summarized below is a list that describes both the state-wide and TPEP pilot site data collections OSPI directed during the 2010-11 development year.

Teacher and Principal Evaluation Project (TPEP) Pilot and Statewide Survey and Reports

State-wide surveys and interviews

- **OSPI Statewide Evaluation Data DOED Survey**
  12/15/10 – 1/21/11
  - This collection of educator evaluation data was required by the DOED upon receipt of ARRA funds.
  - 294 of the 295 districts participated in this survey.

- **TPEP Statewide Evaluation Electronic Survey**
  2/15/11 – 3/15/11
  - This electronic survey was disseminated to all certificated classroom teachers, principals and district administrators in Washington State.
  - Purpose: To better understand current evaluation practices and how to best support implementation of E2SSB 6696.

- **TPEP Statewide Interviews**
  4/1/11 – 4/15/11 – interviews; 4/15 – 4/30 – interview analysis
  - Follow-up to Statewide Evaluation Electronic Survey.
  - Purpose: These interviews were completed to follow-up to the electronic survey in order to gather feedback and experiences from districts regarding the current evaluation system and hopes for the future.

- **TPEP Educator Forums**
  2/7/11 – 3/9/11
  - Participants: 340 educators
  - Purpose: To provide outreach from the state and local pilot site perspective and collect feedback regarding current evaluation practices and hopes for future evaluation systems.

Pilot-specific work

- **Pilot Interviews and Model Review**
  - Interviews with TPEP pilot teachers, principals and district administrators were conducted to gather feedback and experiences from pilot districts regarding the process of developing an evaluation system.
  - The TPEP evaluation model development was analyzed by a group of American Institutes for Research (AIR) evaluation experts and provided to each pilot district.

- **Pilot Focus Group (Consortium)**
  - A similar process was used for the consortium to gather feedback and experiences from the consortium participants regarding the process of developing an evaluation system.

- **TPEP Individual Case Studies and Cross -Case Analysis**
  - A summary report documenting the process and implementation of the new evaluation models will be produced for each TPEP site and a cross-case analysis of the TPEP project will be completed.

- **TPEP Practitioner Panel Review**
  (Teachers, Principals, Superintendents, Professional Development, Human Resource, Data, & Finance Experts)
  - 35 Practitioners from outside of the TPEP sites were selected by the TPEP steering committee to review, analyze, and offer suggestions in the evaluation system development and implementation process for both the TPEP pilot sites and the Statewide TPEP work.

B. Blueprint for Changing and Implementing the Washington State Evaluation System

The TPEP project made it a priority to not only seek input from the selected TPEP pilot sites, but also seek feedback from practitioners outside of the pilots to ensure stakeholders participate in the evaluation system development throughout the entire three-year implementation process. The state-wide data collected through the project produced the following “blueprint” for changing and implementing the Washington Evaluation System. The data also presented three distinct challenges to implementation. This blueprint and challenges demonstrate the congruency between TPEP pilot sites’ development and the input from the general field of practitioners in Washington State.

Blueprint

The data collected from the TPEP project this year determined the new system should be:

1. Focused squarely on improving teaching and learning.
2. Guided by instructional frameworks that reflect the most up-to-date, evidence-based practices for teaching and leading.
3. Supported by rubrics that include clear descriptions of practice and performance; multiple rating levels; and multiple measures to stimulate conversation and reflection that support improved practice for student learning.
4. Providing intensive evaluator training.

Challenges

The challenges to changing and implementing the Washington State evaluation system are grouped in three areas:

- Culture Shifts
- Data
- Professional Development/Training

1. Changing the culture from one of evaluation as compliance to one of professional growth.

    - *78 percent of district administrators responding to the OSPI Evaluation Survey indicated that the primary purpose of their district's teacher evaluation system was compliance.*
    - *Fewer than 25 percent of administrators responding to the OSPI Evaluation Survey report using rubrics to evaluate teachers.*

2. Determining and including student growth and perception survey data are components needing more study.
    - *Telephone Interviews conducted with TPEP pilot participants and educators across the state cautiously suggest OSPI consider including student growth data as one measure for educator effectiveness.*
    - *Telephone interviews and focus group participants overwhelmingly suggest OSPI wait to include teacher value-added scores, unless in the pilot educator evaluation settings.*
    - *Telephone interviews with administrators and teachers reported wanting some flexibility in determining the mix of measures used in a district's evaluation system.*

3. Professional development and training must be a priority for future implementation of the new Washington State evaluation systems.
    - *Data collected from the OSPI educator evaluation survey suggest that classroom and school based observations are inconsistent in timing and value.*
    - *Teachers and administrators indicate they are hard-pressed to adjust and monitor their instructional and leadership practices when the feedback is out-of-date or loosely tied to instructional or leadership challenges.*
    - *Administrators and teachers indicate that in the new evaluation system, time and ratio of evaluator to evaluatee will be a significant factor in the success of the implementation.*

IV. Next Steps

A. TPEP Pilot Sites

The TPEP pilot sites will continue their work over the summer and begin piloting their models fall 2011. OSPI will continue to work with the American Institute of Research, the TPEP Steering Committee organizations, and other supporting partners to support the work of the pilot.

OSPI will be responsible for analyzing the work of the pilots through the 2011-12 SY. This will include looking at the model implementation, use of student data, and evaluator training among other areas. OSPI will continue to work to involve stakeholders in consultation as statewide implementation approaches.

B. State-level Taskforce Committees

A key learning of the work of the TPEP project has been the understanding that this change is one that is ongoing and multi-faceted. A state system should not wait another 25 years before another update. The fields of teaching and leading have changed dramatically over the past 25 years and even over the course of the first year of TPEP development, emerging research has changed our course several times.

Any solid reform enlists a process of evaluation and feedback. The TPEP project should follow suit. It has been the vision of the project from the beginning to have practitioners at the heart of the work. This will continue through three very specific committees. The areas of student growth, evaluator training, and perception survey data have presented challenges to the TPEP districts. In an effort to support the TPEP districts and the rest of the state, the TPEP Steering Committee has recommended three taskforce committees research and present recommendations regarding the inclusion of these areas in our teacher and principal evaluation systems.

Table 15

State-level Taskforce Committees		
Taskforce	Organizational Lead	Supporting Organizations
Student Growth Data	OSPI	WEA, AWSP, WASA, WSPTA, WSSDA
Evaluator Training and Inter-rater Reliability	AWSP	OSPI, WEA, WASA, WSPTA, WSSDA
Perception Survey Data	AWSP/WSPTA	OSPI, WEA, WASA, WSSDA

The TPEP pilot sites have tackled a great deal this year in their work and every district in the state that will follow their work closely over the next year should laud the progress they have made. However, as we have watched in other states and large districts across the country, there are many areas that more study will yield better and more articulated options for State Superintendent Dorn's final recommendation completed July, 2012.

Timeline
The committees will be formed in August 2011 and run through February 2012. They will be comprised of 2/3rds. TPEP practitioners and 1/3rd experts from the field. These committees will present Superintendent Dorn, the TPEP Steering Committee, and the TPEP pilot sites with research-based best practices and guidance around the three areas outlined above.

V. Recommendations

E2SSB 6696 requires the Superintendent of Public Instruction to make recommendations in this report regarding whether a single statewide evaluation model should be adopted, whether modified versions developed by school districts should be subject to state approval, and what the criteria would be for determining if a school district's evaluation model meets or exceeds a statewide model. The report also is to identify challenges posed by requiring a state approval process.

Superintendent Dorn has based the following recommendations on the work of the TPEP Pilot sites and the TPEP Steering Committee input.

1. One or Multiple Models

Districts should be encouraged to select from a limited number of state approved teacher and principal evaluation models. However, it is recommended that a state approval process be developed for those districts not wanting to select one of the state approved models.

2. Evaluation System Components

The Superintendent will require that all systems have specified components that will include, but not be limited to:

a. A research-based instructional framework (teachers) or leadership framework (principals) that clearly describes practice and performance. The Superintendent will approve a limited number of frameworks based on the results of the pilot. Other frameworks, including modified research-based frameworks, will be subject to a case-by-case approval process. The framework must incorporate the state evaluation criteria and must clearly “map-back” to the state evaluation criteria;

b. The use of the state-adopted evaluation criteria and definitions for both teachers and principals. A definition for each of the criteria will be finalized at the conclusion of the pilots;

c. Rubrics, which are based on the instructional framework (teachers) or leadership frameworks (principals), that clearly define the continuum from unsatisfactory through effective teaching and leading practices;

d. A four-level rating system that describes performance along a continuum that indicates the extent to which the criteria has been met or exceeded. At the conclusion of the pilot, the title and definition for each of the levels will be determined by the Superintendent for statewide use. The Superintendent will also establish the delineation between “not satisfactory” and “satisfactory” performance in the four-level system;

e. Multiple measures for determining effective teacher and principal performance. Currently, the pilots are considering a number of tools, including observation, self-assessment, perception surveys, and student growth. At the conclusion of the pilots, the effectiveness and implementation challenges of these tools will be evaluated and minimum requirements will be established;

f. Professional development for principals and classroom teachers regarding the content of the new evaluation systems. The professional development must include information regarding the instructional framework, evaluation criteria, scoring rubrics, and the tools that will be used to measure performance; and

g. Evaluation training for all evaluators (e.g., principals, district administrators) involved in the new evaluation systems.

School districts must be able to demonstrate that teachers, principals, parents, and others were involved in the decision-making process for the new evaluation system within the school district.

As discussed above, based on the outcomes of the TPEP pilots a final set of recommendations with specific approval criteria will be included in the report completed by July 1, 2012.

3. State-Approval Process

At the conclusion of the pilots, the Superintendent shall finalize the components and requirements that must be included in evaluation systems. Districts will be required to include all of the required components as specified by the Superintendent.

During the 2012-13 school year, school districts should be required to submit a description of their proposed evaluation systems that they intend to use beginning in the 2013-14 school year. The description of the system shall include how they will address each of the required components, which will be subject to a thorough, rigorous state review process.

If the system includes one of limited number of frameworks and meets specified criteria for the other components, it shall be approved. If other frameworks are used, the framework shall be evaluated on a case-by-case basis.

The initial review will be conducted by OSPI staff, who will make recommendations to a Review Panel consisting of representatives of teachers, principals, parents, school directors, and school district administrators.

4. Challenges to a State Evaluation Approval Process

The challenges to a state review process rest in two areas: time and resources. The capacity over the next two years to approve all models will be time consuming and require expertise at the state level to remain intact and enhanced through continued consultation with evaluation experts and practitioners. If the system is to be functioning at a high level during the 2013-14 statewide implementation year, serious consideration will need to be given to providing the resources to prepare all the districts in an intentional way for the new teacher and principal evaluation system.

VI. Conclusion

As with other states overhauling their evaluation systems across the country, there is a deliberate sense of urgency within the TPEP project. This urgency is tempered by the commitment to also get this right for students and not to rush to quick fixes that will not have long lasting impact. There are two important aspects of our project that should be noted as standing out among the deluge of state action around educator evaluation.
First, Washington State is taking the bold and important step in redesigning, piloting and implementing both the teacher and principal evaluation systems at the same time. Washington is one of only 12 states that has legislation requiring evaluation reforms. Although this has more than doubled the workload in the TPEP pilot sites, it has been a consistent message all year by everyone involved that the two are inextricably linked and must be implemented together.

Second, the partnerships and collaboration around this work is unprecedented. The coalition that has formed around the TPEP work from the state to the local level has made a profound impact on the hope and promise that the new evaluation systems will be a meaningful and intentional professional growth and accountability model for years to come. As with all new reforms, the new teacher and principal evaluation systems are going to face challenges. The ultimate goal of the TPEP project is to improving teaching and leading for all students in Washington State.

VII. Bibliography

Advocating for teacher quality in all 50 states*. Work in Progress*, The Joyce Foundation (2010). Retrieved June 24, 2011 from http://www.joycefdn.org/content.cfm/advocating-for-teacher-quality-in-all-50-states

Barber, M., & Mourshed, M. (2007). *How the world's best-performing school systems come out on top.* New York: McKinsey and Company. Retrieved September 15, 2009, from http://www.mckinsey.com/locations/UK_Ireland/~/media/Reports/UKI/Education_report .ashx

Brandt, C., Thomas, J., & Burke, M. (2008). *State policies on teacher evaluation practices in the Midwest Region.* Issues & Answers Report, REL 2008-No. 004). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved September 15, 2009, from http://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/techbrief/tr_00408.pdf

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching (2nd edition).* Alexandria, VA: Association for Supervision and Curriculum Development. ISBN: 978-1416605171

DuBois, C.C., & Brandt, C. (2011). *Washington teacher principal evaluation pilot cross-case analysis.* Washington, D.C.: American Institutes for Research.

Evaluating principal leadership in a performance-based school: The eight new principal evaluation criteria. Olympia, WA: Association of Washington School Principals (2010). Retrieved June 23, 2011, from

http://www.awsp.org/Content/awsp/PrincipalsHandbook/PrincipalEvaluationCriteria/Principal_Leadership_1.31.11_web.pdf

Fetters, J., & Behrstock-Sherratt, E. (2011). *Washington's teacher/principal evaluation pilot (TPEP) educator forums: Final report.* Washington, D.C.: American Institutes for Research.

Five dimensions of teaching and learning (5d). Seattle, WA: Center for Educational Leadership. Retrieved June 23, 2011 from http://www.k-12leadership.org/services/5-dimensions

Goe, L., Holdheide, L., & Miller, T. (2011). *A practical guide to designing comprehensive teacher evaluation systems: A tool to assist in the development of teacher evaluation systems.* Washington, D.C.: National Comprehensive Center for Teacher Quality. Retrieved June 23, 2011, from http://www.tqsource.org/publications/practicalGuideEvalSystems.pdf

Leithwood, K., Louis, K., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning (review of research)*. New York: The Wallace Foundation. Retrieved September 15, 2009, from http://www.wallacefoundation.org/SiteCollectionDocuments/WF/Knowledge%20Center/Attachments/PDF/ReviewofResearch-LearningFromLeadership.pdf

Marzano, R. (2011). *The art and science of teaching causal teacher evaluation model.* Blairsville, PA: Learning Sciences International.

Reeves, D. (2008). *Assessing educational leaders: Evaluating performance for improved individual and organizational results.* Cheltenham VIC, Australia: Hawker Brownlow Education. Retrieved June 23, 2011 from http://www.hbe.com.au/PUBLIC/HBEItemImages/pdf/CO8273.pdf

State Legislation: Emerging trends reflected in the state phase 1 Race to The Top applications. Naperville, IL: Learning Point Associates (2010). Retrieved June 24, 2011 from http://www.learningpt.org/pdfs/RttT_State_Legislation.pdf

Washington state school district teacher and principal evaluation system survey results. Olympia, WA: Office of Superintendent of Public Instruction (2011). Retrieved June 23, 2011 from http://www.k12.wa.us/Communications/StimulusPackage/SEES_Summary_PDF_SY2009-10.pdf

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on difference in teacher effectiveness.* Brooklyn, NY: http://widgeteffect.org/downloads/TheWidgetEfect.pdf

VIII. Appendices

Appendix A

- Teacher Evaluation Model Visual Diagram

Appendix B

- Principal Evaluation Model Visual Diagram

Appendix C

- Washington TPEP Cross-Case Analysis:

Appendix D

- Anacortes Case Study & Evaluation Models
- Central Valley Case Study & Evaluation Models
- Kennewick Case Study & Evaluation Models
- NEWESD 101 Consortium Case Study & Evaluation Models
- North Mason Case Study& Evaluation Models
- North Thurston Case Study & Evaluation Models
- Othello Case Study & Evaluation Models
- Snohomish Case Study & Evaluation Models
- Wenatchee Case Study & Evaluation Models

Appendix E

- Washington TPEP Statewide Data Analysis Report:

Appendix F

- Practitioner Workgroups Proposal April 2011
- TPEP Model Review Templates
- TPEP Practitioner Panel Participants
- TPEP Practitioner Panels System & State Feedback

Appendix G

- OSPI School Employee Evaluation Survey PDF Summary
- OSPI School Employee Evaluation Survey Data File

Appendix H

- TPEP Pilot Site Interview Compilation Video

APPENDIX A

Teacher Evaluation Model Visual Diagram

RCW 28A.405.100 Teacher Evaluation Criteria

1 2 3 4 5 6 7 8

Defined in RCW

Defined in WAC — not determined until conclusion of the pilot

Defined by in draft by TPEP pilots

Proposed Satisfactory/ Not satisfactory line

* Required by RCW

Criteria Definitions
are developed to create a common definition and understanding for each criterion.

Rubrics
are linked to each criterion and its descriptors and developed using the district's comprehensive instructional framework. Each criterion is scored by it's own set of measures and evidence. Each measure or piece of evidence may carry differing weight in determining the final score for each criterion.

Comprehensive Instructional Frameworks
are district's model of instruction and common language surrounding their instructional practice, aligned to the 8 criteria in RCW. Rubrics developed using instructional frameworks linked back to clear criteria definitions.

Measures & Evidence
To be defined.

Observation*

Other Measures & Evidence

Ongoing Analysis & Discussion of Measures & Evidence (Teacher & Principal)

Self Assessment

Student Growth Data

"When student growth data, if available and relevant to the teacher and subject matter, is referenced in the evaluation process it must be based on multiple measures that can include classroom-based, school-based, district-based, and state-based tools. As used in this subsection, "student growth" means the change in student achievement between two points in time."

"The four-level rating system used to evaluate the certificated classroom teacher must describe performance along a continuum that indicates the extent to which the criteria have been met or exceeded."

This line indicates a probable cause for non-renewal. The superintendent will also establish the delineation between "not satisfactory" and "satisfactory" performance in the four-level system.

Final Summative Evaluation

Level 1 | Level 2 | Level 3 | Level 4

DRAFT

APPENDIX B

Principal Evaluation Model Visual Diagram

RCW 28A.405.100 Principal Evaluation Criteria

1 2 3 4 5 6 7 8

Defined in RCW

Defined in WAC — not determined until conclusion of the pilot

Defined by in draft by TPEP pilots

Proposed Satisfactory/ Not satisfactory line

Criteria Definitions
are developed to create a common definition and understanding for each criterion.

Rubrics
are linked to each criterion and its descriptors. Each criterion is scored by it's own set of measures and evidence. Each measure or piece of evidence may carry differing weight in determining the final score for each criterion.

Leadership Framework
In addition to using "Evaluating Principals in a Performance-Based School," districts also used standards from the national organizations for principals, National Board for Professional Teaching Standards, Val-Ed, and the work of Dr. Marzano and Doug Reeves.

Measures & Evidence
To be defined.

Observation

Other Measures & Evidence

Ongoing Analysis & Discussion of Measures & Evidence (Principals & District Administrators)

Perception Data

Student Achievement Growth Data

"When available, student growth data that is referenced in the evaluation process must be based on multiple measures that can include classroom-based, school-based, district-based, and state-based tools. As used in this subsection, "student growth" means the change in student achievement between two points in time."

"The four-level rating system used to evaluate the principal must describe performance along a continuum that indicates the extent to which the criteria have been met or exceeded."

The superintendent will also establish the delineation between "not satisfactory" and "satisfactory" performance in the four-level system.

Final Summative Evaluation

Level 1 | Level 2 | Level 3 | Level 4

DRAFT

APPENDIX C

Washington TPEP Cross-Case Analysis

AIR®

AMERICAN INSTITUTES FOR RESEARCH®

Education,
Human Development,
and the Workforce

Washington Teacher Principal Evaluation Pilot Cross-Case Analysis

CC DuBois, Research Associate

Christopher Brandt, Senior Researcher

May 2011; Revised June 21, 2011

Contents

Page

District Overview and Description 1

Methodology 3

Interviews 3

Focus Groups 3

Document Review 4

Teacher Evaluation 5

Setting the Evaluation Framework and Goals 5

Establishing Evaluation Criteria, Identifying Measures, and Articulating Performance Levels 7

Measuring Performance 9

Using Results for Professional Development 12

Communicating Results 13

Pilot Timeline and Participation 13

Principal Evaluation 14

Setting the Evaluation Framework and Goals 14

Establishing Evaluation Criteria, Identifying Measures, and Articulating Performance Levels 15

Measuring Performance 16

Using Results for Professional Development 17

Communicating Results 17

Pilot Timeline and Participation 18

Summary and Conclusions 19

Appendix A. Sample Elements 21

Appendix B. Interview Protocol 23

Appendix C. Focus Group Protocol 26

District Overview and Description

Since the summer of 2010, eight school districts and a district consortium are participating in the Washington Teacher Principal Evaluation Pilot (TPEP). Each district and the consortium are working with the Washington Office of Superintendent of Public Instruction (OSPI) to develop nine new and innovative teacher and principal evaluation systems that comply with new state legislation, Senate Bill 6696, and lead to better teaching and learning. The 16 participating districts are diverse in their geography, size, and student composition. Some relevant demographic characteristics for the 9 pilot sites are shown in Table 1.[1]

Table 1. Demographics of Pilot Sites Participating in TPEP

Pilot Site	Approx. No. of Students	No. of Schools	American Indian or Alaskan Native Student Percentage	Asian or Pacific Islander Student Percentage	African American Student Percentage	Hispanic Student Percentage	Percentage Eligible for Free or Reduced-Price Lunch
Anacortes	2,737	8	2.1	4.1	1.6	5.9	35.3
Central Valley	12,398	22	1.3	2.0	1.2	3.3	38.4
Kennewick	15,969	25	0.9	2.5	2.2	29.7	48.1
North Mason	2,222	6	3.0	1.7	1.3	7.0	39.5
North Thurston	13,952	21	2.8	13.0	9.2	11.3	38.6
Othello	3,690	7	0.2	0.6	0.4	82.6	79.1
Snohomish	9,961	17	1.0	5.2	1.7	5.8	19.7
Wenatchee	7,712	14	0.5	1.3	0.5	43.2	54.9
Northeast Washington Educational Service District Consortium							
Almira	84	1	0.0	0.0	2.4	2.4	42.4
Davenport	595	2	1.2	0.7	0.3	2.4	54.7
Liberty	476	2	2.5	1.1	0.6	2.7	44.0
Medical Lake	2,100	6	0.6	4.7	4.6	6.1	31.5
Pullman	2,398	5	1.7	10.5	3.7	4.6	25.9
Reardan-Edwall	673	2	4.3	0.5	0.5	0.7	38.7
Ritzville	342	2	2.6	1.8	0.0	5.6	41.1
Wilbur	252	2	16.3	2.4	3.6	1.6	40.9

[1] Information for Table 1 was found at the Washington State Report Card, http://reportcard.ospi.k12.wa.us/DataDownload.aspx.

The eight independent pilot districts serve between 2,200 and 16,000 students and operate between 8 and 22 schools. The percentage of minority students served within these districts ranges widely; from 7.8 percent in Central Valley to 83.8 percent in Othello. In nearly all districts, the minority population consists primarily of Hispanic students, the exception being North Thurston. The percentage of Hispanic students across the pilot districts ranges from 5.8 percent in Snohomish to 82.6 percent in Othello. The percentage of students who qualify for free or reduced-price lunch ranges from 19.7 percent in Snohomish to 79.1 percent in Othello. The Northeast Washington Educational Service District (NEWESD) consortium is made up of eight rural districts that are pooling their collective resources to develop a common evaluation system that will be available to every district in the consortium. These districts operate between 1 and 6 schools and serve between 84 and 2,398 students.

Methodology

The Office of the Superintendent of Public Instruction in Washington (OSPI) partnered with the American Institutes for Research (AIR) on a project to understand the current state of evaluation in Washington, document pilot sites' development process, examine pilot sites' experiences and perspectives during development, and provide formative feedback to help pilot sites implement high-quality evaluation systems. The objective of this work is to learn from the pilot sites so that OSPI can determine how to best support districts' implementation of Senate Bill 6696 statewide, which establishes new criteria for evaluating educators.

The purpose of this report is to document pilot sites' progress in the development of their teacher and principal evaluation systems. The report relies on three data sources: (1) interviews, (2) focus groups, and (3) a document review.

Interviews

All of the pilot sites developed and designed their new educator evaluation systems with a committee of educators and other stakeholders. AIR interviewed four actively involved evaluation committee members from each of eight pilot districts, but not the consortium. Due to the unique nature of the consortium, AIR instead conducted three focus groups (see next section below). In an attempt to gain a diverse range of institutional perspectives, AIR interviewed one district-level administrator, one school-level administrator, and two teachers in each district. The interviews were recorded and transcribed. Transcripts were then inductively coded to capture the overall themes with respect to (1) how the pilot districts' performance evaluation systems were being developed and (2) participants' experiences with the process. The interviews were conducted in January and February 2011. Appendix B contains the protocol used for the interviews.

Focus Groups

AIR conducted three focus groups with consortium superintendents, principals, and teachers, respectively, who were actively involved in developing the consortium's new teacher and principal evaluation systems. Each focus group lasted 1–1.5 hours, and all groups were recorded and transcribed. Transcripts were then inductively coded to capture the overall themes with respect to (1) how the consortium's performance evaluation system was being developed and (2) participants' experience with the process. The focus groups were conducted in February 2011. Appendix C contains the protocol used for the focus groups.

Document Review

With explicit guidelines for what to submit and how to submit it, each pilot site was asked to submit to AIR draft documents used in the design and development of their teacher and principal evaluation systems to receive formative feedback. The documents were reviewed by AIR staff members, who are considered to be experts in the areas of teacher and principal evaluation. Specific checklists and processes were used to review the submitted documents to determine

documented evidence of each pilot's actions on the seven key components of a comprehensive educator evaluation system:

1. Communication and Stakeholder Investment
2. Evaluation Format and Multiple Measures
3. Instructional Framework/Leadership Framework and Strength of Measures
4. Selecting and Training Evaluators
5. Frequency of Evaluation
6. Pilot Implementation Tracking and Reporting
7. Data Uses and Tracking

The pilots were given written feedback about their teacher evaluation system and their principal evaluation system, as well as 60-minute individual consultations with an AIR staff member. The purpose of the consultation discussion was to clarify questions raised by the document review results and offer formative feedback to all of the pilots on the design of one or more areas of their developing system. Pilots had the opportunity to submit supplementary documents following the consultation meeting. The document review process occurred in March 2011.

Once all data collection was complete, AIR developed individual case studies for each pilot site. The case studies were then coded and inductively analyzed to create this cross-case analysis. Throughout this report, the district consortium is treated as one individual pilot site, making the number of pilot sites equal to nine. The findings that follow synthesize pilot sites' progress between the start of the pilot in summer 2010 and the end of data collection in March 2011. Pilot sites have continued to develop their teacher and principal evaluation systems beyond the data collection period. **This report, then, serves as a snapshot of progress between July 2010 and March 2011 for the 9 pilot sites as their work continues to unfold and develop as they prepare for full pilot implementation in the Fall of 2011.**

This report will provide information, first for the teacher evaluation systems and then for the principal evaluation systems, on the progress pilot sites have made in the following areas:

- identifying an instructional framework;
- articulating goals;
- establishing evaluative criteria;
- identifying measures;
- articulating performance levels;
- measuring performance;
- using results of the evaluation system for professional development;
- communicating results;
- developing the pilot timeline; and
- identifying pilot participants.

The conclusion summarizes thoughts interviewees shared on the strengths of their new teacher and principal evaluation systems as well as the challenges of development and implementation.

Teacher Evaluation

Setting the Evaluation Framework and Goals

According to interviews, six pilot sites had an instructional framework in place to guide teachers before the pilot began, and three sites are developing the framework this year. Although only two of the six sites submitted a framework for the document review, interviewees within these six sites consistently described a similar framework in use within their districts. The following instructional frameworks were being used in these six pilot sites prior to the start of TPEP. (The number of districts that adopted each of these frameworks is given in parentheses.)

Instructional Frameworks Utilized Within Pilot Sites

- An internally developed framework grounded in the Effective Schools research (1)
- Quality Teaching and Learning (QTL) framework (1)
- An amended versions of the Charlotte Danielson model (2)
- The Five Dimensions of Teaching and Learning, developed by the University of Washington Center for Educational Leadership (1)
- Purpose, Engagement, and Results (PERR), developed in consultation with the Harvard Institute for School Leadership (1)

Sites currently developing instructional frameworks are basing them on the following models:

- Robert Marzano's Classroom Instruction That Works (1)
- BERC Group's Powerful Teaching and Learning (1)
- Charlotte Danielson (3)

One site is using both the Marzano and the Danielson models to inform the development of their instructional framework, and one site is adopting elements of the BERC Group and Danielson models.

In addition, interviewees from six pilots stated that they had aligned their instructional framework to their new teacher evaluation rubrics, although a document review was able to confirm this in only one site. Two of the other three sites indicated that they were currently working on aligning their framework to their evaluation rubrics or planned to begin the process soon. Similarly, in the last site, which was the consortium, the collaborating districts plan to begin linking their new evaluation rubrics to a broader district or school instructional framework once the core components of a new evaluation system is complete. Although their instructional framework has been established, the consortium is still in the process of refining the details of the teacher rubric.

Evaluation System Goals

When pilot site representatives were asked about the main goals of their teacher evaluation systems, several themes emerged.

- **Using evaluations to inform professional development—**Five interviewees discussed the need for more meaningful and collaborative conversations between school administrators and teachers to support ongoing instructional improvement. Specifically, six pilots indicated that administrators need to understand how evaluation data can be used to (1) target specific areas for growth, (2) identify appropriate next steps for improvement, and (3) align specific professional development opportunities with individual and schoolwide goals. One pilot stated that more intensive support should be reserved for teachers whose performance is less than satisfactory to ensure they receive the support they need to improve their practice. One pilot stressed the need to provide appropriate supports at all levels so that teachers will be encouraged to not settle for being good but to strive for being great by working on specific areas of need. Twenty-nine respondents across the pilot sites discussed the importance of the new system being a growth model that improves instructional practice and fosters professional development for educators.
- **Ensuring that evaluations produce credible and trustworthy results—**Four interviewees reiterated the need for an evaluation system that ensures quality by gathering richer and more diverse information on teachers' practices and establishing a consistent teacher evaluation process and set of expectations across schools. Five respondents expressed the importance of having a teacher evaluation system with clear, understandable, and transparent criteria and rubrics for all faculty and administrators.
- **Creating an overall framework to guide the evaluation process—**Three interviewees were hopeful that a new, well-articulated evaluation framework would establish a common definition of quality teaching and learning. In addition, the pilots want the new system to drive improvement in professional development, support, and effective use of time, thereby ensuring that a proficient teacher is in every classroom and that students are learning. In line with district strategic planning, three interviewees insist that the new teacher evaluation system should help teachers be more purposeful in their instruction, providing them with tools, developing clear instructional standards, and clarifying how teachers can be promoted up a tiered model of effectiveness.
- **Maintaining a focus on teaching and learning—**Interviewees suggested that, at the core, the main goal of the new teacher evaluation systems is to improve teaching and learning. By promoting professional growth and improving how teachers deliver instruction, two interviewees are hopeful that student performance will improve.

Using Evaluation Results to Place and Incentivize Teachers

Three pilot sites suggested that teacher evaluation results could be used to inform transfer assignments, continuing contract decisions, and teacher promotion or advancement. One respondent felt that the teacher evaluation results would be a good indicator of a teacher's readiness for promotion or advancement. Another respondent thought that evaluation results

might eventually influence whether a teacher should be counseled out of the profession, if it is the case that the teacher consistently receives unsatisfactory evaluations. Four pilot sites related that their teacher evaluation systems will not be linked to compensation decisions; two sites have yet to discuss these issues; and three sites provided no information on a link between evaluations and compensation. The document review supported these findings, in that no evidence was found that any pilot site would use evaluation results in compensation decisions. One respondent mentioned that, in his district, compensation decisions are currently dictated by union contracts, so even if the district decided to align the pay scale or promotions with the evaluation results, it would have to bargain with the local union to make changes to contract language.

Establishing Evaluation Criteria, Identifying Measures, and Articulating Performance Levels

Senate Bill 6696 established eight criteria to inform the development of new district-level teacher evaluation systems. Most pilot sites initiated the development process by attempting to break down each criterion into a set of observable and measurable elements. After creating the elements, districts plan (or began) to identify measures and instruments that can be mapped to the elements and used to evaluate teachers on each state criterion. In addition, districts must adopt a four-tiered rating system based on the state's eight criteria and are in the process of producing rubrics that describe what performance looks like at each level of effectiveness. A document review confirmed that eight pilot sites had drafted elements to support the state's new criteria, along with rubrics to describe levels of teaching effectiveness. Of these eight sites, four contained measures and instruments for assessing at least one or more of the new evaluation elements and criteria. Appendix A displays draft elements from two of these eight pilot sites.

Throughout the planning process, pilot sites and individual committee members often developed categories to frame their own thinking about what a new teacher evaluation system should measure in light of the new legislation. For instance, an interviewee from one pilot site pointed out categories that the interviewee considered when developing rubrics and describing levels of performance: data use by teachers, clearly articulated and understood teaching standards, and districtwide collaboration to share best practices. Similarly, one pilot site collectively developed three categories—preparation, teaching, self-reflection—and used them to integrate the state's new criteria into their existing instructional framework.

Identifying Who Will Be Evaluated

Although still in development at the time of the case study, pilot sites were in the process of determining who will be evaluated under the new system, interviewees from seven pilot sites suggest that districts will do one of five things. (The number of pilot sites employing this design is given in parentheses.)

- Case One: Apply the new teacher evaluation rubric to all general education teachers, some special education teachers, and possibly specialists (e.g., counselors, speech and language pathologists) (1)
- Case Two: Apply the newly designed rubric to all teachers with different tools being developed for specialists (2)

- Case Three: Apply the new rubric to all classroom teachers but not to specialists (2)
- Case Four: Apply the newly designed rubric to all general education teachers and develop rubrics for all other staff (1)
- Case Five: Apply the newly designed rubric to a percentage of certified staff (1)

Determining the Frequency of Evaluation

None of the pilot sites has determined the number of evaluations annually for provisional and continuing contract teachers. Several sites pointed out that, currently, evaluation frequency is codified in state law (E2SSB6696, Part II, §202(3)(a)). According to this law, provisional employees receive more frequent observations than continuing contract teachers. Most sites indicated that under the new system they intended to observe provisional teachers more often that continuing contract teachers, but they had not settled on a specific number of observations. Two respondents from one site referred to the district's long-form and short-form options, and one respondent expressed hope that the district would continue to have both options. A couple of sites pointed out that they will need to negotiate language on frequency of observations with their local union.

The document review revealed that one pilot site may be planning to place teachers into one of four tracks, with the number of annual observations differing by track.[2] Specifically, provisional contract teachers (track 1) and newly continuing contract teachers (track 2) will be observed twice a year; teachers on a professional growth track (track 3) will be observed once during the year, but observation will be supplemented by meetings with an administrator to check on professional growth; teachers on track 4, the assistance phase, will be provided significant amounts of help to improve their practice.

Developing a Tiered System

While each pilot site was required by the 6606 legislation to have a four-tiered teacher evaluation system, the labels applied to each level vary by site. Table 2 shows the labels for each district's new four-tiered system. The majority of results were informed by interviews, and four sites provided documentation to corroborate interview results, as well as were further verified by information submitted to OSPI in June 2011.

Table 2. Labels Applied to Each Pilot Site's Four-Tiered Evaluation System

Pilot site	Level 1	Level 2	Level 3	Level 4
1	Unsatisfactory	Emerging	Proficient	Exemplary
2	Not demonstrated/ Unsatisfactory	Developing	Proficient	Accomplished
3	Unsatisfactory	Emerging	Proficient	Exemplary
4	Unsatisfactory	Basic/Emerging	Proficient	Distinguished

[2] The document from which this information was pulled was in draft form at the time of analysis and is subject to change.

Pilot site	Level 1	Level 2	Level 3	Level 4
5	Unsatisfactory	Basic	Proficient	Distinguished
6	Unsatisfactory	Basic	Proficient	Innovative
7	Unsatisfactory	Emerging	Proficient	Exemplary
8	Unsatisfactory	Basic	Proficient	Distinguished
9	Unsatisfactory	Basic	Proficient	Distinguished

Differential Ratings

Six pilot sites provided information on differential ratings, according to which teachers are rated in comparison to each other. Five sites reported that they will not use differential ratings. One site is still discussing the issue and has not ruled out using differential ratings on some measures.

Identifying Proficiency Cutoffs

Interviews revealed that five pilot sites have at least tentatively defined a proficiency cutline for their new teacher evaluation systems, and the document review corroborated this finding in two of the five sites. Although the establishment of a cutline suggests disciplinary action or additional supports for those who fall below the cutline, districts are only beginning to plan what actions will be taken. Four sites reported that the cutline would be between levels one and two, and one site reported that the cutline would be between levels two and three. The pilot site that placed the cutline between levels two and three also related that for provisional teachers, or experienced teachers on new assignments, it would be acceptable to be a level two for a period of time to be specified later. The pilot site continues to have discussions about the meaning of this cutline and its implications for all staff members. One of the four pilot sites that have not yet determined a cutline is concerned about what the consequences will be for teachers who continue to receive lower ratings. For example, the pilot committee has not yet reached a consensus on how long a teacher would be allowed to remain at a rating of two without showing enough improvement to advance to a level three.

Measuring Performance

At the time of the interviews, all nine pilot sites were still in the process of developing measures for their teacher evaluation systems. A number of respondents, however, provided a sample of measures that the pilot committees had been discussing. One district is considering feedback meetings between teachers and principals, in which teachers discuss how they are using data from formal tests, classroom-based assessments, and parent logs to differentiate instruction. Table 3 presents the range of measures being considered among pilots.

Table 3. Measures for Teacher Evaluation Under Consideration

Pilot Sites	Observe/ Drop in	Portfolio	Lesson Plans	Prof. Growth Plans	Student Data	Parent Surveys	Student Surveys	Student Artifacts	Peer Evaluation	Feedback Meetings	Self-Assessment
1	X	X	X	X	X	X	X		X		X
2	X		X		X			X		X	
3	X	X			X	X	X				X
4	X	X									X
5	Respondents did not indicate which measures are currently under consideration.										
6	Respondents did not indicate which measures are under consideration; the pilot committee did, however, submit a draft document that specifies a list of potential quantifiable measures.										
7	X		X			X					
8	X		X								
9	X	X									

Selecting Multiple Measures

The seven pilot sites that have begun identifying evaluation measures are planning to conduct formal and informal classroom observations. Interviewees from four sites identified portfolio and lesson plan review as possible measures, three are considering the use of self-assessment or parent surveys, and two are considering the use of student surveys. Other measures mentioned during the interviews were peer evaluation, ongoing feedback meetings between teachers and their principals, student artifacts, and professional growth plans. Of particular note, when asked about the use of multiple measures, broadly, interviewees from only three sites out of seven indicated that they are considering student data as a possible measure for evaluating teachers. However, when specifically asked, eight sites commented on using some form of student growth measures to evaluate teachers.

Four pilot sites submitted a list of multiple measures for the document review. One pilot site has developed a draft document that specifies a list of possible measures. The list, "meant to be a natural harvest of teacher work," emphasizes artifacts of teaching, such as student work and assessment data. Under the district's current plan, teachers will undergo mandatory observations and turn in work samples, assessment data, and other evidence sources to be used in their evaluations. Teachers will choose from a limited selection of acceptable measures to demonstrate that they have met the criteria. Respondents concurred that all teachers, core instructional area teachers and noncore instructional areas alike, would use the same list of measures.

Concerns About the Use of Certain Measures

As pilot districts consider the use of multiple measures for teacher evaluation, concerns have emerged about the use of certain measures. For instance, interviewees from two pilot sites were hesitant to require teachers to create portfolios. These respondents mentioned that portfolios might make the evaluation process too time-consuming and cumbersome for teachers and administrators. In addition, respondents from one pilot site mentioned that their pilot committee is struggling to find an objective measure to assess teachers' collaboration with peers.

During the interviews, respondents from four pilot sites provided information on whether the same measures will be used for teachers in core and noncore subjects. Respondents from three sites related that the measures would be the same, but respondents from one site are still undecided, although many respondents from the site hoped that the measures would be applicable across levels and across the curriculum.

In sum, respondents expressed concerns about the feasibility of using certain measures, the time burden some measures place on teachers and principals, as well as the objectivity inherent in measures for certain criteria.

Using Student Growth Measures

Although five sites have reportedly decided that teachers will be required to use student data to differentiate instruction, only one pilot site provided written guidance on how student data will be incorporated into teachers' evaluations. When specifically asked, eight sites commented on

using student growth measures to evaluate teachers, but no final decisions on what measures to use have been made.

There are signs of internal disagreement over specific growth measures to be considered in the evaluation and the weight to be applied to these measures within the overall evaluation rating for teachers. For instance, a respondent from one site felt that state benchmark assessments (e.g., Measures of Academic Progress testing) and other state exams should be used to inform the teacher evaluation system. Another respondent within the same site was adamant that student test score data be used only to inform teacher planning, not as a measure of teacher success. A third respondent at the same site agreed, positing that although test scores should be used to inform instruction, using student test scores to evaluate teachers is taboo. Within a second pilot site, one respondent expressed a desire to see more use of common assessment data to evaluate student improvement; the respondent was wary, however, of placing too much emphasis on a specific test. At a third site, respondents reported that teachers will be held accountable for using data to differentiate their instruction, but they would not be rated by their students' growth. Similarly, six sites have either decided against the incorporation of value-added scores as the indicator of student growth or are extremely hesitant to consider it.

Training for Evaluators

None of the pilot sites had yet finalized plans for evaluator training, but all were in the process of developing those plans. All pilot sites, however, expressed the importance of providing robust training for evaluators. Of the three pilot sites that have an evaluator training strategy in place, one will conduct training during their annual Administrative Leadership Academy in the summer; one has already conducted two rounds of evaluator training for all evaluators in the district; and the other has contracted McREL to provide training.

Respondents from all nine sites mentioned the importance of ensuring interrater reliability. Several sites mentioned the use of videotaped lessons as a potential training exercise to measure interrater reliability. One site plans to develop a glossary of terms to ensure that principals and teachers are communicating effectively about the teacher evaluation system. Another site proposed to integrate ongoing evaluator training into administrators' regular monthly professional growth meetings. Respondents from a third site mentioned plans to provide evaluator training for principals and teachers as part of their annual instructional conferences and academies.

Using Results for Professional Development

Respondents from all nine pilot sites intend to develop procedures for using evaluation results to inform professional development. No sites have begun devising these procedures, however. Respondents from four pilot sites mentioned that the main impetus driving professional development will probably be one-on-one discussions between the teacher and the principal. Respondents from one pilot posited that, because teacher ratings will be provided on each of the eight criteria, the new system will provide greater clarity on strengths and weaknesses, allowing for more targeted professional development. Respondents from the same pilot discussed how trends in evaluation results would also guide buildingwide or districtwide professional

development activities. Two respondents from another site envisioned that common themes will emerge from the goal-setting that takes place as part of the new evaluation system, and that various opportunities for collaborative professional development will be based on these trends. Respondents from several sites expressed concern over professional development budget cuts and the impact these cuts may have on the quality and depth of professional development the district will be able to provide teachers.

Although no documented information was provided from the pilot sites about plans for training in data use, interviewees from eight pilot sites indicated that they plan to provide data use training for teachers and evaluators. Seven sites already provide training for teachers on the use of data to differentiate instruction, and respondents in these sites were optimistic that these training sessions would continue. One site is currently developing a data dashboard and will conduct training sessions once that tool is complete.

Communicating Results

During the interviews, respondents from eight pilot sites discussed their strategy for communicating the pilot committees' progress and plans to teachers, principals, and relevant stakeholders. Seven of these sites already had a communication strategy in place, and one site is currently developing a strategy. Seven of these eight pilot sites have a component for ensuring parent/community feedback. From the interviews, it was evident that a majority of the pilot sites have been very diligent about communicating the results of their work to teachers, principals, and community stakeholders. Many respondents considered this crucial for transparency and creating buy-in.

Five pilot sites submitted information on their communication strategy for the document review. These documents provided evidence that a number of robust communication strategies are in place; how stakeholder feedback will be incorporated into the development process is still an ongoing task of the pilot sites.

Pilot Timeline and Participation

Respondents from all nine pilot sites stated that they plan to follow the state's timeline for piloting and implementation. That is, the pilot sites are using the current academic year to develop their teacher evaluation systems. They will pilot this system in schools during the 2011–12 academic year. Several sites plan to roll out the pilot in phases. For instance, two pilot sites began holding prepilots during the spring 2011 semester, during which volunteer teachers were evaluated with the new rubrics in an effort to test the rubric and gather feedback. Three sites plan to select a sample of schools and principals to be involved in the pilot during 2011–12 with a full rollout in 2012–13.

To gather enough teachers for the pilot, two sites plan to ask teachers to volunteer, and three sites will randomly select teachers to participate. Respondents from three sites discussed strategies for incorporating feedback from the pilot year, and respondents from two sites mentioned the possibility of amending their rubrics to accommodate other school staff (e.g., librarians, speech pathologists, counselors) after the pilot year.

Principal Evaluation

During the period of data collection for the case studies, the majority of pilot sites had more fully developed their teacher evaluation systems and had only just begun discussions of the principal evaluation system.

Setting the Evaluation Framework and Goals

Six pilot sites provided information during the interviews on the district's leadership framework. Of the six sites, three had a leadership framework in place prior to TPEP, and the other three sites are currently developing a leadership framework to aid in their development of the principal evaluation system. The document review was able to confirm either an existing framework or a framework process in development in three instances. Of the three sites with leadership frameworks in place, one is based on the Association of Washington School Principals model and the other two are guided by Robert Marzano's 21 responsibilities of the school leader. One of the districts using the Marzano model also has relied on resources from the Association for Washington School Principals (AWSP) to develop their evaluation rubric. The three sites that are currently developing a leadership framework are all basing their systems on the resources provided by AWSP. One respondent explained that the AWSP has had a committee that has been proactively working on the design of a model for principal effectiveness for several years.

During the interviews, respondents from two pilots mentioned that their leadership frameworks will be closely aligned with their principal evaluation systems, and the document review found evidence that a third pilot site is also aligning its leadership framework with its evaluation system.

Evaluation System Goals

When respondents were asked about the main goals of the new principal evaluation system, several themes emerged:

- **Professional growth—**As with the teacher evaluation system, the most cited (six respondents) main goal of the principal evaluation system was professional growth. Four respondents hope that the new evaluation system will result in improvement in leadership practice, both instructionally and managerially; that principals will recognize that teacher growth goes hand in hand with principal professional development; and that visionary instructional leadership and self-reflection will be the norm. Another respondent mentioned that the ability to focus professional development efforts on unique individual needs is also a main goal.
- **Quality assurance**—Respondents from three pilot sites reiterated that a main goal of the new evaluation system is to promote consistency across schools by providing clear expectations about what is expected from an exemplary principal. A hope is that this will result in greater accountability and will provide an illustration of how principals are performing.

- **Instructional impact—**Respondents from two pilot sites believe that a main goal of the new principal evaluation system is to improve teaching and learning, thereby affecting classroom teaching and student achievement. One respondent mentioned the use of student data as an important component of the principal evaluation system.

Using Evaluation Results to Place and Incentivize Principals

A respondent from one pilot mentioned that evaluation results could be used to inform promotion and advancement; the respondent was very skeptical, however, about whether they would be used to inform compensation. Respondents from three other pilot sites mentioned that the evaluation committee had not yet discussed whether principal evaluations would be used for purposes other than professional growth, such as determining compensation or other human capital policies. The document review provided no further information on this topic.

Establishing Evaluation Criteria, Identifying Measures, and Articulating Performance Levels

During the interviews, respondents from eight pilot sites confirmed that they have developed or are in the process of developing elements for the state's eight mandated criteria. The document review was able to confirm that five pilot sites have developed rubrics that list elements of the state's criteria. Although none of the elements was shared during the interviews, a respondent from one pilot identified four themes: (1) to ensure that systems are in place to support teachers' assessment programs and provide instructional materials and professional development to assist teachers in differentiating their instruction; (2) proficiency in assessing teacher performance in the classroom; (3) coaching, mentoring, and developing teachers; and (4) accountability—ensuring that a standard of quality is maintained across the school. A respondent from another pilot mentioned several indicators of competency—establishing relationships with teachers, understanding student and teacher data, understanding good lesson design, managing resources, being a good communicator, assessing and monitoring programs, understanding the principal's role as an evaluator, and initiating professional growth dialogue with teachers.

As was mentioned previously, several pilots are using the AWSP model for guidance in the development of their principal evaluation rubrics. One respondent noted that the AWSP model includes a ninth criterion, in addition to the state's eight, on principal accountability for closing the achievement gap. There have been discussions among pilot committee membership on whether they should include this criterion as well. In addition to examining the AWSP standards, a respondent from one pilot site posited that they looked at best practice evaluation models from around the country. The committee used those examples to choose specific language for the rubrics.

Determining the Frequency of Evaluation

A number of pilot site interviewees reiterated that their pilot committee had not progressed in the development of their principal evaluation system enough to have determined the frequency of evaluations. The document review provided no additional information.

Developing a Tiered System

In compliance with the 6696 legislation, each pilot site has been developing a four-tiered principal evaluation system; labels applied to each level vary by site, however. Each pilot decided to use the same labels for their principal evaluation system as they used for their teacher evaluation system, which were given in Table 2 in the Teacher Evaluation section of this report.

Identifying Proficiency Cutoffs

The interviews and document review provided information on the cut line for only one pilot site. That district has determined that the cut line will fall between a level one and level two. No information was provided on what actions are taken when a principal receives a level one rating.

Measuring Performance

During the interviews, eight of the pilot sites addressed the question of performance measures. Respondents from the eight sites reported that the pilot committees have yet to make any final decisions about performance measures. The document review was able to confirm that all nine pilot sites had yet to make final decisions about measures. Several respondents provided examples of the types of measures the pilot committees were discussing: formative and summative assessment results, building-level student achievement data, artifacts, school failure rates, school safety and climate plans, and portfolios.

Using Student Growth Measures

When asked about the inclusion of student growth measures for principal evaluation, four pilot sites provided information during the interviews. Although their pilot committee had not yet determined whether test scores will be included in the principal evaluation system, a respondent from one pilot noted that student test score data can be used under their current evaluation system to meet the “knowledge of student progress” criterion. A respondent from another pilot noted that there is more direct language from the state about improving student test scores in the principal evaluation than in the teacher evaluations. The respondent believed that evaluators will want principals to demonstrate an awareness of student test scores and trends and how the principal is responding to these trends. The respondent noted, however, that this would not be a change from current practice. A respondent from yet another pilot pointed out that student growth measures are already an integral part of the principal evaluation system in his district. This evaluation measure even played a role in determining principal compensation as part of a recently revoked performance pay system in which 10 percent of a principal’s salary was based on goals for student achievement. Principals are still required to report their levels of academic attainment to the school board every year, and the evaluation process assesses whether goals for student growth were achieved, but the measure is no longer used to make compensation decisions. Two of the respondents suggested that the evaluation process for principals in the new system may not differ too markedly from the process that is already in place.

Training for Evaluators

During the interviews, all nine pilot sites commented on training for evaluators. Three sites have a training plan in place, but the document review was able to confirm only one of these. Two of these pilot sites will hold their training sessions over the summer in conjunction with training for the teacher evaluation system. As for the other pilot sites, principal evaluators have received ongoing training on the contents of the rubrics as they develop and change. According to one respondent, they are currently receiving training to ensure interrater reliability. The other six pilot sites are in the process of developing a plan for training evaluators; two sites mentioned that the training would occur over the summer.

Using Results for Professional Development

Respondents from seven pilot sites mentioned the importance of using evaluation results to inform professional development; only one district, however, has identified a mechanism to link the two. From their evaluation results, this district will place administrators in one of four cycles (professional, observation, plan of improvement, or probationary), which will dictate the types of professional development on which administrators should focus. A respondent from one district discussed the possibility of developing internal training or professional development units specific to each evaluation rubric. In this case, there would be specific resources within the district for a principal that is struggling with a particular criterion. A respondent from another district expects that principal evaluation results will inform professional development on both an individual and a district level. As with teachers, principals will have a meeting with their evaluator, who will identify areas of strength and weakness and ways to engage in effective professional development. These procedures have not yet been formally adopted, however. A respondent from another district mentioned that the new evaluation system is intended to be used primarily for principals' professional development. Principals currently receive monthly professional development, which is expected to continue under the new system. Another respondent from the district noted that the professional development under the new evaluation system will need to be tied to the district's 21 responsibilities for leaders. It is intended that professional development will be individualized to meet each principal's growth needs and then grouped to provide learning opportunities as needed for principals across the district.

The document review was able to provide no further information on principals' professional development.

Communicating Results

During the interviews, respondents from seven pilot sites provided information on the site's strategy to communicate the new principal evaluation system to relevant stakeholders. Five pilot sites stated that they would communicate the requirements of the principal evaluation system at the same time as the teacher evaluation system. Because the number of individuals affected is so small in the case of principals, it does not appear from the interviews that any formal communication strategy has been considered necessary in one pilot site. One pilot had developed and disseminated a slide presentation detailing the planned changes for their principal evaluation system and also has distributed feedback forms to principals and certified staff to gauge their

reactions and answer any questions that arise. This pilot site had not yet communicated the upcoming changes to parents and other community stakeholders but does plan to do so.

Pilot Timeline and Participation

During the interviews, respondents from four pilot sites commented on the pilot timeline. Three have already developed a timeline for implementation and one is currently developing a timeline. According to the document review and interviews, three pilot sites have indicated that the pilot of the principal evaluation system will begin in the 2011–12 academic year. One pilot site will require all principals in the district to participate, and a respondent from another pilot site noted that about 50 principals and assistant principals are participating in the pilot this spring, with all of them expected to be evaluated under the new model during the 2011–12 pilot year. McREL will then assist in collecting data to assess the principal evaluation pilot results. Respondents from another district related that both the superintendent and assistant superintendent have used the district's draft principal rubric in midyear evaluations during the current school year (2010–11).

Summary and Conclusions

When respondents were asked about the strengths of their new teacher evaluation systems, the following themes emerged.

- **Professional growth—**Ten respondents from eight pilot sites felt that the new teacher evaluation system would engage teachers and principals in valuable conversations on effective instruction, thereby allowing the principal to provide more robust feedback to teachers. One respondent mentioned that a great strength of the new system is that it will reflect not only what a teacher does, but what a teacher needs to do to progress. There is a hope that viewing the evaluation reforms as a systems approach will assist teachers and principals in receiving constructive criticism, defining professional goals, and focusing professional development efforts, resulting in continual growth. Also, being able to differentiate teaching beyond satisfactory and unsatisfactory so that both struggling and high-performing staff can be recognized and provided with professional growth assistance was perceived as a strength.
- **Clarity of language and expectations—**Seven respondents from four pilot sites felt that the new evaluation systems will provide a clearer and more specific language about teacher competencies and standards, which will allow teachers and principals to assess their performance levels. By clearly articulating what good instruction looks like, teachers can change their instructional practice for the better.
- **Focus on multiple measures—**One respondent noted that a great strength of the new evaluation system is that the rating system provides a continuum of growth and will incorporate multiple measures to inform teacher performance and professional development. Two respondents from different pilot sites noted that the new evaluation systems are much more deliberate about including student data elements and measuring whether teachers are differentiating instruction on the basis of that feedback. This has resulted in the new evaluation system focusing more on student growth.
- **Model development—**Two respondents from the same pilot site mentioned that the fact that their new evaluation system builds upon their previous system is a great strength. Two respondents from different pilot sites also pointed out that they have developed research-based models. A strength in the view of one respondent was that the system was locally developed by teachers and principals in the district with feedback from parents and community stakeholders. One respondent mentioned that the district and union are working together to revise the evaluation system, united by the desire to improve teacher and principal professional growth opportunities.

Respondents also were cognizant of the challenges facing them when designing and implementing the new evaluation systems. Several themes emerged when describing these challenges:

- **External mandates—**Several respondents from different pilot sites expressed concern that the state will eventually mandate the use of value-added measures and impose high-stakes consequences on schools and teachers on the basis of these measures. Similarly, one respondent was concerned that the state would use results from the TPEP pilot project to select and eventually mandate that all districts utilize one rubric for evaluating

teachers. Another respondent noted that the state seemed to be shifting directions from its original requirements and was worried that the work of the pilot sites would not end up being implemented.

- **Stakeholder buy-in—**Two respondents from different pilot sites feared that the new system could be met with anxiety and fear on the part of teachers because there can be issues of trust between school administrators and teachers. One respondent was concerned that the economic cutbacks might make staff wary that as professional supports are taken away (e.g., larger class size, less financial compensation, and fewer resources), they will be evaluated and expected to perform at higher levels. All four respondents from one district were apprehensive about the impending paradigm shift required to evolve from a satisfactory or unsatisfactory rating system to a four-tiered system in which less than 10 percent of teachers are expected to receive a rating of four.
- **Scope of implementation—**Professional development was identified as a challenge, because a large number of teachers and principals will need to be trained to make the best use of the new system. Several respondents from various pilot sites pointed out that principals will need to commit more time to ensuring that robust conversations fostered by the evaluation systems can occur. One respondent thinks it will be challenging to help teachers understand that it is their responsibility to provide evidence of professional growth and student achievement.
- **Resources—**Several respondents from various pilot sites felt that time and resources would be great challenges. One respondent cited the fact that the new system will be more resource-intensive in a fiscally austere environment.
- **System design—**Six respondents from two pilot sites cited interrater reliability as a major challenge for maintaining the integrity of the new evaluation systems. One interview participant expressed concern over the authority that principals have to drive achievement and viewed the new model as increasing opportunities for principals to gain resources. Five respondents were concerned about ensuring the clarity of terms and rubrics for teachers and principals.

Overall, respondents seemed enthusiastic about the project and touted the benefits of well-designed teacher and principal evaluation systems. At the same time, they recognized the challenges not only of design but of implementation. It was obvious that a great amount of time and thought had been directed toward the development of these principal and teacher evaluation systems; a number of pilot sites, however, still have some big decisions about performance indicators and measures. Although, on average, the new principal evaluation systems are not as fully developed as the teacher systems, respondents were hopeful that work on the principal systems would proceed more quickly.

Appendix A
Sample Elements

Here are presented examples from two pilot sites of elements developed to help them observe and measure teachers' progress toward the new state teaching criteria. The state criteria are listed in the first column and the elements in the second column.

Sample 1

Criterion	Elements
1. Centering instruction on high expectations for student achievement	a. Alignment of instructional practices to expectations b. Facilitation of student engagement
2. Demonstrating effective teaching practices	a. Lesson organization b. Instructional practices c. Adherence to alignment of content standards
3. Recognizing individual student learning needs and developing strategies to address those needs	a. Assessment use b. Responsiveness to student learning c. Adaptation of instruction
4. Providing clear and intentional focus on subject-matter content and curriculum	a. Content knowledge b. Use of standards c. Pacing
5. Fostering and managing a safe, positive learning environment	a. Creating an environment of respect and rapport b. Classroom procedures c. Management of physical space d. Management of student behaviors
6. Using multiple student data elements to modify instruction and improve student learning	a. Assessment of data b. Assessment design c. Data-driven lesson design
7. Communicating and collaborating with parents and school community	* At the time of the interviews, this pilot site was still developing elements for criterion 7.
8. Exhibiting collaborative and collegial practices focused on improving instructional practice and student learning	a. School improvement and district initiatives b. Professional growth and development c. Professional learning communities

Sample 2

Criterion	Elements
1. Centering instruction on high expectations for student achievement	a. Lesson planning b. Lesson delivery c. Lesson assessment
2. Demonstrating effective teaching practices	a. Questioning and discussion techniques b. Incorporating higher order thinking questions c. Instructional delivery d. Flexibility and responsiveness e. Student engagement
3. Recognizing individual student learning needs and developing strategies to address those needs	a. Knowledge of student needs b. Planning for individual student needs c. Addressing student needs d. Using data to differentiate instruction
4. Providing clear and intentional focus on subject-matter content and curriculum	a. Knowledge of content b. Use of content c. Intentional focus on academic vocabulary d. Staying on task
5. Fostering and managing a safe, positive learning environment	a. Routines and procedures b. Ensuring a positive learning environment with respect and rapport c. Managing student behavior d. Physical environment
6. Using multiple student data elements to modify instruction and improve student learning	a. Knowledge of available data b. Responsiveness to data
7. Communicating and collaborating with parents and school community	Participation in parent/school community
8. Exhibiting collaborative and collegial practices focused on improving instructional practice and student learning	a. Participation in professional learning communities b. Collegial and professional practices and interactions with colleagues

Appendix B
Interview Protocol

Presented here is the interview protocol used to interview four pilot committee members from each of eight pilot districts, excluding the district consortium.

For each of the questions below, probe separately for

1. The teacher evaluation system and principal evaluation system
2. What has been implemented, what is planned, and what is yet to be determined

The purpose of this interview is to gather information about the new educator evaluation system that is currently being planned in your district. We would also like to know what you consider to be the perceived strengths of the educator evaluation system and the challenges of implementation. The information you provide will be used to help the Washington Office of the Superintendent of Public Instruction understand how to better support districts' implementation of Senate Bill 6696, which established new criteria for evaluating educators.

The discussion today is voluntary and confidential. No individual or their school will be identified in any reports. If there is any particular background information that you wish to share but do not want to be associated with your specific district in the report, please let us know during the focus group. We would like to tape-record our discussion with you. The discussion would be recorded solely for the purposes of ensuring data-collection accuracy. We will destroy the tape once a transcript has been finalized. The interview will take approximately 60 minutes.

Framework/Goals

1. Has your district developed or adopted an instructional framework to guide teachers? If so, please describe the development process.

   Was the framework's design influenced by models that have already been established?

   Examples: Teach for America, University of Washington Center for Education Leadership (CEL), Charlotte Danielson, Marzano model

   a. What instructional standards or elements has your district developed?

   b. Has your district linked its instructional standards/framework to its teacher evaluation system?

2. What are the major goals of your district's teacher evaluation system?

   Examples: Continuing contract decisions, hiring/firing, PD, compensation, assignment decisions, promotion/advancement, identifying teacher leader

3. What are the goals of your principal evaluation system?

   Examples: Continuing contract decisions, hiring/firing, PD, compensation, assignment decisions, promotion/advancement, identifying teacher leader

Criteria/Evaluation Format

4. What type of criteria will you use to measure performance? For teachers? For principals? *(Specific criteria for evaluating teachers/principals should be embedded in the rubrics being developed in pilot districts.)*

5. Who will be evaluated under this new system and set of criteria? *(e.g., general education teachers, core teachers, noncore teachers)*

6. Will the frequency of the evaluations differ for provisional and continuing contract teachers? If so, how will it differ?

7. Is your district's evaluation system based on ratings? If so, please describe the rating system or future plans for developing a rating system.
   a. How many levels does the rating system include?
   b. Are there specific benchmarks that determine rating levels? For teachers? For principals?
   c. Are measures included that will base a teacher's performance on how well they performed compared to their peers?

      Example: A teacher's value added score might be used to rank a teacher against other teachers in the school, district, or state.

Measures

8. What measures will your district use to evaluate performance?

   Examples: teacher value-added scores, observations, portfolios, peer observations, lesson plans, attendance at committee meetings

   a. Do you plan to use student growth measures in the evaluation? If so, for whom?
   b. Describe the measure, how it is calculated, and how it will be used to measure performance.

9. Will the measures differ for core and noncore teachers? If so, how?

Training

10. Describe the types of training evaluators will receive.

11. Will your evaluators be receiving any training on the following?
    - Ensuring compliance
    - Using instruments reliably
    - Ensuring reliability of scores across raters and evaluators

- Evaluation of content

12. Does the district plan to offer training on data use to inform student improvement? If so, please describe the training.

Professional Development

13. Will evaluation results be used to inform professional development needs? If so, please describe.

Communication

14. What is your district's timeline for implementing the new evaluation system?
    a. When will you implement a new rating system?
    b. When will your district start using a new set of teacher rubrics? Principal rubrics?
    c. When will professional development be provided?
    d. Other elements to be included?

15. How do you plan to communicate the new system requirements to principals, teachers, and other stakeholders?

General Strengths and Weaknesses

16. What are the biggest strengths of your new evaluation system?

17. What do you anticipate will be the biggest challenges for implementing and sustaining your new evaluation system?

Thank you very much for your helpful information.

Appendix C
Focus Group Protocol

Presented here is the focus group used for the consortium focus groups.

For each of the questions below, probe separately for

1. The teacher evaluation system and principal evaluation system
2. What has been implemented, what is planned, and what is yet to be determined

The purpose of this focus group is to gather information about the new educator evaluation system that is currently being planned in your districts. We would also like to know what you consider to be the perceived strengths of the educator evaluation system and the challenges of implementation. The information you provide will be used to help the Washington Office of the Superintendent of Public Instruction understand how to better support districts' implementation of Senate Bill 6696, which established new criteria for evaluating educators.

The discussion today is voluntary and confidential. No individual or their school will be identified in any reports. If there is any particular background information that you wish to share but do not want to be associated with your specific district in the report, please let us know during the focus group. We would like to tape-record our discussion with you. The discussion would be recorded solely for the purposes of ensuring data-collection accuracy. We will destroy the tape once a transcript has been finalized. The focus group will take approximately 90 minutes.

Framework/Goals

1. Has your district developed or adopted an instructional framework to guide teachers? If so, please describe the development process.

   Was the framework's design influenced by models that have already been established?

   Examples: Teach for America, University of Washington Center for Education Leadership (CEL), Charlotte Danielson, Marzano model

   a. What instructional standards or elements has your district developed?

   b. Has your district linked its instructional standards/framework to its teacher evaluation system?

2. What are the major goals of your district's teacher evaluation system? *(e.g., continuing contract decisions, hiring/firing, PD, compensation, assignment decisions, promotion/advancement, identifying teacher leaders)*

3. What are the goals of your principal evaluation system? *(e.g., continuing contract decisions, hiring/firing, PD, compensation, assignment decisions, promotion/advancement)*

Criteria/Evaluation Format

4. What type of criteria will you use to measure performance? For teachers? For principals? *(Specific criteria for evaluating teachers/principals should be embedded in the rubrics being developed in pilot districts.)*

5. Who will be evaluated under this new system and set of criteria? *(e.g., general education teachers, core teachers, noncore teachers)*

6. Will the frequency of the evaluations differ for provisional and continuing contract teachers? If so, how will it differ *(e.g., provisional teachers are formally observed twice a year and continuing contract teachers will be observed only once; it will be up to the principal to decide)*

7. Is your district's evaluation system based on ratings? If so, please describe the rating system or future plans for developing a rating system.
   a. How many levels does the rating system include?
   b. Are there specific benchmarks that determine rating levels? For teachers? For principals?
   c. Are measures included that will base a teacher's performance on how well they performed compared to their peers? For example, a teacher's value added score might be used to rank a teacher against other teachers in the school, district, or state.

Measures

8. What measures will your district use to evaluate performance?

   Examples: teacher value-added scores, observations, portfolios, peer observations, lesson plans, attendance at committee meetings

   a. Do you plan to use student growth measures in the evaluation? If so, for whom?
   b. Describe the measure, how it is calculated, and how it will be used to measure performance.

9. Will the measures differ for core and noncore teachers? If so, how?

Training

10. Describe the types of training evaluators will receive.

11. Will your evaluators be receiving any training on the following?
    - Ensuring compliance
    - Using instruments reliably
    - Ensuring reliability of scores across raters and evaluators

- Evaluation of content

12. Does the district plan to offer training on data use to inform student improvement? If so, please describe the training.

Professional Development

13. Will evaluation results be used to inform professional development needs? If so, please describe.

Communication

14. What is your district's timeline for implementing the new evaluation system?
    a. When will you implement a new rating system?
    b. When will your district start using a new set of teacher rubrics? Principal rubrics?
    c. When will professional development be provided?
    d. Other elements to be included?

15. How do you plan to communicate the new system requirements to principals, teachers, and other stakeholders?

General Strengths and Weaknesses

16. What are the biggest strengths of your new evaluation system?

17. What do you anticipate will be the biggest challenges for implementing and sustaining your new evaluation system?

Thank you very much for your helpful information.

APPENDIX D

Anacortes Case Study

Central Valley Case Study

Kennewick Case Study

NEWESD 101 Consortium Case Study

North Mason Case Study

North Thurston Case Study

Othello Case Study

Snohomish Case Study

Wenatchee Case Study

Anacortes Public Schools
Case Study Report

District Overview/Description[1]

Anacortes Public Schools serves about 2,700 students each year through its four elementary schools, one middle school, two high schools, and one home education partnership. The school district employs approximately 300 staff. Anacortes is located 60 miles north of Seattle, and its student population is as follows:

- 2.08 percent of Anacortes students are American Indian or Alaskan Native.
- 4.09 percent are Asian or Pacific Islander.
- 1.61 percent are African American.
- 5.92 percent are Hispanic.
- 35.34 percent are eligible for free or reduced-price lunch.

Case Study Method

This study was undertaken to help the Washington Office of the Superintendent of Public Instruction (OSPI) better understand how to support districts' implementation of Senate Bill 6696, which establishes new criteria for evaluating educators. For this component of the study, four individuals affiliated with the Anacortes School District were interviewed in an effort to collect information concerning the process of developing and implementing the district's new teacher and principal evaluation systems. Each of the participants has played an active role in the development of the new systems. In an attempt to gain a diverse range of institutional perspectives, we interviewed one district-level administrator, one school-level administrator, and two teachers. All responses were coded based on topic and related through the following case study.

Moreover, the district was requested to submit pertinent documents concerning the development and implementation of its teacher and principal evaluation systems. These documents were then reviewed by AIR staff who documented evidence of the district's actions.

Teacher Evaluation

Setting the Evaluation Framework and Goals

The new teacher evaluation system in Anacortes is heavily linked to its instructional model, the Five Dimensions of Teaching and Learning developed by the University of Washington Center for Educational Leadership. The five dimensions include purpose, student engagement, assessment for student learning, classroom environment and culture, and curriculum and

[1] Information concerning Anacortes School District was found at the district's website (www.asd103.org/) and the Washington State Report Card (http://reportcard.ospi.k12.wa.us/DataDownload.aspx).

pedagogy. As a pilot district for this model, Anacortes has been in the process of training staff on this framework for several years.

One interview participant, who had been involved in the process of choosing this instructional model, described the process of adopting the framework as follows: One group of Anacortes elementary teachers and one group of Anacortes high school teachers were charged with using it for professional development for one year. They met monthly, receiving training on the framework, and had the opportunity to observe classrooms and have conversations about what they saw. Based on their positive experiences with it, at the end of the year the two groups jointly recommended to district leaders that the instructional framework be adopted districtwide. Since then, additional cohorts of teachers and principals have been trained in using the model.

One interview participant viewed the district as now being about 70 percent of the way toward fully implementing the model. Meanwhile, the pilot committee has used the model as a basis for the new evaluation system. One respondent noted that, in fact, most of the committee's time has been spent conducting crosswalks between the five dimensions to the state's eight evaluation criteria to identify gaps and making tweaks to fill them in. The committee has been working with several consultants from the University of Washington to do this.

All interview participants agreed that the main goal of the teacher evaluation system is teacher growth. One respondent described how the first major goal is to develop a tool that itself will foster growth, and the second major goal is to make this tool a catalyst within a larger systemwide approach for improving teacher effectiveness through professional development, support, and effective use of time. One respondent believes a focus on growth would affect a larger number of people than other focuses, such as termination. The respondent also noted that the committee felt that if its goal had been to fire teachers more easily, the district simply needed to apply the system it already had rather than create a whole new one. Another respondent noted that under the current system, there is no incentive for teachers to improve except for an intrinsic desire to do so, but the new system is intended to encourage teachers not to settle for being good but rather to strive for being great by working on specific areas.

The pilot committee came to these decisions by trying to start the process by consulting research, with no preconceived ideas of what the new system would look like. One respondent described how committee members visited Tennessee to learn about merit pay and looked at study after study so they could use research in the best interests of the district's staff and students.

Establishing Evaluation Criteria

As the law states, Anacortes is adopting the eight criteria set forth for teacher evaluation. The current draft evaluation rubric includes five dimensions, with multiple elements under each dimension. One respondent described how committee members initially struggled to maintain consistency with their instructional model while meeting the legal requirements because the eight criteria provided by the state did not fit their own categories clearly. In response, the district created three categories—preparation, teaching, and self-reflection—and organized the eight criteria, as well as several additional categories the district created, within these three areas.

The new evaluation system will apply to all teachers. All interview participants agreed that this would be the case, although their responses varied somewhat. One respondent stated that everybody at every level will use the same system. Two respondents noted that all those instructing students, including music, drama, physical education, and special education teachers, will be evaluated under the new model, with different tools being developed for school counselors, occupational therapists, and speech/language pathologists. A fourth respondent reported that all classroom teachers will be included.

Teachers will not be rated against one another but rather against the standards on the rubric, on the following four-tiered system:

- Level 1: Unsatisfactory
- Level 2: Basic/Emerging
- Level 3: Proficient
- Level 4: Distinguished

Two respondents noted that the cutoff point between satisfactory and unsatisfactory lies between Levels 1 and 2. But they are hoping to include in the contract that even a Level 2 teacher is likely to be considered satisfactory only for a specified amount of time, after which the teacher will undergo a review and perhaps a plan of improvement, or even revert back to Level 1.

The pilot committee currently is in the process of considering these details and refining the language that defines each rating level, as well as what the evidence would look like at each level.

Frequency of Evaluation

The frequency of evaluations has not yet been determined for provisional or continuing contract teachers; however, three of the interview participants suggested it was likely that these two groups of teachers would and should be treated differently. This element is codified in state legislation (E2SSB6696, Part II, § 202(3)(a)).

Measuring Performance

The documents provided showed that there are no descriptions of the components of the evaluation rubric to ensure their common understanding, and the descriptions of behavior differentiating the levels of performance may be too subjective to obtain reliable evaluations. For almost all of the elements, the levels differ only by frequency with which the teacher engages in a particular activity (rarely, occasionally, frequently, or always).

At the time of the interviews, the committee members had only brainstormed about the measures that would be used to evaluate teachers, but they planned to have a more in-depth conversation about the issue soon. One respondent expressed a desire that a self-assessment, combined with reflection, be part of the evaluation system and that classroom observations, a portfolio based on the educator's goals, and a variety of forms of evidence of performance also be featured in the

final system. Another respondent stated that observations and the use of artifacts to deal with nonobservable qualities are the only measures that have been confirmed so far.

The district has not yet determined how or if student academic data will be included in the evaluation system or how those data will be linked to teachers. It was noted by one respondent, who believed that student survey data is an important component of teacher growth, that the issue had been discussed and probably would not be used because the union disagreed with the idea. Another respondent noted that, although the committee had not yet addressed test scores, student growth still could be used if weaved into the criteria about teachers' use of data. A third respondent posited that, although using student test scores has not been ruled out, value-added scores are something the committee is not interested in, except perhaps as a formative tool. Another respondent noted that the committee had looked at ideas for including a student growth measure, but no decisions have been made about this yet. One respondent noted that, regardless of which measures were chosen, the 6696 legislation requires that multiple measures be used so that would clearly be a part of their final decision.

There has not been discussion of differential weighting of measures or types of data that will be collected. The district also has not yet progressed enough to have information about data use and tracking.

Training for Evaluators

The process for selecting evaluators has not yet been decided. The training that teacher evaluators in Anacortes will receive also has not yet been determined. All interview participants agreed that training will take place, and three respondents emphasized the need for it to be robust training. For example, one respondent believes the professional development will have to be more than just one day. One respondent believes it is likely to include training on ensuring compliance but not training on evaluation of content, because of the small size of the district.

Two respondents brought up the work that the district has been doing with experts from the University of Washington and Western Washington University on their instructional model, and one noted that these experts also might be able to help with training on interrater reliability, which was brought up by two interview participants as an important consideration. The other believed the Five Dimensions model and the crosswalk they developed already had tremendous interrater reliability.

The interview participants suggested that the principals would likely be responsible for conducting teacher evaluations. One respondent noted that about half of the district's principals serve on the pilot committee, so they will require less training than those principals who are less familiar with the instrument.

Using Results for Professional Development

All interview participants viewed professional development as a key outcome of the evaluations, although no policies regarding this aspect of the system have yet been decided. One respondent viewed professional development as the driver behind the system.

One interview participant noted an idea that had been discussed, that is, for teachers at Level 3 to be able to take a professional growth option the following year rather than have observations; meanwhile, teachers who are struggling would perhaps be placed on an improvement plan. Two respondents posited that, because ratings will be provided not just overall but also for each of the eight criteria, the new system will provide greater clarity about where teachers are within the teaching model, allowing more targeted professional development to be provided. One respondent stressed the need to move beyond a one-size-fits-all approach to a more individualized form of professional development.

Two respondents discussed how trends in evaluation results also would guide buildingwide or districtwide professional development activities. Anacortes already provides training on using data to inform instruction through the assessment component of its Five Dimensions instructional model, which will form the basis of the evaluation system. This training will continue when the new evaluation system is implemented.

Communicating Results

The Anacortes pilot committee has yet to determine a number of issues, including the number of observations that will be required and who or how many observers there will be. One respondent noted that some teachers association issues need to be worked out first. The committee's next steps are to determine the multiple measures evidence and then to focus on the evaluation process. Three respondents noted that in the next month or so, the district will begin testing the rubric with videos, which are provided by the University of Washington, to determine if the various parts of the rubric are observable and can be "score-able." Two respondents stated that subsequently, this spring, the rubric will be unofficially tested with teacher volunteers in their classrooms. The goal is for the tool and process to be in place for the official pilot in 2011–12. One respondent noted that, although it may be ambitious and will probably require quite a bit of time spent during the summer to refine the system, the district plans to have everything in place for full implementation in 2012–13.

Each of the interview participants indicated that they believed the district already had done a great deal to communicate with teachers and other stakeholders. They each relayed that the superintendent and a teacher representative visited each school building in the district in the autumn, as well as a second round of visits more recently to update staff on the criteria. They also recounted the efforts of the pilot committee and what is taking place in the pilot districts across the state.

In addition to these school building visits, the teacher representative conducted one-on-one listening sessions through a "Listening Campaign" to hear concerns and to further inform education staff about the reforms. Also, the pilot committee includes two representatives from each building who are available to answer questions. One respondent noted that his or her school's representative presented an update during a staff meeting and that teachers had a lot of questions. Outside of the schools, the committee has met with the school board and held a number of community open informational meetings for parents and other community members.

On top of these face-to-face meetings, the committee has created a website and blog that are updated after each committee meeting and whenever new information emerges. There is a space on the website for people to ask questions and receive answers. The website also includes documents collected from the committee's research. In addition, information is exchanged via a newsletter and a survey that was conducted to solicit initial feedback about the evaluation system. So, as one respondent pointed out, the lines of communication are kept open. It is not simply a matter of the committee communicating to stakeholders. To do all of this, one respondent pointed out, a teacher has been given half-time release to work exclusively on this project. However, the type of feedback and any actions taken as a result of that feedback are not apparent. In addition, the details of the pilot and how feedback from it will be used to improve the system also do not appear to have been determined.

One interview participant noted that the feedback received from staff at first was very positive but that staff members have become more concerned as the time has neared for the new system to take effect. The respondent said that the committee intends to address ongoing concerns by continuing to visit buildings, holding at least one more community or parent event, and continuing to add new information to the website and blog. Another respondent added that there will be some professional development prior to the pilot and throughout the year.

Principal Evaluation

The Anacortes pilot committee has devoted less attention to its principal evaluation system compared with its teacher evaluation system. Two respondents attribute this to the fact that the Association of Washington School Principals (AWSP) has been so proactive in developing a principal evaluation model for districts to adopt.

Setting the Evaluation Framework and Goals

Anacortes School District does not have a leadership framework available, but interview respondents conveyed that they will be relying heavily on the AWSP's model for their principal evaluation reforms. An AWSP committee has been proactively working on the design of a model for principal effectiveness for several years. The model is closely aligned with the eight criteria for principal evaluation established by the state. One interview participant reported that the district has adopted the AWSP model, and another noted that the district's model will be linked to the AWSP model. One respondent noted that the district also is working with the state and Interstate School Leaders Licensure Consortium (ISLLC) standards, and two respondents noted that the Five Dimensions model for teacher effectiveness, which the district has been piloting for two years, also will be linked where appropriate to principal evaluation.

When asked about the goals of the principal evaluation system, respondents agreed that, as with the new system for teachers, the principal evaluation system is intended to be a growth model. One respondent clarified that the goal was to achieve growth for principals, for teachers, and for students. Another respondent explained that the main goal is to impact classroom teaching and student achievement but added that using data, being an instructional leader, and being a visionary are also goals.

The district has not yet discussed whether principal evaluations should be used for purposes other than growth, such as for determining compensation or other human capital policies.

Establishing Evaluation Criteria

The criteria used in the principal evaluation model will include the state's eight mandatory criteria. One respondent stated that the AWSP model the committee is using also includes a ninth criteria that holds principals accountable for closing the achievement gap, around which the committee has had some discussions. One respondent viewed as an intriguing feature of the AWSP's model the fact that alongside the criteria were columns for supports and authority for the principal to operate in that area.

Principal evaluations in Anacortes will use the following four-tiered system:

- Level 1: Unsatisfactory
- Level 2: Emerging
- Level 3: Proficient
- Level 4: Exemplary

One interview respondent stated that the rubric language has not yet been developed for what constitutes Levels 4, 3, 2, or 1. It does not appear that objectives, rubrics, and measures have been identified, or that the evaluation procedures for principals have been determined. One respondent noted that the committee will likely look at the rigor of the teacher evaluation system, after it is more fully developed, and ensure that the same things are applied in the principal evaluation system.

Frequency of Evaluations

The committee has not progressed enough in the development of its principal evaluation system to have determined the frequency of evaluations, which will be a determining factor in the validity of the principal evaluation system.

Measuring Performance

The AWSP model includes the evidence that can be used to demonstrate whether each criterion was met; however, the district has not yet determined how evidence will be used to demonstrate that the criteria have been met. One respondent noted that a lot of evidence sharing between principals and their evaluators already takes place in Anacortes because it is a smaller district and so may require less structure in terms of evidence collection than would be the case in a larger district.

The committee has not yet determined whether test scores will be included in the principal evaluation system. One respondent noted that student test score data can be used now with the current evaluation system under the knowledge of student progress criterion. Whether a data-

tracking process will be developed or what it might look like also have not been decided at this point.

Training for Evaluators

The pilot committee provided neither evidence that an evaluator training process has been designed nor a timeline for its development.

Using Results for Professional Development

Although the district plans to use the evaluation system to promote principal growth, there is no evidence to date of how evaluations would be linked to professional development or to other components of a comprehensive human capital management strategy, such as progressive discipline or compensation policies.

Communicating Results

Limited information exchange appears to be in progress, including school-based listening sessions and consultation with researchers. However, it does not appear that Anacortes has progressed to the point of developing a plan for its pilot testing of its principal evaluation system, tracking of evaluation results, or communicating the reforms to stakeholders. It does not seem that an independent evaluator or other methods for gauging satisfaction, fidelity, reliability, or utility of the new principal evaluation system have been identified.

Summary and Conclusions

One interview respondent believes the greatest strength is that the new system clearly articulates what good instruction looks like so that teachers can genuinely change their instructional practice for the better. Likewise, principals can read the eight standards and clearly see what it looks like to be an effective principal and what they need to do to get there. Two interview participants see the greatest strength as being able to differentiate teaching beyond satisfactory and unsatisfactory so that both struggling and high-performing staff can be recognized and attended to and can themselves see where they stand and how they can improve. Another strength is that the respondents are viewing their evaluation reforms as a systems approach to assist staff in continual growth. In addition, they started from square one with a blank slate. As such, they agreed as a committee to look at everything taking place across the country to identify the most cutting-edge approaches. They are doing their homework and reading the research available to them.

All interview participants were concerned about time and resources. Finding the time and resources to do things right, to have meaningful conversations and multiple observers and multiple occasions were seen by two participants as among the greatest challenges ahead. One thought that reducing reporting requirements or allowing some teachers' evaluations to take place every other year could free up time needed to do evaluations correctly. Another respondent thought that supporting sustained professional development for all current and future staff involved, so that the system is understood by all, would be a challenge. A fourth respondent was

concerned that the economic cutbacks might make staff weary that as professional supports are taken away from them (e.g., larger class sizes, less financial compensation, and fewer resources), they will be evaluated and expected to perform at higher levels. One interview participant expressed concern over the authority that principals have to drive achievement and viewed the new model as opening up opportunities for principals to gain resources and engage in conversations with superintendents about how they can do so. Another viewed skepticism among faculty as a major challenge. Some staff in the district are suspicious that the reforms are an effort to "whack teachers over the head," and others in the community want to make it easier to fire poor-performing teachers and will not see the reforms as aggressive enough.

In all, the teacher and principal evaluation reforms in Anacortes School District were seen as very much a work in progress, but those involved expressed confidence and enthusiasm for the final result. One interview participant referred to the process of developing the reforms as one of the best professional discussions and opportunities experienced during that respondent's career in education.

Central Valley Public Schools
Case Study Report

District Overview/Description[1]

Central Valley School District serves nearly 12,400 students each year through its 22 schools, including 12 elementary schools, five middle schools, three high schools, one nontraditional elementary/middle school, and one home-link partnership. Serving an 80-square-mile area in the Spokane Valley, Central Valley School District is a largely Caucasian school district, with the following demographic composition:

- 1.31 percent of Central Valley students are American Indian or Alaskan Native.
- 2.02 percent are Asian or Pacific Islander.
- 1.16 percent are African American.
- 3.25 percent are Hispanic.
- 38.41 percent are eligible for free or reduced-price lunch.

Case Study Method

This study was undertaken to help the Washington Office of the Superintendent of Public Instruction (OSPI) better understand how to support districts' implementation of Senate Bill 6696, which establishes new criteria for evaluating educators. For this component of the study, three individuals affiliated with Central Valley Public Schools were interviewed in an effort to collect information concerning the process of developing and implementing the district's new teacher and principal evaluation systems. Each of the interview participants has played an active role in the development of the new systems. In an attempt to gain a diverse range of institutional perspectives, we interviewed one district-level administrator, one principal, and one teacher. All responses were coded based on topic and related through the following case study.

Moreover, the district was requested to submit pertinent documents concerning the development and implementation of its teacher and principal evaluation systems. These documents were then reviewed by AIR staff who documented evidence of the district's actions.

Teacher Evaluation

Setting the Evaluation Framework and Goals

In Central Valley, there is no formal instructional model for teachers, but all respondents did view the district's evaluation system as linked to a mix of Robert Marzano's Classroom Instruction that Works strategies and the workshop model Gradual Release of Responsibility. Two interview participants mentioned Marzano's instructional strategies as informing their

[1] Information concerning Central Valley School District was found at the district's website (www.cvsd.org/) and the Washington State Report Card (http://reportcard.ospi.k12.wa.us/DataDownload.aspx).

evaluation work, and another mentioned his idea of a guaranteed and viable curriculum. Aiming for consistent curriculum, consistent scope, and sequence timelines are among the elements included. Two respondents described how the district has been working with Denver-based McREL both on the implementation of Classroom Instruction that Works and the development of its evaluation system, for which the district receives ongoing feedback. In this sense, all respondents agreed that their evaluation system was linked to the district's instructional standards, although one noted that the evaluation system is not tied directly to Marzano or anyone else's vocabulary, and that this had been decided because specific terms or models change quickly in the education field.

One respondent referred to an instructional emphasis on protecting instructional time, lesson delivery, and consistent vocabulary. Two respondents additionally referred to a workshop model for reading, writing, and math, which is emphasized at the elementary level, as well as gradual release of responsibility, which is emphasized at the middle school level. The use of these models emerged from requests from teachers and principals; principals then took the information back to their teachers to solicit broader input and then coordinated train-the-trainer sessions for lead teachers, who trained their peers on the materials before they became part of the expectations of principals during walkthroughs.

Interview participants agreed that the new evaluation system is intended to be a growth model, including individual goal setting, self-reflection, identification of areas of strength and weakness, and collaboration between evaluators and those being evaluated. One respondent noted that clearly defining what quality teaching and leadership look like was a goal and that fulfilling a part of the district's strategic plan also was a goal. One respondent viewed the first major goal as simply to meet the requirements of the law.

Establishing Evaluation Criteria

The work of establishing teacher evaluation criteria began with a committee of 50–60 teachers, principals, central administrators, staff developers, and parents, who worked to define what each criterion meant and developed observable descriptors for each. All respondents agreed that Central Valley was using the eight state criteria exactly as they are written in the law, and then expanding and refining the descriptors of each. The rubric that has been developed is intended for use both as an observation instrument and as an intermediary step on the way to a summative rating. The instructional framework lists performance criteria by standard and component, with general descriptions of performance at each level; the checkboxes and language that suggest application of the rubric over multiple points in time may make this tool difficult to use reliably across multiple evaluators.

Central Valley has adopted a four-tiered rating system composed of the following:

- Level 1: Not Demonstrated/Unsatisfactory
- Level 2: Developing
- Level 3: Proficient
- Level 4: Accomplished

Participants receive a rating for each of the eight criteria. One respondent described how a teacher must meet all of the developing criteria to be considered developing; all of the developing and proficient criteria to be considered proficient; and then all of the developing, proficient, and accomplished criteria to be considered accomplished. The descriptors mentioned above help constitute evidence of a teacher's rating. This rating system will replace the existing binary system whereby teachers are simply rated as *Satisfactory* or *Unsatisfactory*.

All Central Valley teachers rated as *Developing*, *Proficient*, or *Accomplished* meet expectations for the district, regardless of their career stage. However, it appears from the documents currently available that the *Unsatisfactory* rating in the instructional framework, which teachers receive if they are not at least developing skills, has been eliminated for the summary ratings; only *Developing*, *Proficient*, A*ccomplished*, and *Not Demonstrated* are used on the document titled Teacher Evaluation Form A. The elimination of the *Unsatisfactory* category for summary ratings could be confusing; for example, if a teacher received an *Unsatisfactory* rating during an observation, what standard-level rating would he or she receive? In addition, this omission could skew the distribution of ratings upwards as it is unclear how any teacher would ever be rated *Unsatisfactory* because *Not Demonstrated* is not equivalent to *Unsatisfactory*.

All interview participants agreed that teachers would be evaluated on the basis of the criteria and not compared with one another. One respondent, however, noted that the district hopes to incorporate peer coaching and that comparisons of teachers' ratings could potentially be used in those decisions.

With respect to which teachers would be evaluated under the new system, respondents were in agreement that essentially all teachers who have regular contact with a classroom of children would eventually be evaluated under the new system. This includes physical education teachers, music teachers, and media specialists. One respondent noted that only occupational therapists, counselors, and speech and language pathologists would be excluded.

Frequency of Evaluations

The frequency of teacher evaluations has not yet been determined, with one interview respondent noting that the district and union decided to wait until the spring to have conversations about whether the frequency of evaluations for provisional and continuing contract teachers would differ. Another respondent stated that, as per the law, provisional teachers will have one additional observation within the first 90 days of the school year. Another respondent noted that, although the legislature has not specified the number of observations that must take place, Central Valley is envisioning the process involving more than two conversations a year, and ongoing reflection, growth, and goal setting. Two respondents referred to the district's existing long-form and short-form options; one respondent hoped that they could continue to have both options, such that not all teachers needed to be evaluated on the long-form every year. Evaluation frequency is codified in the state legislation (E2SSB6696, Part II, § 202(3)(a)).

Measuring Performance

Central Valley has not yet decided which measures of teacher effectiveness will be used in the evaluation system; however, for each of the eight criteria, the committee has identified several

pieces of evidence or artifacts that could be used to demonstrate unobservable practices and, according to one respondent, the plan is to identify more during the pilot. One interview participant confirmed that lesson plans and walkthroughs would be among the types of evidence incorporated. Another stated that it is unlikely that student or parent surveys would be used, but that evidence of a teacher's outreach and communication with parents would be included.

Within the instructional framework, the artifacts and sources of evidence listed for each component include parent logs, lesson plans, posted standards, and other evidence. The exceptions were student work and student assessment data, but respondents were in agreement that student test scores would not be directly tied to teachers' evaluations. Rather, teachers will be expected to differentiate instruction for students based on what the ongoing data—being collected by the teachers—conveys about students' strengths and their misconceptions. Two respondents shared that teachers will not be expected to create portfolios. One of these respondents explained that the goal was to not require teachers to do extra work, but rather that the evidence collected would be a part of their regular classroom work, presented in an honest, safe, comfortable conversation in the teacher's own classroom where the needed evidence should be readily available.

Central Valley School District plans to use the same measures for core and noncore teachers. One interview participant agreed that applicability to both core and noncore teachers was the reason the district did not specifically define the measures to be used, particularly how formative and summative classroom assessments were to be used.

It was noted in the documents reviewed that Central Valley School District is working with McREL to design teacher evaluation systems and tools and that some of this work is around tool reliability and validity.

Training for Evaluators

All interview participants shared that McREL also would be providing evaluator training in March. This one-day training would focus on how the tool works. Two respondents believed that the pilot teachers being evaluated would be attending the same or a similar training so that all involved in the pilot would hear the same message and learn together about how the tool works. One respondent noted that the initial training would involve significant conversations taking place around what the descriptors mean, what they look like in practice, and how effective practices will be measured, and that additional training on the tool will be provided as the system is rolled out.

Two respondents believed that the McREL training would address issues of ensuring compliance, but a third respondent did not view that as being a significant piece of the training and saw ensuring compliance as equally important to the training of both those being evaluated and the evaluators. There was otherwise little indication of the training content that would be included. Furthermore, it was unclear from the documents available at the time whether evaluations would involve classroom observations, portfolio reviews, parent and student surveys, or other means of gathering evidence; therefore, it was unclear who would require training in gathering the information needed or in issuing summative teacher evaluation ratings.

One respondent stated that, to date, nothing concrete had been planned to train evaluators for interrater reliability, but that the committee members have several ideas. One idea on the table is to have teachers volunteer to videotape some lessons that principals can use for ongoing training to ensure interrater reliability.

Using Results for Professional Development

The respondents were in agreement that the focus of the new system will be on professional development, although there are still more conversations to be had about how this will look. One respondent described this scenario: No matter where a teacher is in his or her performance, there will be an opportunity to be supported in growing as a professional. Two respondents envisioned that common themes will emerge from the goal setting that takes place as part of the new evaluation system, and that various opportunities for collaborative professional development will be provided based on these trends.

One respondent expressed concern over professional development budget cuts and other budget cuts, and noted that the district will have to be innovative in its approach. For example, two respondents referred to within-district teacher experts or staff developers providing this professional development. Book studies were cited by one respondent as an approach that might be taken. Another respondent noted that, although there will be many conversations and much input, the professional development will be driven by the individual and so will be meaningful to each staff member.

It does not appear that specific training on using data to improve instruction will be directly tied to the new system, and all respondents agreed that this is something the district already does. Two respondents noted that the district's guaranteed viable curriculum training around Classroom Instruction that Works includes this piece. One respondent referred to a one-hour collaboration time scheduled each week for this type of training. Another respondent described how the district works with school principals who then work with their school improvement teams on the use of data.

Communicating Results and Ensuring Stakeholder Investment

Central Valley is engaging in a four- to eight-week pilot in April and May with 50–60 volunteer teachers and about 50 principals and assistant principals. During the summer, the committee will make adjustments based on feedback from participants in the pilot and from data collected by McREL. McREL will manually gather data from evaluator-teacher interactions, conduct Web-based surveys, and perform one-on-one interviews of participants.

One respondent noted that the district and union intentionally decided to wait until the spring to have many of the conversations about the specifics of the new system so that they could focus first on developing the rubrics and the pilot. A second, yearlong, field test will take place during the 2011–12 school year with all of the district's principals and 125–150 teachers randomly selected by McREL. One participant noted that the goal is for the system, including the tool and wider components of the system, to be near finalization in time for this yearlong field test.

During the summer, the data from the initial pilot will be used to inform changes for the next phase of the reform.

One participant noted that feedback would be collected during the yearlong pilot and further adjustment made such that OSPI would have a fairly complete package to present to the legislature. One respondent noted that, depending upon how negotiations go, the district may enact a partial adoption in 2012–13 for those who opt in, with the expectation that everyone would be on board in 2013–14.

All respondents agreed that a large amount of effort already has gone into communicating these changes to stakeholders. One respondent spoke of the importance of creating a shared understanding among all staff to make what may seem like a drastic change less overwhelming or uncomfortable for teachers. All respondents mentioned the sharing of draft rubrics through several visits to every school in the district by a core team of seven committee members, including the teacher and principal union presidents. This approach is intended to ensure consistency in the message delivered.

Two respondents noted that the district is really trying to be transparent, with all information posted on the district's website, including a continually updated frequently asked questions page. These two respondents also noted the important role of having representation from every school building on the district committee, so that after each meeting the representatives can bring a slide presentation and talking points back to their buildings to present to staff there. One respondent also noted that this is in addition to including all parties—teachers, principals, district administrators, PTA/PTO members, and community stakeholders—on the committee, so that all can be involved in and aware of the process.

Another respondent noted that the district also regularly presents updates to the school board about its progress and that a district newsletter is sent to all staff after each school board meeting, so this material gets to teachers as well. This respondent noted that the teachers association also publishes a monthly newsletter that almost always includes information about progress. This respondent believes that they are continually talking about the new system because they do not want any surprises. Another respondent commented on the power of the conversations of the 60-person committee and the importance of communicating the tool-creating process to staff.

In addition to these formal channels for communication, one respondent noted that during the yearlong pilot in 2011–12, even teachers who are not randomly selected to participate will be able to observe what is taking place and discuss the new system. All in all, the breadth and depth of communication thus far was impressive.

Principal Evaluation

Principal evaluation in Central Valley is being developed in much the same way as for teachers.

Setting the Evaluation Framework and Goals

All interview participants reported that the principal evaluation system was linked to a growth model, with two respondents noting that principals in Central Valley are guided by the 21

responsibilities around balanced leadership, which the district has spent three years working on with McREL. One respondent noted that this model has been coupled with a document provided by the Association of Washington School Principals (AWSP). However, the committee has not adopted a clearly defined leadership evaluation framework. Respondents agreed that, as for teachers, the goal of the principal evaluation system is to improve practice. Two respondents cited self-reflection as a goal. Encouraging goal setting and outlining what quality leadership looks like also were mentioned as goals for the new system.

Establishing Evaluation Criteria

The teacher and principal evaluation systems have been developed simultaneously by the same diverse 60-person task force and using the same process of developing descriptors for the eight criteria handed down by the state. One respondent noted that the committee looked comprehensively at the descriptors using exactly the same process as for teachers. Another respondent noted that, for principals, the committee was able to rely more heavily on the document provided by the AWSP when developing the descriptors.

The committee has developed a set of evaluator and self-assessment rubrics that include major domains of effective principal practice and generally align with state and national (Interstate School Leaders Licensure Consortium [ISLLC]) standards and priorities. However, definitions of key terms and criteria levels are not yet available, and criteria include high-inference language that may challenge rater reliability. The committee intends to address these matters through upcoming training and professional development.

Principals in Central Valley will be rated on the same four-point scale as teachers:

- Level 1: Unsatisfactory/Not Demonstrated
- Level 2: Developing
- Level 3: Proficient
- Level 4: Accomplished

There is a list of indicators within each level. The levels build upon each other, so the *Accomplished* principal shows evidence of doing all things that a *Proficient* principal does and more. The Central Valley School District rubric appears applicable to principals at different stages of their career and is free from bias. However, the "cut points" between levels are not well-defined, with inconsistencies across domains, particularly the *Proficient* and *Accomplished* levels. It appears that the difference between the *Developing* and *Proficient* domains is the difference between thinking/believing/articulating and doing. However, this difference appeared to the reviewer to be inconsistent across domains.

Frequency of Evaluation

The committee is not far enough along in its process to have determined the frequency of formal and informal evaluation processes for principals, which will be a determining factor in the validity of the principal evaluation system.

Measuring Performance

One respondent noted that the same logic regarding the process of measuring performance would exist for principals as for teachers—namely, rather than having principals go to the district office, evaluators and principals will meet in the principals' buildings where any artifacts or evidence would be readily available.

One challenge may be that, although some of the domains include observable and measurable verbs and other language, many do not. For example, "Fosters school and community commitment to the principles of sustainability and continuous improvement" can be interpreted many ways. This type of language could create a burden and ambiguity for evaluators and self-assessors.

The committee has begun to identify data sources for the criteria but has not identified specific surveys or other measures. One respondent noted that there is more direct language from the state about improving student test scores in the principal evaluations as compared with teacher evaluations. How Central Valley interprets this language will be part of the pilot. The respondent believes that evaluators will want principals to demonstrate an awareness of where their test scores are, what trends appear, and how the principal is responding to these trends. However, the respondent further noted that this would not be a change from current practice.

The committee was not sufficiently far along in its principal evaluation system development to provide evidence of a data-tracking process being developed.

Training for Evaluators

The training of evaluators of principals, as well as the criteria for selecting and evaluating evaluators, have not yet been determined, but one respondent reported that the individuals evaluating principals also will attend the March training that principals attend to learn more about their own tool. (See the Training for Evaluators section in the above description of the teacher evaluation system for more details.)

Using Results for Professional Development

The new evaluation system is intended to be used primarily for principals' professional development as well as teachers' professional development. Principals currently receive monthly professional development, which is expected to continue under the new system. One respondent noted that all professional development under the new evaluation system will have to be tied to the district's Balanced Leadership 21 responsibilities for leaders. It is intended that professional development will be individualized to meet each principal's growth needs, and then grouped to provide learning opportunities identified as needed for principals across the district. No clear links, however, have yet been drawn to show how evaluations will inform professional development. Nor have elements of a comprehensive human capital management strategy (e.g., a strategy for addressing and aligning all principal quality policies—hiring, induction, compensation, etc.) been developed in connection with evaluation results.

Communicating Results and Ensuring Stakeholder Investment

Interview respondents noted that about 50 principals and assistant principals are participating in the pilot this spring, with all of them expected to be evaluated under the new model during the 2011–12 pilot year. McREL then will assist in collecting data to assess the principal evaluation pilot results.

The committee provided no documentary evidence that a communications and stakeholder investment plan was completed or under development. However, interview participants noted that reforms will be communicated to principals in the same way as is described for teachers. One difference is that there is no negotiated agreement with principals. Principals are on a somewhat faster timeline in that they all are supposed to be piloting the new system in 2011–12.

Summary and Conclusions

Interview participants believed that the new evaluation system will lead to improvements in the quality of teaching that takes place in Central Valley. Two respondents believed one of the greatest strengths to the new system was creating clear, high-quality definitions of what high-quality teachers and leaders look like in their context. This was considered by one respondent to be an eye-opening experience and by another respondent a good way to show what is expected of people who work in the district. Two respondents also cited as major strengths the types of powerful professional conversations they believe will take place under the new system and the constructive feedback that educators will receive as a result. One respondent—who believed the new system's biggest strength is that it will allow teachers and leaders to be reflective—described how it also will take some time and education for people to get used to not being told their practice is perfect, but that this is a good thing. A final strength in the view of one respondent was that the system is being locally developed by teachers and principals in the district with the voice of parents as well.

Alongside these opportunities, the interview participants acknowledged that there will be challenges. Two respondents noted that their greatest concern is that this new system will consume much more time than evaluation does currently. This is the case both for teachers and for principals. The additional time required for conversations, reflection, feedback, goal setting, sharing, and visits was viewed as a challenge. A second challenge mentioned by one respondent was providing the staff development needed to help teachers struggle through the change from a binary model to a four-tiered model. Another respondent cited as a third challenge providing the professional development that teachers and principals identify as needed to improve on their areas of weakness, in light of limited resources. A fourth and final concern mentioned by one respondent was that the district might not be able to use the tool it spent time and energy developing if OSPI mandates that a different one be used statewide.

The interview participants felt grateful to have the opportunity to participate in developing this new tool for the district. One respondent expressed that it will be a significant change from the past to have an evaluation tool that has meaning and substance. Another respondent expressed concern for other districts that do not have the opportunity to develop their own tools, believing that the conversations that went into the development of the Central Valley tool lay behind its power and the excitement surrounding its use.

Kennewick School District
Case Study Report

District Overview/Description

The Kennewick School District, located in the southeastern part of Washington state, serves almost 16,000 students in its 15 elementary schools, four middle schools, and nine high schools. Its student demographics are as follows:

- 0.88 percent American Indian or Alaskan Native
- 2.49 percent Asian or Pacific Islander
- 2.22 percent African American
- 29.71 percent Hispanic
- 60.95 percent white
- 48.14 percent eligible for free or reduced-price lunch

The district employs more than 800 teachers, 68.6 percent of whom have at least a master's degree.

Case Study Method

This study was undertaken to help the Washington Office of the Superintendent of Public Instruction (OSPI) better understand how to support districts' implementation of Senate Bill (S.B.) 6696, which establishes new criteria for evaluating educators. For this component of the study, four individuals affiliated with Kennewick Public Schools were interviewed in an effort to collect information concerning the process of developing and implementing their new teacher and principal evaluation systems. Each of the respondents has played an active role in the development of the new systems. In an attempt to gain a diverse range of institutional perspectives, we interviewed two district-level administrators and two teachers. All responses were coded based on topic and related through the following case study.

Moreover, the district was requested to submit pertinent documents concerning the development and implementation of its teacher and principal evaluation systems. These documents were then reviewed by AIR staff who documented evidence of the district's actions.

Teacher Evaluation

Setting the Evaluation Framework and Goals

Prior to its participation in the Teacher-Principal Evaluation Pilot (TPEP), the Kennewick district had already consulted with the Harvard Institute for School Leadership to develop an instructional framework called Purpose, Engagement, Rigor and Results (PERR). Parts of the instructional framework are based on Marzano criteria, and Fred Jones provided the foundation

for the classroom management dimension. Over the last eight to ten years, the district has been designing and honing specific rubrics to support the PERR criteria for effective teaching, even going so far as to develop videos that exhibit exemplary instructional practices according to each rubric. Throughout the process of designing a new evaluation system to pilot in the 2011–12 school year, the district has been working with Charlotte Danielson to align this existing instructional framework and its teaching standards to the requirements set forth in Senate Bill 6696. The district has adapted some of its own language to more closely resemble the rubrics in the Danielson framework and also has heeded her advice to move certain skills into different criteria categories. The evaluation committee is still working on adding rubrics and evidence indicators to all eight of the state-mandated criteria.

The document review, conducted later in the district's planning process, found that the language in the rubrics is clear and connects to the four components of the PERR framework. The rubrics define both teacher and student indicators and also provide guiding questions for evaluators. However, some of the documents that reference the rubrics, such as the Teaching and Learning Task Force "Talking Points" document, use vague terms such as Essential Academic Learning Requirements and Grade-Level Expectations without explicitly explaining how they align with the rubrics.

The district has not yet clearly defined how the final version of the standards will be linked to the evaluation process. The current language of the teachers union contract does not align instructional standards with the teacher evaluation system. However, there are plans to draft revised contract language in the spring to reflect any changes required to pilot the new evaluation system in the 2011–12 school year. Negotiations are expected to take place with the union over the summer. Two interview participants mentioned that there have been some discussions about tying the evaluation system to transfer assignment language in the union contract, but no other employment or compensation decisions. The document review did not find any documentation indicating that teacher performance will be linked to student achievement. It also found no evidence that teacher tenure or dismissal would be based on evaluation results.

All four interviewees indicated that the overarching goal of the new evaluation system is to promote professional growth, improve how teachers deliver instruction, and thereby improve student performance. A related goal that one respondent mentioned is to ensure that the criteria and rubrics of the evaluation system are clear, understandable, and transparent for all faculty and administrators, which will help focus conversations between principals and teachers about specific areas for professional growth.

Establishing Evaluation Criteria

The state mandates that teachers receive ratings on a four-point scale for each of the eight criteria established in S.B. 6696. The Kennewick district decided that a rating of Level 1 on a particular skill represents unsatisfactory performance. A teacher who receives a Level 2 rating is exhibiting basic competency in performing a skill but still needs to make significant improvements. A rating of Level 3 indicates that a teacher is demonstrating effective practices, and Level 4 characterizes exceptional performance. Schools will be required to report the

distribution of ratings across teachers, and fewer than 10 percent of teachers are expected to be rated overall as a Level 4.

According to the interview respondents, the Kennewick district has defined rubrics for the eight criteria but is still in the process of choosing evidence indicators on a four-point scale. One respondent mentioned that for several of the criteria, the committee has found it easier to define a rating of "1" or "4" than to distinguish between Levels 2 and 3. That respondent stated:

> I think it's hard … to develop the 2 and the 3. … Sometimes I don't think there's a big enough gap between each of the four standards and … it's going to be pretty subjective for the administrator using the teacher evaluation.

In addition to defining the evidence indicators, the district is also still debating what the consequences will be for teachers who continue to receive lower ratings. For example, the committee has not yet reached consensus on how long a teacher would be allowed to remain at a Level 2 rating without showing enough improvement to advance to a Level 3. One respondent felt that the guidelines should account for certain circumstances in which a teacher might be expected to remain at a Level 2 for a longer period of time—for example, if a teacher had to switch from teaching fifth grade to teaching kindergarten and learn all new curriculum. The district is still grappling with how to incorporate all of these possible scenarios into the evaluation system guidelines.

Even though the rubrics are still in development, the district does plan to evaluate teachers in a variety of subjects using the new criteria in the 2011–12 pilot, including those who teach subjects such as music, physical education, and vocational education. However, one of the goals of the pilot is to determine whether the criteria and rubrics "apply equally to a math teacher or social studies teacher as they do to an auto mechanics teacher or a PE teacher." One interviewee thought that the state would likely give districts more direction on whether noncore teachers need to be evaluated after a pilot program is adopted statewide. Additional support staff such as counselors, librarians, and speech language pathologists will not be evaluated under the new evaluation system.

State law and union contract language currently dictate how frequently teachers are observed as part of the evaluation system. State law stipulates that provisional employees receive more frequent observations than continuing contract teachers. S.B. 6696 did not make any changes to those requirements, but it did increase the length of time that teachers are under provisional status from two years to three years. There is an understanding between the Kennewick district and the local union that language regarding the frequency of evaluations will be negotiated over the summer in order to successfully implement the pilot next year. However, interview participants did not mention a target or goal for how much more frequently teachers may be evaluated in the new evaluation system.

The document review found that the summative rubric captures what the district believes to be the major components of effective teaching. The district has not yet determined how frequently teachers will be formally evaluated. Based on their previous evaluation rating, teachers will be placed in specific evaluation "cycles"—either a professional growth cycle, observation cycle,

plan of improvement cycle, or probationary cycle. No information about the frequency of evaluations was included for each probationary cycle.

Measuring Performance

Evaluation committee members have not settled on which measures will be used to evaluate teachers, but they are certain that at least one measure will be observations. One respondent acknowledged that the breadth of areas covered by the state-mandated criteria would likely require multiple measures of effectiveness to collect an adequate amount of information for thorough evaluations. Two respondents expressed hesitation about using value-added scores in teacher evaluations because of concerns about the measure's inconsistency from year to year. If the district were to use value-added data, it would have to be in a way that was "fair to teachers" and that "[accounted] for the differences in starting points of the students." In addition, one respondent thought that if value-added scores were tied to compensation or job security, there might be unintended consequences, such as teachers shying away from working in low-income schools or schools with high levels of student mobility.

Two respondents also thought that it was unlikely that the district would use extensive portfolios to evaluate teachers. These respondents mentioned that portfolios might make the evaluation process too time consuming and cumbersome for teachers and administrators. One respondent described some of the tradeoffs that the evaluation committee is considering between the thoroughness of different measures and the time required to prepare those measures for evaluations:

> I want [teachers] to be able to defend [their] practice, but at the same time, I don't want [teachers] to … have to put together [a] 5-inch portfolio. … It's hard for an administrator to come in for 30 minutes and then say, "You're a number 4" or "You're a 3 or a 2." … [Therefore] I think it has to be a balance. … We don't want [the evaluation system] … to replace what [teachers are] supposed to be doing in the classroom, spending more time on getting ready for [their] evaluation than [they] are preparing for [their] own students.

Some measures that the committee is considering using include classroom artifacts such as lesson plans, grade books, and teachers' communication tools with parents. Two respondents mentioned that the committee is perhaps struggling the most with finding an objective measure for the criterion that assesses a teacher's ability to collaborate with peers.

The document review found that teachers have the option of providing additional measures of effectiveness to the evaluator that go above and beyond the evaluation rubrics. However, the district does not specify what types of artifacts are appropriate to submit, and no rubrics have been developed to evaluate those additional measures. This omission leaves teachers without any guidance on what to present to the district or how to align those additional measures with student growth. As a result, there may be a variety of supporting evidence provided by teachers, which threatens uniformity and presents coordination challenges.

Kennewick has just begun to discuss the possibility of assigning different weights to the various criteria of teacher effectiveness (e.g., assigning more weight to classroom management than classroom instruction), but no final decisions have been made. Correspondingly, the document review did not find any documentation regarding how to weight rubrics differently.

Training for Evaluators

Although the district is still in the very early stages of planning for evaluator training, two respondents emphasized the critical role that the training will play in successfully implementing the new evaluation system. One respondent stated:

> We talked about the need for … commonality … so that when one administrator is looking at a teacher's performance, another administrator would also see the performance in the same way. We know that extensive training is going to be involved.

The steering committee has discussed the possibility that the training will require observers to code videos of a lesson to check for interrater reliability. These sessions would emphasize the importance of achieving consistency across evaluations and understanding "the intent of the language in the rubrics" and the "kind of evidence that supports [each of the rubrics]."

Under the current PERR evaluation system, the district uses Instructional Conferences—meetings that the district holds with its principals four times a year—to train administrators on how to evaluate teachers consistently within each of the four areas of the framework. The district intends to expand these training efforts under the new evaluation system to include the broader additional evaluation criteria. Similarly, the district plans to use Instructional Academies—its equivalent meetings with teachers that occur three or four times a year—to clearly communicate to teachers the specifics of the new evaluation system. These training sessions also will likely involve watching videos of examples of exceptional teaching techniques that reflect the new evaluation system criteria.

In order to help ensure interrater reliability, the district also is considering assigning a second principal from another school to conduct an additional round of evaluations of a school's teachers, which would allow the district to compare those results with the actual principal's scoring. However, this type of arrangement would have to be negotiated as part of the union contract. The ultimate goal is to assuage any fears that teachers may have that administrators are subject to bias in their assessments or may not "know what good teaching looks like," as one respondent stated.

At the time that the document review was conducted, the district did not have any documents describing the evaluator training available for review.

Using Results for Professional Development

Although all interview participants stated that promoting teachers' professional growth was the primary goal of the new evaluation system, the district has not yet specified how the evaluation results will inform professional development practices. One respondent explained that clearly defining the rubrics will help ensure that "a principal can sit down with a teacher and look at the rubric and say, 'Where are you now?' And then, 'What's the next step in your development as a professional?'" Another respondent mentioned that there is one principal in the Kennewick district who has previous experience with aligning the rubrics of the Danielson framework with teacher professional development, so the steering committee will likely utilize that principal's

expertise, along with advice from Danielson herself, in designing this piece of the evaluation system.

The document review found evidence that teachers will be placed in different evaluation "cycles" (professional growth, observation, plan of improvement, or probationary) depending on their previous summative rating. In the professional growth cycle, teachers will create a plan for professional growth with their supervisors, which will serve as the foundation for both an interim check-in and an end-of-year summative evaluation. Teachers placed in the improvement cycle will complete a Planning Conference Worksheet and Plan of Improvement Worksheet. The district is still finalizing documents that describe the observation cycle and the probationary cycle.

The Kennewick district already has a system in place to train principals on analyzing and interpreting student data to inform instructional practices. The district holds regular training sessions in which principals are taught how to use various spreadsheet functions to analyze student-growth data in different ways. Principals also are assigned district administrators to act as their "learning partners," who then offer extra feedback and support on data issues. For example, the learning partners recently helped their principals analyze their students' fall and winter Measures of Academic Progress (MAP) test scores. According to respondents, these practices will continue in the new evaluation system and help link the results of evaluations with the design of professional development programs.

The document review did not find any information about how the Kennewick district intends to collect and store data.

Communicating Results

One of the goals that the district set forth as a priority early in the planning process was to make the committee's decisions as transparent as possible for relevant stakeholders. The committee recently created a general slide presentation describing the district's progress in designing the pilot, which the Education Association (EA) president then presented to all building administrators. The EA president also gave a joint presentation with the district lead to the Kennewick Education Association's teacher representatives and all the principals. The district intends to continue creating and delivering slide presentations throughout the planning process. In addition, the association president includes updates about the committee's progress in weekly newsletters called *Quick Notes*, which are sent to members. Despite these efforts to communicate with stakeholders, one respondent questioned whether some teachers would have been made aware of the committee's progress if they had not attended association meetings or their principals had not updated them.

Although the document review found the slide presentation developed for stakeholders, it found no evidence of communication or engagement with the larger community (e.g., parents). On the other hand, the district has made efforts to solicit feedback from teachers. The district's Rubric Feedback Loop document states that there are plans to establish a specific time frame for teachers to review the instruments and provide feedback, which the district then will incorporate into the design of the instrument.

The timeline for the district going forward will be to continue tweaking the rubrics and choosing evidence indicators for each of the different rating levels over the next few months. In March and April of 2011, the district will review the rubrics with principals at their administrator meetings and with teachers at their building meetings. In early May, the district plans to hold informational meetings with potential pilot participants, whose applications will be due in mid-May. In June, the district will solicit feedback from teachers and principals and incorporate any modifications into the rubrics. Negotiations to change union contract language will occur over the summer, with evaluator training tentatively slotted for August 2011. The pilot is scheduled to be implemented in the 2011–12 school year, followed by full implementation throughout the district in the 2012–13 school year.

The document review surfaced plans to hold an informational meeting about the rubrics and pilot process on May 2. On May 10, the district will notify the pilot sites, which will have 40 voluntary participants. Teachers and principals taking part in the pilot will be required to meet periodically throughout the year to provide feedback to the steering committee.

Principal Evaluation

Setting the Evaluation Framework and Goals

To help redesign its principal evaluation system, the Kennewick district is referring to the Association of Washington School Principals' model for an effective principal or school leader. The document review also found evidence that the district is aligning its rubrics to evaluate principals with the Interstate School Leadership Licensure Consortium (ISLLC) standards.

Establishing Evaluation Criteria

Similar to the requirements for the teacher evaluation system, the state provided eight criteria on which districts must evaluate principals, but the district can choose the specific indicators and rubrics on a four-point scale. Similar to the teacher evaluation system, the district determined that a Level 1 rating for a principal reflects unsatisfactory performance; Level 2 means that the principal is exhibiting basic competency but still needs to make significant improvements; Level 3 represents an effective demonstration of that skill; and Level 4 characterizes exceptional performance.

According to the interview participants, the district is still in the process of determining specific benchmarks and evidence indicators for all eight of the criteria. The document review, which was conducted later in the planning process, found evidence that the district had developed a detailed and comprehensive four-point rubric scale to evaluate principals on all eight of the criteria, including examples of evidence indicators. The district has not yet decided how frequently principals will be evaluated under the new system, and the document review did not find any reference to how frequently evaluations will occur.

Measuring Performance

According to the respondents, the district has not settled on which measures it will use to evaluate principals. Respondents stated that there are several indicators in the criteria, such as

"has a well-thought-out school safety and climate plan," that present challenges for designing an objective measure of effectiveness. The committee has been considering requiring principals to submit a portfolio with different artifacts; however, no final determinations have been made.

Student-growth measures are already an integral part of the principal evaluation system in the Kennewick district. This evaluation measure even played a role in determining principal compensation as part of a recently revoked performance pay system in which 10 percent of a principal's salary was based on goals for student achievement. Principals are still required to report their levels of academic attainment to the school board every year, and the evaluation process assesses whether goals for student growth were achieved, but the measure is no longer used to make compensation decisions. Two of the respondents suggested that the evaluation process for principals in the new system may not differ too markedly from the process that is already in place.

The document review indicates that the district has developed a list of several types of measures that principals can include as part of the evaluation, but did not find any specific procedure to weight the different measures for a summative rating or score.

Training for Evaluators

The district is in the very early stages of discussing evaluator training for the principal evaluation system. Currently, only two people evaluate principals in the Kennewick district: the Assistant Superintendents of Secondary and Elementary Education. The district has not yet discussed how it plans on changing this process in the new evaluation system. Correspondingly, the document review did not find any documentation detailing who the evaluators will be and what type of training they will receive.

Using Results for Professional Development

Similar to its documentation regarding the teacher evaluation system, the district provided documents stating that administrators will be placed in one of four evaluation cycles (professional, observation, plan of improvement, or probationary) based on their rating in the previous year's evaluation. The document review did not find any information on how evaluation data will be collected and stored.

Communicating Results

The district developed and has disseminated a slide presentation detailing the planned changes for its principal evaluation system and also has distributed feedback forms to principals and certified staff to gauge their reactions and answer any questions they may have. The district has not yet communicated the upcoming changes to parents and other community stakeholders but does plan to do so. The pilot will begin in August 2011 with 40 teachers and administrators.

Summary and Conclusions

Overall, all four interview participants expressed enthusiasm about designing and implementing the pilot for the new evaluation system, especially the goal of improving opportunities for

teacher and principal professional growth. One respondent's remarks aptly characterize the sentiments similarly expressed by the others:

> The current evaluation system of the state is old and needs to be replaced. And we think this is going to be better for us. It's going to be a little bit of a challenge to turn it to the other side. But, in the end, when we get to that other side, the evaluation system will be viewed, hopefully, entirely differently by teachers. And we can use it in a way to help people become better at what they do as opposed to just saying, "Here's your annual review; sign here."

Another strength that the respondents mentioned about their pilot program is that the district and union are working together to revise the evaluation system, united by the desire to improve teacher and principal professional growth opportunities. One respondent suggested that in the past, the district and the union had struggled to have meaningful conversations about "[moving] forward from sort of the compliance-type evaluation process to one that's more professional growth." However, when presented with the opportunity to participate in this pilot program, the two organizations actually wrote the proposal for the state together and continue to flesh out the design of the evaluation system jointly.

At the same time, several concerns also were noted. All four respondents indicated that it will be a significant challenge to ensure that the rubrics are absolutely clear to evaluators and teachers, and that ratings are consistent across schools. All respondents also were apprehensive about the impending paradigm shift required to evolve from a "Satisfactory or Unsatisfactory" rating system to a four-tiered system, in which fewer than 10 percent of teachers are expected to receive a Level 4 rating. One respondent explained:

> Everybody thinks they're a [Level] 4. Culturally, that's going to be a big change for us. Now, you're either satisfactory or unsatisfactory, and to move to that scale is going to be a big cultural shift.

Finally, one respondent noted that the state seems to be shifting directions in its requirements for districts, and that committee members were wondering if the state may soon provide more specific definitions for the rubrics. As a result, the respondent was worried that the work they had done so far on the rubrics would not end up being implemented in the pilot.

Despite the challenges already encountered and anticipated in the coming months, the district is continuing to move forward with designing its pilot, and the respondents believe that the end results will be worth the effort.

NorthEast Washington Educational Service District 101 Consortium Case Study Report

District Overview/Description

The NorthEast Washington Educational Service District (NEWESD) 101 Consortium is composed of eight rural districts working collaboratively to design a new principal and teacher evaluation system that will be used jointly across all participating districts. Each district varies in student demographics, number of staff, and number of schools. Table 1 presents each of the districts participating in the consortium, along with the number of schools in each district.

Table 1. Districts Participating in the Educational Service District 101 Consortium

District	PK–3	PK–5	PK–6	PK–8	K–9	K–6	K–8	4–6	6–8	7–8	7–12	9–12
Almira							1					
Davenport			1								1	
Liberty				1								1
Medical Lake	1			1				1		1		2
Pullman		3							1			1
Reardon-Edwall					1							1
Ritzville							1					1
Wilbur						1					1	

Case Study Method

This study was undertaken to help the Washington Office of the Superintendent of Public Instruction (OSPI) better understand how to support districts' implementation of Senate Bill 6696, which establishes new criteria for evaluating educators.

Focus Groups. For this component of the study, three focus groups were conducted on February 11, 2011, at the educational service district with superintendents, principals, and teachers who are developing new teacher and principal evaluation systems in the consortium. Each focus group lasted 1–1½ hours, and all groups were audiotaped and transcribed. Transcripts were then inductively coded to capture the overall themes around (a) how the performance evaluation systems were being developed and (2) participants' experience with the process.

Document Review. The consortium submitted its draft evaluator and principal rubrics to the American Institutes for Research (AIR) to receive formative feedback on ways to refine the rubrics.

Teacher Evaluation

Setting the Evaluation Framework and Goals

Focus Group Feedback. Before the districts' collaboration as a consortium, each district, prior to the passage of Senate Bill 6696, evaluated teachers' performance annually using either a short or long form that provided ratings of satisfactory/unsatisfactory on a number of indicators. In the majority of districts participating in the consortium, teachers' performance evaluation was not linked to a broader district or school instructional framework. In addition to completing the short or long form, principals were required to conduct classroom observations twice a year: 60 minutes using the long form and 30 minutes using the short form. While each district minimally used either the short or long form to guide the performance evaluation, in a few of the districts, evaluators went above and beyond the requirements of the form by providing teachers with detailed feedback based on their classroom observations.

Most teachers participating in the consortium indicated that they were evaluated annually based on the minimum requirements mandated by the state. One teacher remarked about the current evaluation system, "Our previous principal did the same statements for all of the teachers and just changed the name. Then sometimes [the principal] would forget to change the pronoun from she to he." Another teacher reported having a previous administrator who "put the evaluation inside my box to sign, and it was for a date where I was not present at school."

Document Review. The consortium did not provide evidence related to frequency of evaluation. Considerations need to be made for educators at different points in their career. For instance, will the number of evaluations differ for a first-year teacher and a 10-year veteran? It is recommended that all teachers be evaluated at least yearly, but the consortium should determine what, if any, conditions may exist under which the number of observations per teacher may differ. In addition, the capacity of evaluators to complete the number of required observations is another issue the consortium will need to address. An analysis of the number of observations currently required across schools in the consortium should be conducted and compared with how the number may change under the new evaluation system. A broad stakeholder group, including evaluators, needs to look closely at the data to determine whether or not evaluators will have the capacity to complete the number of required observations. In addition, plans should be made to potentially identify additional staff who may be able to provide support in the event that an evaluator is unable to realistically complete the number of required observations. Lastly, timelines and requirements related to required feedback also should be considered. If a teacher is observed, what is a reasonable amount of time to provide meaningful feedback to the teacher, keeping in mind that timely and constructive feedback will help the teacher grow professionally and potentially impact student achievement? Does the collective bargaining agreement for districts include language about timelines and deadlines as they relate to teacher evaluations?

Establishing Evaluation Criteria and Performance Measurement

Focus Group Feedback. Teachers' previous experiences with their performance evaluation have influenced the ways in which they are working to design the new system. As of February 2011, the consortium was using Charlotte Danielson's instructional framework to systematically

guide the consortium's new system and to align it with the state criterion, which are presented below:

- Centering instruction on high expectations for student achievement.
- Demonstrating effective teaching practices.
- Recognizing individual student learning needs and developing strategies to address those needs.
- Providing clear and intentional focus on subject matter content and curriculum.
- Fostering and managing a safe, positive learning environment.
- Using multiple student data elements to modify instruction and improve student learning.
- Communicating and collaborating with parents and school community.
- Exhibiting collaborative and collegial practices focused on improving instructional practice and student learning.

In addition to Danielson's model, the consortium also consulted several models already in place, including a model from North Carolina. The consortium participants are aware that the new performance evaluation system needs to be tied to an instructional framework; however, as of February 2011, they were still in the process of refining the details of the teacher rubric.

To design the teacher performance evaluation, consortium participants met monthly, starting in August 2010. The purpose of these meetings began with (a) trying to gain a clearer understanding of the state's requirements; (b) researching successful models that have already been in place in other states, districts, or schools; and (c) using the established models to inform their own design.

After looking at established models, the consortium decided to establish a four-tiered framework, with the following ratings: 1, 2, 3, 4, along with key phrases to describe each tier (such as *routinely*). A percentage of certified staff, including continuing, provisional, core, and noncore teachers, will be evaluated using the established framework.

To do this, each of the eight districts in the consortium took one of the eight criteria designated by the state and worked as a small team to develop indicators and ratings within that criterion. For example, one district took *Centering instruction on high expectations for student achievement* and drafted the performance indicators and ratings within that criterion. After all districts drafted their indicators, their work was consolidated into one master draft and reviewed by all participants in the consortium.

At the time the focus groups were held, the consortium was just beginning to engage in a broader conversation about how teacher performance would be measured. Participants in the focus groups indicated that they were looking at using classroom observations, with some discussion around using portfolios.

Document Review. Upon review of evaluation format, it appears that the consortium has carefully considered the summative rubric for teacher evaluation (Final Report Out and Steering Committee Recommendations). However, the rubric descriptors are focused mainly on teacher actions, and only in Tier 4 are students referenced. The consortium may want to consider revisiting the descriptors to include more evidence of student learning. The cut line has been identified between Tiers 1 and 2, which will likely result in a wider distribution of teachers across the four tiers. The consortium participants have chosen not to name the tiers (i.e., excellent, satisfactory, needs improvement, ineffective). Instead, they feel strongly that a numbered system, which includes key phrases describing each tier, would be more effective. Additional detail surrounding this decision was not available within the provided documents.

It is not clear from the information provided whether the Consortium plans to differentiate the format of the evaluation depending upon a teacher's career phase. Ensuring that the system accounts for this will be important and will send a clear message that the consortium understands that, as with students, teachers too come into the profession with different levels of readiness, which should be accounted for within the evaluation tool. Evidence for multiple measures was limited. Documents provided related to multiple measures (Methods of Collecting Data for the Teacher Evaluation Criteria and Teacher Evaluation Matrix) had not been completed. In some documents, it was clear that specific districts within the consortium had considered multiple measures in their planning (Liberty, Reardan, and Ritzville). There was no evidence related to weighting of measures for the teacher evaluation system.

The consortium has attempted to explicitly define indicators under each of the evaluation criterion to ensure a more objective approach to the evaluation. As the consortium moves forward, it would be helpful to carefully review each of the indicators to ensure there is no redundancy within the evaluation criteria (Revised MASTER Teacher Criteria Measures Jan. 21). For example, *Pacing* is referenced in criterion two and then *Effective use of time* is listed as an indicator in criterion three. In the current draft, it is unclear how these indicators differ. This would be an opportunity either to condense the two into one or to define more explicitly each of the indicators so that evaluators do not misinterpret how they should be capturing evidence. Criterion three provides another example of potential confusion for the evaluator. *Interventions for all students* and *Adaptation of instruction* both reference differentiation or modifying instruction based on individual needs in addition to ensuring legally required accommodations. It may be necessary to define more explicitly these indicators or collapse them into one that covers both differentiation and compliance.

Most of the indicators as identified within the rubric can be reliably measured. The consortium has not yet explicitly identified the multiple measures that will be used within its teacher evaluation system. As a next step the consortium should identify measures by reviewing the criteria and indicators to determine the evidence necessary to effectively evaluate them as identified within the consortium's rubric.

Training for Evaluators

In February, the consortium had not yet reached the point of determining where, how, or when evaluators would be trained. Consortium participants did, however, engage in a discussion

around the need for evaluators across the state to be calibrated in order to ensure that the ratings they assigned teachers were fair and equitable statewide.

Using Results for Professional Development

When asked about whether the results from the teacher performance evaluation would be used to inform professional development needs, all participants expressed that as the ultimate goal. However, many expressed concern around how professional development would be funded, and strongly suggested that if the state was mandating a new performance evaluation system that it also should provide funding to ensure that teachers receive professional development tied to their needs.

Ensuring Effective Communication and Ensuring Stakeholder Investment

Focus Group Feedback. Consortium participants indicated that they plan to begin piloting the teacher evaluation system in fall 2011. With implementation scheduled for the fall, they would like professional development on using the tools to begin in summer 2011; however, the specific times and dates have yet to be established.

Document Review. Upon review of documentation submitted by the consortium, there was limited evidence of how the consortium ensured that communication and stakeholder involvement were clearly integrated into the design of the evaluation system. Stakeholder engagement evidence was limited, but it appears that several meetings were held for those interested learning about the evaluation system and offering feedback on progress made thus far (*Educator Forum*; *ERNN Conference Jan 2; NEWASA presentation Dec. 2*). However, it is unclear whether or how the feedback received from stakeholders was considered by the consortium evaluation committee. The consortium communicated with stakeholders using several different methods including memos, meetings, and forums (Eval Pilot Announcement; NEWASA presentation Feb. 17). In addition, one district included evidence of a survey sent to all staff members (e-mail to all certified teachers re: Teacher Evaluation System). It is not clear whether the community fully understands the importance of an improved teacher evaluation system as the provided documents did not include evidence of community involvement.

Principal Evaluation

Setting the Evaluation Framework and Goals

Focus Group Feedback. The consortium approached designing the principal evaluation system by first ensuring that it aligned with the components of the teacher evaluation system, and that the language in both systems was clear. The consortium began by asking, “What is good performance for a principal?” and establishing criterion around those indicators. The specifics of how and when principals would be evaluated was yet to be established.

Document Review. There was no evidence related to the frequency of the principal evaluation within the consortium documentation. Decisions related to frequency of evaluation should take into consideration a number of factors, including the principal’s stage in his or her career. For

instance, a new principal most likely will require additional support in the first few years. This may come in the form of a mentor or consulting principal who can provide guidance and support in a structured way related to specific growth areas. For a veteran principal, frequent evaluations are still necessary, but will likely look somewhat different from those of a new principal.

The consortium also will need to consider the specific requirements and timelines for evaluators to provide feedback to principals. As with any feedback, it should be timely, consistent, and constructive. Evaluators should be interacting with the principals on a frequent basis, whether to observe the principal in action or to analyze specific data related to a particular evaluation competency. The process also should take into consideration the frequency of observations or visits and the time it will take the evaluator to prepare and engage in a meaningful conversation with the principal.

Establishing Evaluation Criteria and Measuring Performance

Focus Group Feedback. To define the evaluation criteria that influenced the principal evaluation, the consortium began by reviewing the criterion established by the state, which are as follows:

- Creating a school culture that promotes the ongoing improvement of learning and teaching for students and staff.
- Demonstrating commitment to closing the achievement gap.
- Providing for school safety.
- Leading the development, implementation, and evaluation of a data-driven plan for increasing student achievement.
- Assisting instructional staff with the alignment of curriculum, instruction, and assessment with state and local district learning goals.
- Monitoring, assisting, and evaluating effective instruction and assessment practices.
- Managing both staff and fiscal resources to support student achievement and legal responsibilities.
- Partnering with the local school community to promote student learning.

After the criteria were reviewed, the consortium then discussed the principal evaluation model that was already established by the Association of Washington School Principals (AWSP). Because this AWSP resource was already provided, the consortium used that document as a guide. The criteria by which principals will be evaluated include school safety, managing resources, communication, involvement in the community, and staff development. *How* these criteria will be evaluated and measured is still under development.

Document Review. The consortium has worked to create a summative rubric that addresses each of the eight criteria set forth by the state (Revised MASTER Principal Criteria and Rubrics Feb 14). The rubric aligns well with the criteria set forth by the state but does not always clearly articulate expectations across the competency levels. For example, within criterion one, the

indicator of *Difficult issues* (page 2), there is not a clear distinction between Tiers 3 and 4, which could cause confusion when an evaluator is scoring. The consortium should continue to revise the rubrics to ensure that the examples of each tier are explicit and consistent.

The cut line for the principal evaluation system is between one and two. The consortium should work to ensure that the established line will clearly distinguish between those principals who are ineffective and those who may be developing. The consortium may want to consider following the same process utilized when establishing the cut line for teachers, which is set between Tiers 1 and 2 to allow for distribution of ratings across the tiers. In addition, the consortium will need to determine how it will accommodate for principals at different phases of their career. Will the frequency of evaluations be the same or different depending upon experience and prior levels of effectiveness? Within the provided documentation, it does not appear that the consortium has established multiple measures for principals. In identifying multiple measures, it will be important to align the criteria to potential measures and determine the feasibility of each measure within the different districts. It appears that the consortium has a document (Principal Criteria and Measures) that it may use to determine which measures will be used to evaluate each of the eight criteria. Weighting of measures is another consideration for the consortium. When multiple measures are identified, it will be important to identify appropriate weights for the measures. For example, what portion or percentage of a principal's evaluation will be based upon available student achievement data?

It appears that the consortium used an AWSP document (Principal Leadership DRAFT 09152010) to guide the development of the criteria and rubrics for the principal evaluation system. This document uses the eight criteria set forth by the state to provide guidelines related to knowledge and skills, evidence and measures, and system considerations. Within its criteria and rubrics documentation (Revised MASTER Principal Criteria and Rubrics Feb. 14), the consortium may want to consider expanding its definition of indicators so they are more explicit. This will help to lower inference and support a more objective approach to principal evaluation. The examples of measures and evidence provided in the AWSP document could be reliably measured. The consortium will need to use this document and others to determine which of the measures it will use for its system and how it will weight them within the overall evaluation.

There was no evidence provided related to data use and tracking.

Training for Evaluators

As with the teacher evaluation system, the consortium had not yet reached the point of determining where, how, or when evaluators would be trained but indicated that it would happen during the summer of 2011, roughly at the same time as the training on the teacher evaluation system.

Using Results for Professional Development

Principals in the consortium indicated that they would like the principal evaluation to be used as a tool that guides reflection on their performance; however, the degree to which the feedback will be tied to profession development is yet to be determined.

Ensuring Effective Communication and Ensuring Stakeholder Investment

Focus Group Feedback. The principal performance evaluation system will also be rolled out in fall 2011; however, there was no discussion around how this information has been communicated to stakeholders.

Document Review. It is unclear from the documentation provided whether there was broad stakeholder involvement in the design of the principal evaluation system. There was some evidence of stakeholder meetings or forums (NEWAS Presentation Dec. 2 and NEWASA Presentation Feb. 17) that appeared to address both the teacher and principal evaluation systems. These events were communicated via e-mail, memo, or press release (Educator Forum Invitation and Eval Pilot Announcement). There was no evidence of stakeholder feedback within the provided documentation. It is unclear whether the school community understands and values the contribution of the evaluation system at this time.

There was no evidence provided related to pilot implementation tracking and reporting.

Process Feedback

Throughout the course of each focus group, participants reflected on the process of designing a teacher and principal evaluation system. From this discussion, several themes emerged.

All consortium participants, regardless of role, expressed frustration with how the development of the teacher rubric had unfolded over time. Initially, when the consortium first met in the fall, participants were eager to begin working on designing the new system, particularly teachers who would have a hand in influencing how the rubric(s) would be rolled out during the pilot year. However, this enthusiasm began to dwindle as participants became more steeped in designing the rubrics and received feedback that they were not quite on track.

Communication from the state around the process has been ambiguous, leaving consortium participants feeling less certain about the state's credibility. As noted, participants were initially fully on board to begin designing the evaluation system but began to lose morale as they received messages that the way they were approaching the rubrics needed to change, without clarity around *how* they needed to change. As one teacher noted:

> I think it's been slowly swirling down the toilet as we've gone because every time we think, "Okay, we're doing a good job; we're right on track," we hear, "Nope, you've got to start over. Erase that."

Participants did respond favorably to Charlotte Danielson's workshop held in February 2011 because they felt she gave them concrete information about how to proceed. After attending Danielson's workshop, participants indicated they that they left with useful information because it "was the first time we had seen other peoples' work." Up until meeting with Danielson, participants indicated that the roadmap for where the rubrics could go was vague, which left them uncertain about how to proceed. Danielson's workshop helped the participants solidify a direction, and after the workshop, the consortium agreed to use Danielson's model as the primary framework for their rubrics.

Participants expressed ambivalence about the ultimate utility of the performance evaluation rubrics. Most consortium participants indicated that if the teacher rubric would be used as a reflective, focused, professional growth tool, they would welcome that feedback. However, participants expressed a deeper concern that teachers' performance evaluations would be used as means to "get rid of bad teachers." This sentiment was echoed repeatedly, and captured in the following statement made by a teacher:

> I love the professional growth model. I want that to go. I just want to make sure that [nothing happens] to hamstring that—to distract me from professional growth in order to protect me from professional elimination.

Participants have felt that the process has been too rushed for them to design a quality tool. Superintendants, principals, and teachers all expressed concern that the push to design the teacher performance rubric has been too swift. By February, the morale of the group appeared to be low, and at one point during the teacher focus group, one participant stated that she felt the group had been "set up to fail."

Summary and Conclusions

Based on the information gleaned from focus groups with consortium participants and a review of all documents submitted to AIR, the consortium appears to be halfway through the process of developing the multiple components of its principal and teacher evaluation system.

In general, the consortium has been onboard with the process since August 2010 but went through several fits and starts as it worked to establish the indicators by which teacher and principals should be evaluated and how their performance should be measured. In addition, the consortium did not intend to share and disseminate information about the upcoming evaluations system until the participants, as a group, had a clearer sense of their model.

In addition, the consortium addressed a number of issues that may have impeded its design process, including the following:

- Lack of clarity around the process from the state
- Ambivalence about how performance evaluations actually will be used
- A timeline for completion that felt too rushed

North Mason Public Schools
Case Study Report

District Overview/Description[1]

North Mason Public Schools serves about 2,200 students each year through its six schools, which include two elementary schools, one middle school, one high school, one options high school, and one Homelink program. Located in rural Washington at the eastern tip of the Hood Canal, North Mason is a primarily Caucasian school district with the following demographics:

- 2.97 percent of North Mason students are American Indian or Alaskan Native.
- 1.67 percent are Asian or Pacific Islander.
- 1.26 percent are African American.
- 6.96 percent are Hispanic.
- 39.45 percent are eligible for free or reduced-price lunch.

Case Study Method

This study was undertaken to help the Washington Office of the Superintendent of Public Instruction (OSPI) better understand how to support districts' implementation of Senate Bill 6696, which establishes new criteria for evaluating educators. For this component of the study, three individuals affiliated with North Mason Public Schools were interviewed in an effort to collect information concerning the process of developing and implementing its new teacher and principal evaluation systems. Each of the respondents has played an active role in the development of the new systems. In an attempt to gain a diverse range of institutional perspectives, we interviewed one district-level administrator, one school-teacher, and one union official. All responses were coded based on topic and related through the following case study.

The district also was requested to submit pertinent documents concerning the development and implementation of its teacher and principal evaluation systems. These documents were then reviewed by AIR staff who documented evidence of the district's actions. It is important to note that this case study report captures only the findings from interviews conducted and documents submitted prior to mid-March. It therefore does not capture developments that have taken place since then or documents that were not available at the initial time of review.

Teacher Evaluation

Setting the Evaluation Framework and Goals

Around October 2009, members of a teacher evaluation task force comprising principals and teachers from a variety of grade levels in North Mason began simultaneously developing an

[1] Information concerning North Mason School District was found at the district's website (www.nmsd.wednet.edu/nmsd/site/default.asp) and the Washington State Report Card (http://reportcard.ospi.k12.wa.us/DataDownload.aspx).

instructional model and a teacher evaluation rubric, which are linked. The development of the district's instructional model began with a group of experts from the University of Washington Center for Educational Leadership conducting a needs study for North Mason. They also received training in The BERC Group's Powerful Teaching and Learning, which one respondent reports as most heavily influencing their work. The group also looked to Charlotte Danielson and others' frameworks as guides for their own. Both the instructional model and the teacher evaluation rubric have gone back to the school sites about once a month so that there has been ongoing teacher review and input.

The instructional model, which has not yet been formally adopted, is composed of two parts. Interview participants reported that the first part is about what is expected to take place in a classroom and includes vision of instruction, clear learning targets, clear procedures and routines, focused instruction, student engagement, frequent and ongoing assessment, feedback interventions and enrichment, evaluation, and closure. The second part focuses on planning elements, including learning objectives, the development of common aligned assessments, the development of units, teaching the unit, administering formative assessments, analyzing and adjusting the summative assessments, and lesson or unit closure.

The North Mason School District uses the terms *instructional framework* and *instructional model* interchangeably. The rubric is used for preobservation teacher reflection and teacher-principal conferences, as an observation tool, as a postobservation teacher reflection/conference tool, and as a final rating tool for each formal observation.

When asked about the major goals of the teacher evaluation system, respondents agreed that improving teaching and learning were considered key. More specifically, respondents cited additional goals of helping teachers be more purposeful in their instruction, providing them with tools, developing clear instructional standards, and clarifying how teachers can move up a tiered model of effectiveness. Human capital management functions such as hiring, firing, and compensation have not yet been discussed but are not off the table.

Establishing Evaluation Criteria

North Mason is using the eight criteria provided by the state as the eight domains in its summative teacher evaluation rubric, which is aligned exclusively to the instructional framework, which adequately captures all aspects of effective teaching. The committee is developing a number of more specific elements for each of the criteria. One respondent noted that the process has been somewhat more complex than it had to be because it was begun one year ago, prior to the state's development of eight criteria, and so elements that had initially been developed for a slightly different set of criteria have had to be rearranged. Another respondent noted that, because teachers and principals have been providing input over the past 1½ years, they have developed buy-in.

Depending on where teachers fall on the rubric, they will be rated 1–4, with each level represented as follows:

- Level 1: Undeveloped

- Level 2: Basic/Developing
- Level 3: Proficient
- Level 4: Distinguished

This four-tiered system replaces the existing binary system that rates teachers only satisfactory or unsatisfactory. In addition to receiving an overall rating, teachers will receive a 1–4 rating for each criterion. Respondents agreed that ratings are based on meeting the criteria on the rubric rather than a normative-based system whereby teachers are compared with one another.

Detailed descriptions have been written for what constitutes each of these rating levels, with tiered expectations for teachers at different phases of their careers and evaluation cycles (e.g., professional growth plan, summative, etc.). To create benchmarks for each rating, the committee looked at state regulations, the benchmarks of the National Board for Professional Teaching Standards, and work from other districts. However, the expert reviewer found the language of the summative rating categories somewhat confusing, with differing labels for Level 2 depending upon teachers' professional status.

The committee initially had decided that the cut score would be between Levels 2 and 3. However, for provisional teachers, or experienced teachers on new assignments, it would be acceptable to be at Level 2 for a specified period of time. The expert reviewer felt it was unclear how this varied cut line for "meets expectations" will impact the distribution of summative ratings for the entire population of North Mason teachers. For example, the lower bar for provisional staff could force a more equalized distribution for new teachers by moving the whole distribution downward, while raising the bar for veteran teachers could make a Level 2 rating essentially the same as a Level 1 rating for this group, compressing the ratings distribution. In an attempt to explain the process, the North Mason School District created a document titled Teacher Evaluation—Track Explanation Document. Further, the district continues to have discussions about the meaning of this cut line and its implications for all staff members.

Frequency of Evaluation

It appears that the North Mason School District has detailed the frequency of evaluation observations in line with what is codified in the state legislation (E2SSB6696, Part II, § 202(3)(a)), but that this issue is still being further discussed by the committee. Respondents noted that all teachers, provisional and continuing, will have one summative evaluation per year. Depending on which of four tracks individual teachers fall in, however, the frequency of their observations will differ. An interview respondent stated that teachers on first tier, the provisional track will be observed about six times per year, with those on the second tier, the summative evaluation track, being observed about three times per year. However, the documentation provided at the time of writing (which was subsequently updated) suggests that provisional and newly tenured teachers will be observed just twice per year. The third track, the professional growth plan, involves only one formal observation, but this is accompanied by meetings with an administrator to check on professional growth progress. Finally, the fourth track, the assistance phase, is in place for teachers who score low on the rubric and are then provided significant amounts of help to improve their practice. While the plan appears well-considered, the reviewer

was uncertain whether the evaluation timelines and frequency of observations will be manageable for the principals and assistant principals who will be conducting the evaluations.

Measuring Performance

Although no final decisions have been made concerning which measures will be used to evaluate teacher performance, the committee has developed a draft document that specifies a list of potential quantifiable measures. The list of possible sources of evidence, "meant to be a natural harvest of teacher work," emphasizes artifacts of teaching such as student work and assessment data.

The plan is that observations will be mandatory for all teachers, but in incidents where a measure is not observable, teachers would provide evidence from the multiple measures document. Teachers would have a limited selection of options to demonstrate that they have met the criteria. Respondents concurred that all teachers, core and noncore alike, would use the same list of measures. A committee member noted that, subsequent to recent feedback from AIR and the superintendent, a short, more directive list of measures has been created from a longer menu of choices for teachers completed by the task force. One respondent reported that it may be possible to link these measures to the rubric through a data management system they are exploring with the Washington School Information Processing Cooperative, but as of late March this idea had not yet been developed.

Test scores will be taken into account in the evaluation system in that the instructional domain on the rubric requires that state, district, and classroom testing be used by teachers to inform their teaching and to differentiate instruction as needed. The district uses Measure of Academic Progress (MAP) tests, which are administered several times per year and therefore allow for testing of growth over time. However, teachers primarily would be rated not on whether growth took place necessarily but on how they used the data to inform their instruction.

A closer look at the district's newly developed instructional model (using the document Teacher Evaluation Rubric Revised 3-1-11) reveals some measurement concerns at the descriptor level. First, the evidence being considered varies between rubric levels (e.g., in Standard 2, the *Distinguished* descriptor requires measurement over multiple observations or outside the classroom, and *Undeveloped* pertains to a single lesson). Second, the difference between *Proficient* and *Distinguished* is not always elucidated by the descriptor language. Third, descriptors for *Developing/Proficient/Does Not Meet* do not support tiered application to new versus veteran staff (e.g., the Standard 5A descriptor says teacher "addresses students in a disrespectful manner," which isn't appropriate behavior for any teacher).

At the time of writing, there was no documentation available in the area of data uses and tracking. However, as noted, a task force member indicated that the district is planning to address this through consideration of piloting a teacher evaluation data management system with Washington School Information Processing Cooperative.

Training for Evaluators

From the documents submitted at the time of writing, it is evident only that a group of teacher representatives attended a training on the new instructional framework (T-Eval Training Agenda). However, the contents of this training, whether additional sessions took place, and the involvement of administrators were not readily apparent. The interviews revealed that all evaluators, which includes the five principals in the district, have received two rounds of evaluator training on the eight domains alongside the task force members, so that they could evaluate teachers in the pilot. Through various rounds of input solicitation and through the initial piloting, the evaluators have become familiar with the rubric and its contents. However, one interview participant clarified that the district has not yet conducted any training on interrater reliability, although another respondent clarified that this issue will be addressed during the district's administrative retreat in August 2011.

Using Results for Professional Development

Each interview participant viewed the system as a means of helping teachers to be more effective. Teachers and their principals will be expected to look at the criteria and elements where improvement is needed and target their professional development accordingly.

For teachers who are not rated above the cut score, supports such as coaching or classes will be provided to help them improve more quickly. For teachers who already are satisfactory, the system is designed to have them wondering how they can improve in order to be distinguished. One respondent also noted the district's interest in considering having those who are rated Level 4 model or showcase some of their practices to help other teachers learn and grow.

For several years, principals have been providing training to teachers at their schools on data use. One respondent noted that one possibility being considered with the new evaluation system is looking for areas of widespread weakness where buildingwide professional development opportunities could be targeted.

One participant clarified that if the pilot teacher evaluation data management system described above is implemented, supports for individual teachers could be built into the system such that unsatisfactory performance for any category would result in a list of supports being made available for teachers to select from and track results.

Communicating Results and Ensuring Stakeholder Investment

At present, about 20 volunteer K–12 regular classroom teachers, selected by the task force co-chairs in part because they were considered effective teachers, are piloting the rubric in their evaluations. A memorandum of understanding (MOU) was created for the pilot volunteers to ensure that they would not be negatively affected by their evaluation results. Although this MOU expressly prohibits using the data gathered as part of the pilot to place a teacher on probation, it is not yet clear what data are being collected or how this information can be used to support teachers in the pilot. The committee's goal is to expand the pilot next year to include a randomly selected 25 percent of the teaching force, before putting the new system into effect districtwide

in 2012–13. At that point, all classroom teachers will be evaluated on these new criteria, as required by the law. Interview respondents indicated that a next step for the committee is to determine how to adjust the rubrics so the same tools can be used for evaluating music and physical education teachers and school librarians.

One participant reported that North Mason will take as much time as the state will give it, so that it can make the system as strong as possible before it is put into full effect. In addition to making changes based on the pilot evaluations, after every task force meeting the committee members (which include representation from every school in the district) return to their schools and share the progress made at staff meetings and solicit feedback. Teachers have thus far been engaged through the union-district task force and Teacher-Principal Evaluation Pilot (TPEP) committee and a set of focus groups to provide feedback on the observation tool. In addition, the assistant superintendent and teacher association president have held meetings with external groups, including each of the district's parent groups, the high school Associated Student Body class, and the local Kiwanis club, to share what the district is doing with the wider community; however, the involvement of these groups is limited to providing "advice" as to what the task force should be thinking about in developing the teacher and principal evaluation systems. The advice is collected and presented at TPEP task force meetings and is considered as revisions are made. As a result of the ongoing two-way communication, one participant noted, virtually everyone is on board with the professional growth piece of the new system.

Principal Evaluation

Interview participants reported that North Mason's principal evaluation system is in an earlier stage of development than its teacher evaluation system. Whereas the teacher evaluation reforms began 1½ years ago, the principal reforms began only this year. At the same time, one respondent noted that the pace at which it will be implemented may in fact be faster because there are only five principals in the district, three of whom are new.

Setting the Evaluation Framework and Goals

As was the case for teacher evaluations, all respondents agreed that the main goal for principal evaluation is to improve teaching and learning. One respondent noted that further goals, such as hiring, dismissing, or compensating principals based on their performance, have not yet been addressed and would have to be negotiated at a later point.

Respondents did not view the district as having any model for principal practice similar to the instructional model for teachers. However, they have a leadership framework, grounded in Robert Marzano's 21 characteristics of effective school leaders and in a resource from the Washington Association of School Principals, and these are considered to have been helpful in developing their draft principal rubric.

Although their framework appears to be a comprehensive description of effective school leadership, the rubric provided appears to be missing component labels, which may make it difficult for evaluators to decipher the subparts of each larger standard for measurement and review purposes. Some of the standards and components include overlapping concepts—for

example, Components 6.6, 6.8, and 6.9 are similar, and several of the components in Standards 3, 5, and 8 take a different perspective on the same ideas—which also may lead to difficulty in applying the rubric across multiple schools and contexts.

Establishing Evaluation Criteria

The principals on the task force committee have taken the lead on developing the principal rubric, deciding on the elements under each of the state's eight criteria. Only limited documentation was available at the time of writing on the principal evaluation tools and process currently in development; documentation did not include a summative rubric that captures all aspects of effective school leadership. Rather, it comprised a menu of multiple measures of principal performance, including stakeholder surveys; review of documents such as school improvement plans, budgets, schedules, and professional development plans; and student achievement data.

Frequency of Evaluation

The district has not yet progressed far enough to determine the frequency of evaluation for principals.

Measuring Performance

One interview respondent noted that the superintendent is currently working with a group of principals to develop a list of measures that principals can use to demonstrate their effectiveness, and that more systemwide factors, such as high school failure rates, will be incorporated. Principals will be rated on the same four-point scale as teachers:

- Level 1: Undeveloped
- Level 2: Basic/Developing
- Level 3: Proficient
- Level 4: Distinguished

As with teachers, the expectation is that, after one year, a principal would be at the Proficient level. According to one respondent, specific benchmarks were created for each rating level by looking at the state criteria and regulations and other districts and states.

Training for Evaluators

The superintendent and assistant superintendent, who both have been heavily involved with the development of the new system, will conduct the evaluations for the district's five principals. As mentioned in the teacher evaluation section, all parties are receiving ongoing training on the contents of the rubrics as they develop and change. According to one respondent, they are currently receiving training to ensure interrater reliability, although it is not clear if the training is formal or rigorous, or rather more of a thorough discussion.

Using Results for Professional Development

Results from the principal evaluation system will inform professional development, and it is expected that this will take place both on an individual level and districtwide. As with teachers, principals will have a meeting with their evaluator who will identify areas of strength and weakness and ways to provide effective professional development. However, these procedures have not yet been formally adopted.

Communicating Results and Ensuring Stakeholder Investment

North Mason has developed each level of the principal rubric, and one respondent noted that both the superintendent and the assistant superintendent have used the draft principal rubric in midyear evaluations during the current school year. One respondent noted that members of the TPEP committee and other stakeholders have been included in the same method as described above for the teacher rubrics. The district plans to pilot the principal evaluation system in 2011–12, but the reviewer did not see information at the time of writing about what this pilot will entail. Because the number of individuals affected is so limited in the case of principals, it does not appear from the interviews that a formal communication strategy has been seen as required.

Summary and Conclusions

Respondents view the new system positively. One respondent believes the greatest strength is the ability of the new evaluation system to significantly improve teaching by identifying areas of strength and areas in need of continued growth. On a related note, another respondent views the greatest strength as being that the system will provide clearly defined and articulated standards that will allow teachers and principals to assess their performance levels. Another respondent noted as a strength the fact that the rubric has raised the level of conversation about teaching and learning in a way that those involved had never before experienced.

At the same time, those interviewed were aware of challenges as well. One interview participant cited the general idea that change can be difficult and that it will take some time for teachers and principals to get used to rating themselves and engaging in serious self-reflection. Maintaining ongoing communication with all affected parties also was cited as a challenge. Another respondent viewed achieving interrater reliability as the greatest challenge. A final challenge noted by one respondent was the possibility of the state adopting a model different from the one they have worked on developing and gaining buy-in on for two years; if that were to happen, buy-in was considered to become difficult.

Overall, respondents seemed enthusiastic about the project and its potential. One respondent noted how grateful the committee was to be a part of the pilot, particularly to have had the opportunity to receive feedback from Charlotte Danielson and to learn from the other pilot districts. Nevertheless, the participants recognized the challenges and scope for challenges that lay ahead. It was obvious that a great amount of time and thought had been directed toward the development of the elements within each criteria of the teacher evaluation system, the definitions of the four rating levels, and the importance of ongoing communication with all stakeholders. However, the district still has some decisions to make, particularly regarding the specific performance indicators to include and the training of evaluators.

North Thurston Public Schools
Case Study Report

District Overview/Description[1]

North Thurston Public Schools serves nearly 14,000 students each year through its 13 elementary schools, four middle schools, three comprehensive high schools, and one high school of choice. The school district employs more than 1,700 staff. Located in South Puget Sound, North Thurston is Washington's 23rd largest school district. Moreover, North Thurston is the most ethnically diverse school district in the South Sound region, as follows:

- 2.82 percent of North Thurston students are American Indian or Alaskan Native.
- 12.96 percent are Asian or Pacific Islander.
- 9.22 percent are African American.
- 11.28 percent are Hispanic.
- 38.56 percent are eligible for free or reduced-price lunch.

Case Study Method

This study was undertaken to help the Washington Office of the Superintendent of Public Instruction (OSPI) better understand how to support districts' implementation of Senate Bill (S.B.) 6696, which establishes new criteria for evaluating educators. For this component of the study, four individuals affiliated with North Thurston Public Schools were interviewed in an effort to collect information concerning the process of developing and implementing their new teacher and principal evaluation systems. Each of the respondents has played an active role in the development of the new systems. In an attempt to gain a diverse range of institutional perspectives, American Institutes for Research (AIR) interviewed one district-level administrator, one school-level administrator, and two teachers. All responses were coded based on topic.

Moreover, the district was requested to submit pertinent documents concerning the development and implementation of its teacher and principal evaluation systems. These documents were then reviewed by AIR staff who documented evidence of the district's actions.

Teacher Evaluation

Setting the Evaluation Framework and Goals

According to the interviewees, North Thurston developed an instructional framework grounded in the Effective Schools research[2] that is used in all K–12 schools. It is important to note that this

[1] Information concerning North Thurston Public Schools was found at the district's website (www.nthurston.k12.wa.us/2246106811396587/site/default.asp) and the Washington State Report Card (http://reportcard.ospi.k12.wa.us/DataDownload.aspx).
[2] For more information on Effective Schools, go to www.effectiveschools.com/.

American Institutes for Research®

model goes beyond an instructional framework and encompasses all aspects of a school, beginning with high expectations for all students. All schools have gone through the process of developing power standards based on the state's grade-level expectations. Pacing guides have been developed based on the power standards to ensure that teachers in each school are covering the same material.

One respondent reported that, prior to the pilot project, North Thurston did not have a strong correlation between what the district would characterize as its instructional framework and its evaluation instrument. Given that the district's current instructional model is not based off any single instructional framework, a couple of respondents reported that it has been a challenge—but a welcomed opportunity—to link the instructional framework to the teacher evaluation system. The pilot committee has aligned each area of the Effective Schools framework to the state-mandated evaluation criteria to illustrate how they are connected. One respondent reported that this process has challenged the district to develop a clearer articulation of the instructional frameworks and its outcomes. This has resulted in a feedback loop whereby the pilot committee is assessing what the district values in terms of quality instruction and including this in the teacher evaluation. Moreover, the process of developing the evaluation and learning more about current practices in teacher evaluation have, in turn, informed conversations about the instructional framework.

When interview participants were asked about the main goals of the district's teacher evaluation system, several different responses were elicited. One respondent considered the main goal to be improving instructional practice. Another respondent provided two main goals: quality assurance and professional development. Ensuring that every student has highly qualified, effective teachers is of paramount importance, and providing professional development that is driven by the teacher evaluation system and specific to teacher goals is one way a district can support teacher quality. Another respondent thought that providing better feedback between the principal and teacher is most important, and professional growth also was cited as a top priority. All respondents echoed the same sentiment concerning the importance of professional development and professional growth, and the desire to ensure that the teacher evaluation system is a major component informing that process.

Each respondent noted the problematic nature of linking evaluation with teacher compensation and felt sure that this was not a policy the district intended to pursue. One respondent considered linking evaluation results with teacher compensation antithetical to the philosophical interests of the district. Another respondent noted that stakeholders can find this component of an evaluation system troublesome, fearsome, and problematic. However, the evaluation tool could and probably will be used in making continuing contract decisions. One respondent thought that the teacher evaluation tool could be useful as an indicator of a teacher's readiness for promotion or advancement.

The document review confirms that the district's current instructional framework is grounded in the Effective Schools research. Documentation of the components and competencies of this framework were not provided. Moreover, evidence was provided that the district plans to align its evaluation system to its existing instructional framework.

Establishing Evaluation Criteria

According to the respondents, North Thurston is currently in the process of developing elements for the state's mandatory criteria. When the elements for each criterion have been finalized, the district will develop performance indicators aligned to each element to measure what teachers know and do in their classrooms. These performance indicators will provide guidance to evaluators by illustrating how a Level 1, 2, 3, or 4 teacher would meet each element. Then, rubrics and observation measures will be developed and piloted to assess the extent to which teachers meet the new criteria. The table lists the current set of elements used to unpack each criterion.

Criterion	Elements
1. Centering instruction on high expectations for student achievement	a) Alignment of instructional practices to expectations b) Facilitation of student engagement
2. Demonstrating effective teaching practices	a) Lesson organization b) Instructional practices c) Adherence to alignment of content standards
3. Recognizing individual student learning needs and developing strategies to address those needs	a) Assessment use b) Responsiveness to student learning c) Adaptation of instruction
4. Providing clear and intentional focus on subject matter content and curriculum	a) Content knowledge b) Use of standards c) Pacing
5. Fostering and managing a safe, positive learning environment	a) Creating an environment of respect and rapport b) Classroom procedures c) Management of physical space d) Management of student behaviors
6. Using multiple student data elements to modify instruction and improve student learning	a) Assessment of data b) Assessment design c) Data-driven lesson design
7. Communicating and collaborating with parents and school community	At the time of the interviews, North Thurston was still developing elements for Criterion 7.
8. Exhibiting collaborative and collegial practices focused on improving instructional practice and student learning	a) School improvement and district initiatives b) Professional growth and development c) Professional learning communities

One of the respondents identified three broad themes covered by the criteria and elements:

- A robust knowledge of individual student needs and evidence that a teacher is differentiating instruction and assessing students in an ongoing manner and planning for instruction to meet the needs of all learners
- The active engagement of students in the learning process

- Communication and collaboration with parents, colleagues, administrators, and the community

As for who will be subject to the teacher evaluation model, the pilot committee is planning to development a model that will be applicable to K–12 teachers. All respondents foresee the model eventually being applied to every general education teacher, K–12, as well as some special education teachers. However, the respondents were not sure if specialists (e.g., teacher librarians, counselors, school psychologists) would eventually be subjected to some version of the teacher evaluation tool.

According to the respondents, the frequency of evaluations has yet to be determined for both continuing contract and provisional teachers. At a minimal level, provisional contract teachers will be evaluated more often than continuing contract teachers; however, if an individual teacher needs additional attention, the evaluation tool will certainly make that possible.

The North Thurston teacher evaluation model is a four-tiered system with the following distinctions:

- Level 1: Does not meet expectations
- Level 2: Emerging/needs improvement
- Level 3: Proficient/effective
- Level 4: Distinguished

This model replaces the previous bimodal system of satisfactory and unsatisfactory. According to the respondents, performance indicators have been developed for each element to assist evaluators in rating teachers according to each criterion. The cutline will fall between Level 1 and Level 2. No information was provided as to what action will be taken for teachers who fall below the cutline.

All four respondents related the steering committee's disdain for differential ratings in which one teacher is evaluated in comparison to another. The goal is to measure teachers against instructional standards rather than have a normative-based system.

According to the document review, the teacher evaluation framework submitted includes four rating levels and various elements and descriptions of behavior at each level, for each criterion designated in the state legislation. The documents do not include a description of how the framework will be scored, whether the criteria are weighted differently, or what additional elements (e.g., student academic growth or achievement) will be included in the teacher evaluation system. Moreover, the documents did not include any statements on the frequency of evaluation.

Measuring Performance

Although no final decisions have been made concerning what measures will be used to evaluate teacher performance, respondents posit that the district is considering observations, portfolios,

lesson plans, professional growth plans, student achievement data, perception survey data from students and parents, student artifacts, self-assessments, and peer evaluation. Regardless of what measures eventually will be chosen, the steering committee decided that multiple measures will be used to evaluate teacher effectiveness across all criteria.

When asked about student growth measures, each respondent related his or her belief that student growth data are an important component of any teacher evaluation system but cannot be the sole arbiter of success. The district has curriculum-based assessments and periodic formative assessments that could inform teacher practice. All interview participants indicated that teachers are expected to review student assessment data and differentiate their instruction based on those data. None of the respondents were keen on including a value-added component to the teacher evaluation.

The steering committee has yet to decide whether measures will be different for teachers of core subjects and teachers of noncore subjects; however, several respondents were hopeful that the measures selected for use will be applicable across levels and curriculum.

According to the document review, the teacher evaluation framework does not yet include the types of data that will be gathered as evidence of performance (e.g., classroom observations, lesson artifacts, student surveys). Moreover, the documents do not address how student academic data will be included in the evaluation system or how the data will be linked to teachers.

Training for Evaluators

According to the respondents, evaluators will receive training on effective coaching strategies, observations, data collection strategies, and the use of evaluation tools. They also will receive training to ensure interrater reliability. North Thurston conducts an Administrative Leadership Academy every summer, which will serve as the vehicle for providing evaluator training. Each respondent reiterated the importance of providing a robust evaluator training program to ensure consistency and reliability.

At this point, no evidence was submitted for the document review concerning the selection and training of evaluators.

Using Results for Professional Development

Each respondent reiterated the importance of using the teacher evaluation instrument to inform professional development. Moreover, each respondent hopes that the new evaluation system will include a formative component. Although the specific mechanisms have not been determined, feedback from multiple information sources (e.g., observations, portfolios, etc.) will inform and guide teacher goal setting and professional development opportunities.

North Thurston has provided ongoing training for teachers concerning data use for the past five years. According to the respondents, the use of student data and data teams is a core component of the district's current professional development framework. This will continue with a particular emphasis on how data use, teacher evaluation, and instructional practice are intertwined.

At this point, no evidence was submitted for the document review concerning discussions about data use and tracking.

Communicating Results

According to the respondents, North Thurston will follow the state timeline for implementing the new teacher evaluation system. The district is using this year to develop the evaluation tools and processes. Over the summer, professional development and training will be provided to teachers and principals participating in the pilot year (2011–12). All principals will be participating in the pilot; however, teachers will be participating on a volunteer basis. After the pilot is complete, the district will use findings from the pilot to recommend changes to the evaluation system and begin negotiations with the teachers bargaining association. Throughout the pilot year, the pilot group will report to the steering committee with reflections, recommendations, and insights. The leadership team will make ongoing revisions throughout the process.

At this point, no evidence was submitted for the document review concerning plans to pilot the evaluation system. In subsequent discussions, it was confirmed that the district plans to pilot the evaluation system in accordance with state mandates.

As for communicating the new system requirements and procedures to teachers and community stakeholders, the respondents reported that a three-layer strategy is in place. For the first layer, representatives from each school meet with steering committee members and provide updates to interested parties in their buildings. The second layer consists of formal communications between the district and schools that provide periodic updates regarding the project. The third layer is what North Thurston has termed a "road show." A project director and project facilitator will go to each school and give a presentation of the draft tools at a staff meeting. These presentations will also serve to gather input and feedback from teachers as well as answer questions and address any concerns. Moreover, communication will also occur in a professional development context over the summer. One parent forum has been held for the purpose of gathering input on both the teacher and principal criteria. Additional parent input opportunities are currently being planned, which may include phone surveys and/or PTA/PTO meetings. District administrators have already visited each building in the district to outline the governing legislation (S.B. 6696) and to share that the district will be participating in the pilot program.

The document review provided evidence that North Thurston has made it a high priority to have representation from all schools, school levels (i.e., elementary, middle, and high school), and types of stakeholders (e.g., special education teachers and union representatives) on the pilot committee. Because this approach resulted in a large group, the pilot committee subdivided into a teacher evaluation subcommittee and principal evaluation subcommittee. In addition, the group formed a parent and community subcommittee to assist in model development. The teacher evaluation subcommittee divided its work according to the criteria designated in S.B. 6696. The bulk of its work was undertaken at in-person meeting in which multiple opportunities were provided for substantive feedback in response to summary documents. Documents submitted for the review do not contain any discussion of the extent to which support for the proposed system has been garnered through stakeholder involvement, but it appears that questions and concerns of the committee members are being addressed. The extent to which the school community

understands the value of a quality evaluating system in improving teacher performance is not documented. The district has outlined a plan to involve parents and community members in model development and revision through polls, surveys, and interviews.

Principal Evaluation

Overall, North Thurston has determined many more details of the teacher evaluation system than the principal evaluation system. However, given that principals are already rated using a four-tiered system, the principal evaluation model is more mature than the teacher model.

Setting the Evaluation Framework and Goals

When interview participants were asked about the main goals of the principal evaluation system, several different responses were provided. One respondent cited improved leadership practice, both instructionally and managerially. Another respondent thinks the new system will provide a higher level of quality assurance by helping assess whether a highly effective, competent principal is in every building. Professional growth was again mentioned as an important goal for both principals and teachers. Several respondents pointed out that, because the principals were already being evaluated by a four-tiered model, this is an opportunity for the district to refine its model and ensure that it is aligned with the state-mandated criteria.

The district did not submit documents that summarize its leadership framework; however, according to its timeline, these documents are under development. Moreover, it is evident that the pilot committee has reviewed considerable research thus far. At this point, the committee appears to have drafted a list of data sources and reached out to influential researchers about its thoughts on the framework.

Establishing Evaluation Criteria

According to the respondents, each criterion in the principal evaluation model is accompanied by a number of elements. Recognizing that the principal's function is to facilitate leadership and support the work of teachers in the classroom, one respondent identified four themes: (1) ensuring that systems are in place to support teachers' assessment programs and providing instructional materials and professional development to assist teachers in differentiating their instruction; (2) proficiency in assessing teacher performance in the classroom; (3) coaching, mentoring, and developing teachers; and (4) accountability—ensuring that a standard of quality is maintained across the school.

According to the document review, the district has not yet developed a clear format for the evaluation rubrics.

Measuring Performance

Principals in North Thurston are already rated using the following four-tiered system:

- Level 1: Does not meet expectations

- Level 2: Emerging/needs improvement
- Level 3: Proficient/effective
- Level 4: Distinguished

According to the respondents, these ratings will be retained in the new system. Each element has performance indicators that assist the evaluator in determining the principal's rating for each criterion. In order to receive a Level 4 rating, the principal must demonstrate leadership that extends beyond the confines of his or her own building.

Although the measures have not been decided, one respondent believes that formative and summative assessment results will be utilized in some fashion. However, this would only be one measure in a multimeasure principal evaluation model.

The document review also confirms that multiple measures for determining principal performance have yet to be identified. Moreover, the committee has not provided information on how frequent evaluations of principals will be conducted or whether evaluations will be different at different stages of a principal's career.

Training for Evaluators

According to the respondents, training for evaluators will be provided over the summer in conjunction with training for the teacher evaluation model. More information concerning the details of the professional development offerings is outlined in the Teacher Evaluation section.

According to the document review, the district did not provide evidence that an evaluator training process has been designed; however, the district's timeline suggests that it may be under development. The district's materials also mention that schools are required to report evaluation data for all employee groups, and teachers are to use student data to improve student learning.

Using Results for Professional Development

The respondents believe that results from the principal evaluation system will inform professional development; however, the exact mechanism has yet to be determined.

According to the document review, there is no evidence that the principal evaluation system will be linked to a comprehensive human capital management strategy, with linkages to principal professional development planning, progressive discipline, or compensation policies.

Communicating Results

According to the document review, the pilot testing materials will be developed in 2011–12 and implemented in the 2013–14 school year. The teacher evaluation development process has already begun, but the principal evaluation development process is not yet underway. The district has indicated that it is currently developing a plan to examine issues such as interrater reliability of observations and validity of measures.

According to the respondents, information concerning the principal evaluation system will be provided at the same time as information concerning the teacher evaluation system. Although development of the principal evaluation system is still in its beginning stages, the district plans to have it completed by the time the "road show" begins. For an outline of the district's multiphase communication strategy, please see the Teacher Evaluation section.

According to the document review, the district has outlined an informative and well-thought-out communications plan. The documents show that the district has taken great effort to engage educators, administrators, district-level personnel, and state education agency personnel throughout the development process of the principal evaluation system. The district has taken a number of steps to ensure good communication with stakeholders. For example, individuals responsible for reformulating the principal evaluation system began by recruiting all educators and administrators who would like to be involved in the process. Also, the district superintendent and chief operations officer began dialogue with key stakeholders in the community. Although there has been a demonstrated willingness to communicate with stakeholders and involve them in the development process, there was not much information in the submitted documents concerning the reaction of stakeholders to the proposed plans and whether stakeholder opinions or suggestions were considered during the development process.

Summary and Conclusions

Respondents recognize the strengths inherent in their new teacher evaluation system. One respondent feels confident that the new system will engage teachers and principals in valuable conversations concerning effective instruction. The new system has been touted as a vehicle for a principal to provide more robust feedback to teachers. Professional growth and development will be facilitated through feedback loops, and the evaluation system will provide clearer and more specific language concerning teacher competencies. Moreover, the rating system provides a continuum for growth and will incorporate multiple measures to inform teachers' performance and professional development.

When asked about the greatest challenges facing the implementation of the new teacher and principal evaluation systems, one respondent was concerned about the external imposition of value-added measures and high-stakes consequences. Another respondent recognized that the new system could be met with anxiety and fear on the part of teachers. Professional development was identified as a challenge, given that a large number of teachers and principals will need to be trained to make the best use of the new system. Another respondent considered the two greatest challenges to the initiative to be time and money. All members of the pilot committee are committed to constructing the best teacher and principal evaluation system possible; however, they are constrained by resources. It also was pointed out that principals will need to commit more time to ensuring that the robust conversations fostered by the evaluation systems can occur.

Overall, respondents seemed enthusiastic about the project and touted the benefits of a well-designed teacher and principal evaluation system. At the same time, they recognized the challenges not only of design but also of implementation. It was obvious that a great amount of time and thought had been directed toward the development of overall goals and elements for the

teacher evaluation system as well as the development of a multilayered communication strategy. However, the district still has some big decisions in its future concerning performance indicators and measures. Although the new principal evaluation system is not as fully developed as the teacher model, respondents were hopeful that work on the principal model would proceed more quickly given that a multilevel rating system is already in place.

Othello School District
Case Study Report

District Overview/Description[1]

Located in Adams County, 80 miles southwest of Spokane, the Othello School District had a 2009–10 total enrollment of 3,690 students and employed approximately 192 teachers. Othello's 2009–10 student demographic characteristics are as follows:

- 0.22 percent are American Indian or Alaskan Native.
- 0.62 percent are Asian or Pacific Islander.
- 0.38 percent are African American.
- 82.55 percent are Hispanic.
- 16.21 percent are white.
- 35.16 percent are transitional/bilingual.
- 79.12 percent are eligible for free or reduced-price lunch.

Case Study Method

This study was undertaken to help the Washington Office of the Superintendent of Public Instruction (OSPI) better understand how to support districts' implementation of Senate Bill 6696, which establishes new criteria for evaluating educators. For this component of the study, four individuals affiliated with Othello Public Schools were interviewed in an effort to collect information concerning the process of developing and implementing their new teacher and principal evaluation systems. Each of the respondents has played an active role in the development of the new systems. In an attempt to gain a diverse range of institutional perspectives, American Institutes for Research (AIR) interviewed one district-level administrator, one school-level administrator, and two teachers. All responses were coded based on topic.

Moreover, the district was requested to submit pertinent documents concerning the development and implementation of its teacher and principal evaluation systems. These documents were then reviewed by AIR staff who documented evidence of the district's actions.

Teacher Evaluation

Setting the Evaluation Framework and Goals

According to the interview participants, Othello developed the Quality Teaching and Learning (QTL) Framework based on a comprehensive review of research concerning quality teaching and

[1] Information concerning North Thurston Public Schools was found at the Washington State Report Card (http://reportcard.ospi.k12.wa.us/DataDownload.aspx).

learning. Several district-level teams were organized to review best practices with regards to instruction, assessment, and strategy. The QTL Framework includes rubrics and guidelines for teachers that outline what the district defines as effective practice in three main areas: planning, teaching, and assessment evaluation.

When asked whether the instructional framework was being linked to the teacher evaluation system, three participants replied in the affirmative, and one replied that it was not. One respondent noted that, given the investment the district has made in developing the QTL Framework, it is essential that the framework and evaluation system be linked by common language and themes. The evaluation system will essentially build upon the tools already provided by the framework. Another respondent noted that the pilot committee used the QTL Framework as a guide when developing most of the elements for the criteria.

All four respondents cited professional development as one of the main goals of the new teacher evaluation system. One respondent mentioned that having a common language that defines quality teaching and learning throughout the district is also important. It was continually reiterated that fostering the personal growth of teachers through the continual feedback of school administrators is crucial.

When asked about linking evaluation results to compensation, all four respondents stated that this was something the pilot committee has not discussed.

Establishing Evaluation Criteria

According to the respondents, Othello is currently in the process of developing elements for the state's mandatory criteria. One respondent posited that, when constructing the elements, the pilot committee attempted to identify behaviors associated with effective teaching and then to identify documentation that would provide evidence of these behaviors. Although the elements have not been finalized at this point, one respondent reported a number of elements for each criterion, which are listed in the table.

Criterion	Elements
1. Centering instruction on high expectations for student achievement	a) Lesson planning b) Lesson delivery c) Lesson assessment
2. Demonstrating effective teaching practices	a) Questioning and discussion techniques b) Incorporating higher-order thinking questions c) Instructional delivery d) Flexibility and responsiveness e) Student engagement
3. Recognizing individual student learning needs and developing strategies to address those needs	a) Knowledge of student needs b) Planning for individual student needs c) Addressing student needs d) Using data to differentiate instruction

Criterion	Elements
4. Providing clear and intentional focus on subject matter content and curriculum	a) Knowledge of content b) Use of content c) Intentional focus on academic vocabulary d) Staying on task
5. Fostering and managing a safe, positive learning environment	a) Routines and procedures b) Ensuring a positive learning environment with respect and rapport c) Managing student behavior d) Physical environment
6. Using multiple student data elements to modify instruction and improve student learning	a) Knowledge of available data b) Responsiveness to data
7. Communicating and collaborating with parents and school community	a) Participation in parent/school community
8. Exhibiting collaborative and collegial practices focused on improving instructional practice and student learning	a) Participation in professional learning communities b) Collegial and professional practices and interactions with colleagues

There was some discrepancy among respondents as to which teachers will be subject to the new evaluation system after the pilot year is complete. One respondent stated that every teacher would be involved. Another respondent believed that the new evaluation system would be extended to all general education and special education teachers but was not sure if the system would differentiate between the two groups. One respondent stated that the pilot committee was currently focusing on general education teachers, but that they are thinking about developing something different for special education teachers and specialists. The other respondent stated that the district has not decided who should be subject to the teacher evaluation system after the pilot year.

When asked if the frequency of evaluations would differ between provisional and continuing contract teachers, all respondents stated that this has yet to be finalized.

Othello's new teacher evaluation system will be a four-tiered model with the following distinctions:

- Level 1: Unsatisfactory
- Level 2: Basic
- Level 3: Proficient
- Level 4: Innovative

Although one respondent reported that a cutoff had been set between Level 1 and Level 2, no information was provided as to what actions may be taken when a teacher falls below the cutoff. Two respondents reported that preliminary discussions had taken place concerning the

differential weighting of criteria and elements; however, this had not been finalized at the time of the interviews.

When participants were asked about rating teachers in respect to one another, several different responses were elicited. One respondent thought this might happen at the district level. Another respondent reported that this is something the pilot committee is considering but has not discussed at length. Another respondent thought differential ratings would be incorporated into the teacher evaluation system and that this would be addressed in the near future. Another respondent was not sure whether differential ratings would be included but indicated that the system was not currently set up to compare teachers.

Othello's teacher observation rubric was submitted for the document review; however, the rubric does not indicate whether it will be used for summative or formative purposes.

The document review confirms that the teacher evaluation system will use the aforementioned four levels. However, the rubric submitted by Othello includes subjective language that could be interpreted differently depending on the evaluator. Moreover, the rubric also has inconsistencies with the descriptions associated with different ratings. From the documents provided, it was unclear when observations would be conducted, how often observations would be conducted, if there would be any differentiation between new and veteran teachers, and if there would be any differentiation between effective teachers and teachers in remediation.

Measuring Performance

At the time when the interviews were completed, the pilot committee had brainstormed appropriate measures for the teacher evaluation system; however, no final decisions had been made. One respondent suggested classroom "drop-ins," samples of student work, data, and assessments. The respondent also mentioned feedback meetings between teachers and principals whereby teachers give an overview to the principal of student data being generated and how the teacher is differentiating instruction based on those data. Last year, Othello implemented the shared circle responsibility conversation in which every teacher meets with his or her building principal on a monthly basis to discuss student progress in the class. The respondent felt certain that these meetings would be incorporated into the teacher evaluation system in some form. Another respondent suggested that teacher artifacts such as lesson plans should be included and was certain that observations would be included. Another respondent listed teacher artifacts and student artifacts as necessary measures. Two respondents mentioned the challenge of finding observable measures that provide insight into all the criteria.

As for the use of student growth measures, the respondents provided varying opinions. One respondent felt that state benchmark assessments, MAP testing, and other state exams should be used to inform the teacher evaluation system. Another respondent suggested that test scores could be used either as a teacher artifact or as a means to adjust instruction. One respondent was adamant that student test score data be used only to inform teacher planning, not as a measure of teacher success. Another respondent posited that using student test scores to evaluate teachers is "taboo." The same respondent reiterated that student test scores should inform instruction.

As of now, the same measures will be used for teachers in both core and noncore subjects, according to one respondent. Although using student data to differentiate instruction will probably be an included measure, which data are available and applicable to a given teacher's subject area will be defined by the teacher and principal. As for special education teachers and specialists, three respondents thought that a different rubric would eventually be developed to meet their unique needs.

For the document review, the district submitted a list of teacher actions and artifacts as well as a list of student actions and artifacts as potential measures. It is unclear which teacher and student actions are associated with the artifacts listed. It is also unclear how these artifacts will be used in the overall teacher evaluation process. Moreover, it is unclear if this list of artifacts is conclusive or if teachers can request to submit additional artifacts that are not listed. The district did not submit information concerning the weighting associated with each of the artifacts that reflect teacher and student action.

Training for Evaluators

According to the respondents, training most likely will be provided for evaluators over the summer. All four respondents reiterated the importance of ensuring interrater reliability. One mentioned the use of training videos as a means of practice for evaluators. Another respondent mentioned that the pilot committee had discussed developing a glossary of terms to ensure that all evaluators were using the same language when discussing elements and criteria from the rubric.

All four respondents mentioned the extensive training teachers receive in using data to inform instruction. Districtwide training is offered in the summer and throughout the school year, so teachers can better differentiate instruction and inform parents about student progress.

The document review received no evidence that a comprehensive plan for evaluator selection and training has been developed.

Using Results for Professional Development

When asked about using the results of the teacher evaluation system to inform professional development, three of the respondents discussed their current professional development offerings and how they might fit with the evaluation system. One respondent was uncertain how evaluation results would be linked to professional development. One respondent posited that the district has been very strategic about its professional development offerings and its expectations that teachers implement these practices in their classrooms. Othello has worked with Teachscape[2] for the last three years in an effort to increase student achievement. Another respondent mentioned the circle of concern monthly meetings between each teacher and principal. The respondent felt confident that these monthly meetings would be integrated into the evaluation system and used to make professional development recommendations.

[2] For more information on Teachscape, see www.teachscape.com/html/ts/nps/index.html.

No information was submitted for the document review that informs how the district will use and track data collected from the teacher evaluation system.

Communicating Results

The respondents reported that, during the pilot year, five teachers from each building in the district and one from the district's alternative high school will be randomly selected to participate. Teachers participating in the pilot are required to have at least three years of experience and a satisfactory final evaluation for the 2009–10 year. All teachers on the pilot committee also will participate. The pilot will include no specialists, special education, or new teachers. All teachers were informed by the district that if they are selected by the random draw, they are required to participate.

According to the respondents, pilot teachers will be sampled, and all principals will participate in the pilot. Professional development will be provided for principals and administrators, probably in June, and professional development for the pilot teachers will follow soon thereafter, probably in August. One of the respondents mentioned that the committee has consulted teachers for feedback concerning the evaluation rubric.

The document review confirmed that in order to be eligible to participate in the evaluation pilot, a teacher must have had a satisfactory evaluation for 2009–10 and have taught for three or more years. There will be a random selection of five or six eligible teachers and one ALPS from each school for the pilot. Teacher participation will be mandatory for those randomly selected.

As for communication with relevant stakeholders, according to the respondents, one of the district administrators on the pilot committee has been diligent about providing updates to the entire school district through e-mail. All the respondents mentioned the challenge of including parents and community stakeholders in the development process. One of the respondents mentioned reaching out to the community through print media. Another respondent suggested providing a professional development course on the evaluation system for interested parents.

According to the document review, the district submitted a plan to communicate with and engage relevant stakeholders; however, no evidence was provided to indicate if this plan has been carried out. There was evidence that stakeholders reviewed the observation rubric; however, the feedback that was collected was not incorporated into the rubric. Moreover, no evidence was provided that showed that Othello has communicated the teacher evaluation system's development, goals, or plan of implementation with the larger education community.

Principal Evaluation

Overall, Othello has determined many more details of the teacher evaluation system than the principal evaluation system.

Setting the Evaluation Framework and Goals

Although all four respondents reported that the pilot committee had just begun discussing the principal evaluation system, three respondents provided what they considered to be the main

goals of the system. One respondent thought the principal evaluation system would solidify the philosophy that teacher growth goes hand in hand with principal professional development. Two respondents cited professional growth as the main goal of the principal evaluation system.

One respondent mentioned that evaluation results could be used to inform promotion and advancement; however, he was very skeptical that they would be used to inform compensation.

According to the document review, Othello has not yet provided information concerning the district's instructional framework and the strength of the measures used to inform the development of its principal evaluation.

Establishing Evaluation Criteria

At the time of the interviews, elements had been developed for four of the state-mandated criteria. However, none of the elements were provided during the interviews. One respondent mentioned several indicators of principal competency—establishing relationships with teachers, understanding student and teacher data, understanding good lesson design, managing resources, being a good communicator, assessing and monitoring programs, understanding his or her role as an evaluator, and initiating professional growth dialogue with teachers.

The summative rubric provided for the document review identifies eight areas or "elements" of effective school leadership. However, none of the documents provide a description as to how the evaluator is to use the tool, what the cut-off scores should be, how the data listed in the Menu of Evidence should (or are) weighted, and whether the rubric allows differentiation among principals in different stages of their career.

According to the document review, the district has not yet provided information concerning the frequency of principal evaluations.

Evidence was provided by the respondents as well as the document review that the principal evaluation system will employ the four rating levels used by the teacher evaluation system:

- Level 1: Unsatisfactory
- Level 2: Basic
- Level 3: Proficient
- Level 4: Innovative

Measuring Performance

When asked what performance measures will be used to evaluate principals, all respondents reported that the pilot committee has yet to discuss or make any decisions concerning measures. One respondent thought the measures would be similar to those used in the teacher evaluation system. The respondent also thought that building-level data concerning student achievement will be employed.

The document review provided details concerning the multiple types of data sources that the pilot committee believed should be included as part of the principal evaluation.

Training for Evaluators

According to the respondents, professional development and training for evaluators will be provided over the summer. More information is included in the previous section concerning the teacher evaluation system.

According to the document review, Othello has developed the TPEP Action Plan Rev. and OSD PD Plan for Evaluation, which indicate that all principals will receive training or professional development on the structure and use of the new evaluation rubric. However, no information was provided as to who will evaluate principals, how evaluators will be selected and trained, and how interrater reliability will be assessed and maintained.

Using Results for Professional Development

According to the interviews, results from the principal evaluation system will inform professional development; however, the exact mechanism has yet to be determined.

According to the document review, the district has yet to provide information concerning data usage and tracking as it pertains to its principal evaluation system.

Communicating Results

According to the document review, the district has indicated that the pilot of the principal evaluation system will begin in the 2011–12 academic year and that all principals in the district will participate. However, the district did not include a more detailed and comprehensive plan pertaining to the specifics of the pilot such as timelines for dissemination and implementation; whether (and what type of) data will be collected and analyzed; how, if at all, the evaluation system will be used to inform principal mentoring and induction programs; or if (and how) the district has plans to use the feedback collected from the pilot to make changes to the evaluation system and tools.

According to the respondents, information concerning the principal evaluation system will be provided at the same time as information concerning the teacher evaluation system. Although development of the principal evaluation system is still in its beginning stages, the district plans to have it completed in time for the summer professional development sessions.

According to the document review, the district has included only precursory information as to which stakeholder groups will be involved in planning and/or providing feedback on the principal evaluation process. Although evidence was provided that stakeholder communication has taken place, it is unclear who from the district participated in these meetings.

Summary and Conclusions

When asked about the greatest benefits of the new teacher evaluation system, one respondent cited the use of a common language and the flexibility it will afford teachers to learn and grow. The respondent also thinks the new system will allow principals to provide more strategic advice and guidance to teachers. Two of the respondents thought that the greatest strength of the new system is that it is a growth model. The other respondent believes the rubric will be very informative and helpful for teachers. Rather than teachers being evaluated simply as satisfactory or unsatisfactory, the rubric will offer performance indicators that are aligned with the QTL Framework. This respondent also liked that the new system establishes quality discussion time between teachers and principals.

As for the greatest strengths of the principal evaluation system, one respondent likes the fact that it will allow district administrators to assist principals in their professional growth efforts. As with the teacher evaluation, the performance indicators will provide examples of principal competencies and provide guidance. Another respondent felt that moving toward a growth model was a positive step for the principal evaluation system by providing a vehicle to set goals and receive feedback on professional development.

The respondents also recognized the great challenges facing the new teacher and principal evaluation systems. One respondent cited the challenge of ensuring that the glossary of terms is communicated to the entire staff as well as to principals. The principals and teachers will need to use the same terms, with the same underlying definitions, and thoroughly understand the rubric in order to communicate effectively. The same respondent thinks it will be a challenge to help teachers understand that it is their responsibility to provide evidence of professional growth and student achievement. Another respondent recognized that going through the evaluation editing process will be a challenge for the pilot committee. One respondent mentioned that there are often issues of trust between school administrators and teachers and, if the evaluation system is going to be used to its full capacity, these will have to be overcome. The same respondent recognized the challenges ahead for the pilot committee in making revisions to the evaluation systems and ensuring that all stakeholders are thoroughly aware of the changes. One respondent cited the additional time burden on principals as a challenge of the new evaluation systems.

Overall, respondents seemed enthusiastic about the project and touted the benefits of a well-designed teacher and principal evaluation system. At the same time, they recognized the challenges not only of design but of implementation. It was obvious that a great amount of time and thought had been directed toward the development of overall goals and elements for the teacher evaluation system. However, the district still has some big decisions in its future concerning performance indicators and measures. Although work had just begun on the new principal evaluation system, respondents were hopeful that work would proceed quickly.

On a final note, one respondent mentioned that the district had hired consultants to help organize the pilot process and that this has been a major advantage. Another respondent mentioned the time burden the pilot committee places on teachers who are participating. The same respondent also mentioned that the committee feels stretched to complete all the work that must be done before the pilot year begins.

Snohomish School District
Case Study Report

District Overview/Description[1]

The Snohomish School District is located approximately 30 miles north of Seattle in Snohomish County. The District employs approximately 540 certified teaching staff and 500 classified staff through two high schools, two middle schools, 10 elementary schools, a preschool, an alternative learning program for secondary students, and a cooperative program for home-school families. The district serves 9,961 students in Grades K–12, with the following demographic characteristics:

- 1.0 percent American Indian or Alaskan Native
- 5.2 percent Asian or Pacific Islander
- 1.7 percent black
- 5.8 percent Hispanic
- 85.0 percent white
- 19.7 percent eligible for free or reduced-price lunch

Case Study Method

This study was undertaken to help the Washington Office of the Superintendent of Public Instruction (OSPI) better understand how to support districts' implementation of Senate Bill 6696, which establishes new criteria for evaluating educators. For this component of the study, four individuals affiliated with the Snohomish School District were interviewed in an effort to collect information concerning the process of developing and implementing their new teacher and principal evaluation systems. Each of the respondents has played an active role in the development of the new systems. In an attempt to gain a diverse range of institutional perspectives, American Institutes for Research (AIR) interviewed one district-level administrator, one school-level administrator, and two teachers. All responses were coded based on topic.

Moreover, the district was requested to submit pertinent documents concerning the development and implementation of its teacher and principal evaluation systems. These documents were then reviewed by AIR staff who documented evidence of the district's actions.

Teacher Evaluation

Setting the Evaluation Framework and Goals

According to the interview respondents, Snohomish's current evaluation tool and instructional framework is based on the Charlotte Danielson model. However, each of the respondents was

[1] Information concerning the Snohomish School District was found at the Washington State Report Card (http://reportcard.ospi.k12.wa.us/DataDownload.aspx)and the Snohomish School District website (www.sno.wednet.edu/index.php/district_information/).

quick to point out that the district had adapted the model to fit its unique needs. As one respondent reported, "We looked at her elements and had them reflect our district's culture, values, and standards." The district has decided to retain the Charlotte Danielson model, which it has been using for approximately seven years, for the new teacher evaluation system. Given that the district is currently using a four-tiered model that describes what a teacher would be doing in the classroom at each level, the district pilot committee is simply revising the model to meet the new state standards. This strategy has allowed the committee to review the current model with a critical eye—combining, separating, adding, and deleting elements as needed. Building on the existing model, the new model will be similar yet substantially different in regard to specific constructs.

All four interview participants agreed that the district's instructional standards were linked to the teacher evaluation system. One respondent expressed appreciation that the pilot provided an opportunity for the district to reevaluate its teacher evaluation tool and to update it to ensure a strong alignment with the instructional framework. The same respondent mentioned that this process also has informed the instructional framework through the development of such items as curriculum maps.

The document review confirms that the district's instructional framework is based on the Danielson framework; however, the model has been altered significantly and therefore no longer bears the psychometric properties of the original framework. Moreover, the instructional framework has a number of clear and well-defined descriptions of the components of effective teaching, but there are several cases where language needs clarification or modification.

When asked about the main goals of their teacher evaluation system, all four respondents mentioned the fact that it will assist in professional growth and development. The tool provides suggestions for growth regardless of whether the person being evaluated is a first-year teacher or a veteran. One respondent mentioned that the teacher evaluation system will inform school administrators on how to assist teachers who are on probation or consistently scoring "unsatisfactory" in some levels. The respondent mentioned that the tool is used to drive support, feedback, and growth not only for individuals but sometimes whole departments. As a growth model, the teacher evaluation system should be part of an integrated whole—as the district offers professional development options, school administrators are ensuring that these align with the stated needs of teachers. Three of the respondents mentioned that it is an extremely effective hiring tool as well. The proficient column of the evaluation was used to create a screening tool for applicants. One respondent mentioned the need for continual evaluation of the teacher model and initiating revisions as needed to meet the needs of the district. Another main goal listed by the respondent was to ensure that a proficient teacher is in every classroom and that students are learning.

Two respondents related that compensation is not tied to the evaluation system.

Establishing Evaluation Criteria

According to the respondents, evaluative standards have been developed for the state criteria; however, specific examples for all criteria were not shared during the interviews. The rubric

developed by the pilot committee delineates what observable characteristics a teacher possesses in regards to each criterion at each tier within the model (e.g., *Proficient, Emerging*). Two respondents posited that the district has decided that, to the extent possible, evaluative standards should be more qualitative than quantitative. One respondent mentioned that evaluative standards should go beyond the frequency of an observed standard to an understanding of how well it is embedded in the teacher's daily actions. One respondent pointed out three areas on which the evaluative standards should focus: data use by teachers, a thorough knowledge of instructional standards and being able to clearly articulate those to students, and collaboration among teachers in the district to ensure that best practices are shared.

All respondents related that the district's current teacher evaluation system has an evaluation rubric for all certified staff and educational service agencies. There are different iterations of the evaluation tool for general education teachers, counselors, special education teachers, nurses, librarians, and so on. One respondent pointed out the necessity of having different rubrics for different types of staff because of varying job descriptions and desired competencies. One respondent noted that only classroom teachers will be included in the pilot but that rubrics would eventually be developed for all staff.

For the document review, the district provided its teacher evaluation rubric. Overall, the summative rubric generally captures many aspects of effective teaching. As a whole, these components could create a comprehensive definition of a highly effective teacher. Teachers are instructed to use relevant student data in an effort to differentiate instruction; however, teachers are not evaluated on the basis of student test scores.

Three of the respondents noted that provisional teachers will be observed more often than continuing contract teachers; however, they were not certain as to the details. One respondent noted that after the development of the district's current evaluation tool, the number of informal observations by the principal increased significantly.

The document review found no information concerning the frequency of evaluations or the differences in evaluation systems for teachers at different stages of their careers. It is unclear whether the frequency of observations is manageable for assigned evaluators or if there are specific timelines and requirements regarding evaluator feedback.

Snohomish's new teacher evaluation system will be a four-tiered model with the following distinctions:

- Level 1: Unsatisfactory
- Level 2: Emerging
- Level 3: Proficient
- Level 4: Exemplary

One respondent notes that these ratings are qualitative and not quantitative. Although the cut score has not been determined, the same respondent felt that it should be between Levels 1 and 2.

Two respondents reported that benchmarks have yet to be developed. All four respondents reported that differential ratings will not be used.

Measuring Performance

In regards to measuring teacher performance, all respondents mentioned that observations will be the foundation of the evaluation. One respondent mentioned that a self-reflection component had been added to the district's evaluation tool. The same respondent thought that portfolio evidence may be required but was not certain about that. Another respondent mentioned portfolios as well as peer, parent, and student input. The same respondent cautioned on including peer or student feedback as a point component, noting the qualitative focus of the overall evaluation. One respondent felt that an important measure to consider is the use of student data.

In regards to student growth measures, two of the respondents expressed a profound distaste for the use of value-added scores, and another respondent posited that they will probably not be used. Three of the respondents thought it was important that teachers use student data to inform instruction; however, no specific measures were discussed. One respondent would like to see more use of common assessment data to evaluate student improvement; however, the respondent was wary of placing too much emphasis on a specific test. For the respondent, the ultimate goal of student data is the ability to illustrate that students are meeting specific learning targets.

One respondent reported that evaluation measures will not be different for teachers of core subjects and teachers of noncore subjects. Two respondents did not know whether the measures would be different, and one respondent did not think the evaluative measures should be different.

Training for Evaluators

Two respondents did not know the details concerning evaluator training, but all four respondents felt a robust training plan was crucial. One respondent mentioned that, currently, the district provides general training concerning the use of the teacher evaluation tool. Trainers also work with principals on interrater reliability. For example, time during administrative meetings has been used to watch videos of lessons and compare ratings. The same respondent reported that, at the secondary level, the district often has assistant principals assist with evaluations in subject areas that they have taught or for which they've been endorsed. One respondent mentioned that there has been evidence of low inter-reliability among raters for their current evaluation tool and thinks this needs to be addressed during evaluator training for the new tool. The same respondent stressed that training is essential if the new system is to have any credibility, transparency, and consistency. Another respondent mentioned the current training that is being conducted for the administrative core concerning the identification of instructional best practices.

According to the document review, there is little or no evidence to indicate that the district has addressed the selection of evaluators or ensured that they are trained to give quality evaluations or specific, constructive feedback to teachers.

Three interview participants outlined the district's plan for training teachers in data use, and one respondent was not aware of the district's plans. The district is currently developing a data

dashboard that will contain an array of standardized scoring information, both formative and summative. The goal of the dashboard is to provide any data measure the teacher needs to inform instruction and to make the data consistent and systematized across grade level, department, and school. This will enable the dashboard to provide tools concerning student performance trends. One respondent mentioned that any training in data use should provide a better understanding of the specific purpose of each assessment, how to interpret the data, and how to modify instruction based on this feedback.

The district did not provide any information for the document review concerning data use and tracking plans for its teacher evaluation system.

Using Results for Professional Development

Three of the respondents related how the district is currently using the evaluation tool to inform professional development. No plans for amending this model were shared. Currently, the main impetus driving professional development is one-on-one discussions between the teacher and principal. Based on those discussions, the principal can suggest that the district's teaching and learning department provide individualized services for that particular teacher or a group of teachers who are experiencing the same problems. One respondent mentioned that the district does not conduct large-scale training sessions for teachers.

Communicating Results

When participants were asked about the district's timeline for implementing the new teacher evaluation system, three of the four reiterated that their new system is practically an updated version of their old evaluation system, and they do not foresee any challenges with completing all components according to the state's timeline. This year, the district has been evaluating the state's new standards and rearranging its elements to fit within those standards. The pilot will begin next year. No firm details were provided as to how the pilot year will be organized. One respondent thought that one elementary, middle, and high school principal would participate in the pilot year. The evaluation tool will probably be piloted in their schools. The same respondent thought the year following the pilot would be used to modify the tool and develop amended rubrics for librarians, counselors, and others. By the 2012–13 school year, this respondent foresees the new evaluation system being used throughout the district for classroom teachers, and by 2013–14, the new evaluation system will be fully implemented for all staff.

No information was provided for the document review about the district's plans for piloting the new teacher evaluation system in the fall of 2011–12. It is unclear what types of data will be collected and how the new evaluation model will impact teachers in provisional status or those in a mentoring and induction program. Furthermore, no information was provided as to how the data collected during the pilot will be used.

Two respondents reported that principals have been kept abreast of the progress made by the pilot committee throughout the development of the new evaluation system. Although all four respondents mentioned that nothing had been finalized by the pilot committee, three respondents imagine that the communication strategy will be similar to the one used when the district

unveiled its current evaluation system. When introducing the current evaluation system, district administrators and the teacher association president met with staff from each school and provided training for teachers and administrators. One respondent mentioned that progress on the evaluation system has been shared with the district's school board. Another respondent thinks the pilot committee may invite the PTA or PLUS group to make comments on the new tool.

According to the document review, there is no evidence that the district has developed broad stakeholder involvement in the design of the teacher evaluation system.

Principal Evaluation

Setting the Evaluation Framework and Goals

One respondent reported that Snohomish's current evaluation system is based on the Association of Washington School Principals (AWSP) model. Another respondent is disappointed with the current system and thinks it lacks robustness. One respondent considers the main goal of the new principal evaluation to have a dynamic tool that illustrates how principals are performing and ways in which they can grow professionally. Another respondent posited that the main goal is to provide clear expectations about what is expected from an exemplary principal.

Establishing Evaluation Criteria

For the principal evaluation system, the district is following the same rubric model as the teacher evaluation system. For each of the state's criteria, there is a description that explains how a principal at each level would perform certain elements related to the criterion.

The interviews and document review confirm that the principal evaluation system will employ the four rating levels used by the teacher evaluation system, as follows:

- Level 1: Unsatisfactory
- Level 2: Emerging
- Level 3: Proficient
- Level 4: Exemplary

According to the document review, the district has developed a set of evaluator and self-assessment rubrics that generally align with state and national (Interstate School Leaders Licensure Consortium [ISLLC]) standards and priorities. The rubrics also include a list of indicators for each level of the system. A majority of the indicators have been finalized; however, this process is not complete. The district's levels build upon each other in that an *Exemplary* principal shows evidence of meeting all the indicators that a *Proficient* principal meets, among others. Moreover, the rubric appears applicable to principals at different stages in their careers.

Measuring Performance

None of the respondents knew which measures would be employed to measure principal performance.

According to the document review, although the district has created a rubric that includes major domains of effective principal practice, the pilot committee has not clearly defined the leadership evaluation framework. Definitions of key terms and criterion levels have not been provided. Moreover, the criteria include high-inference language that will present challenges for interrater reliability.

Training for Evaluators

As was noted for the teacher evaluation system, training will be provided to evaluators concerning the new evaluation tools; however, according to the respondents, the pilot committee has yet to develop a formal training plan.

According to the document review, the district's pilot committee did not provide evidence that an evaluator training process has been designed or is under development. Nor did the committee provide criteria by which evaluators will be selected and evaluated.

Moreover, no evidence was provided that a data-tracking process has been designed or is under development.

Using Results for Professional Development

None of the respondents commented on using principal evaluation results to inform professional development offerings.

According to the document review, there is no evidence that the principal evaluation system is being linked to a comprehensive human capital management strategy, with linkages to professional development planning, progressive discipline, or compensation policies.

Communicating Results

The district provided no evidence for the document review that a plan for pilot testing and tracking evaluation results has been developed or is under development. The district has not identified an independent evaluator or methods for gauging satisfaction, fidelity, reliability, or utility of the new principal evaluation system.

According to the respondents, the new principal evaluation system will be discussed with relevant stakeholders and district staff at the same time as the new teacher evaluation system. For more details concerning past practice in this area, see the Communicating Results section for the teacher evaluation system.

According to the document review, the district provided no evidence that a communications and stakeholder investment plan has been completed or is under development.

Summary and Conclusions

In conclusion, the respondents recognized the many strengths that the new teacher evaluation system offers. Two respondents cited the richness of conversations concerning professional growth as one of the system's greatest strengths. Also, having a tool that reflects not only what a teacher does, but also what a teacher needs to do to progress. One respondent thought that a strength of the system is that it builds upon the district's previous system. The same respondent felt that the greatest strength of the principal evaluation system is that it is modeled after the teacher evaluation tool and will be much more dynamic than the previous system. Another respondent also mentioned that the new teacher evaluation system builds upon the old model. The respondent also felt that being much more deliberate about including student data elements and measuring whether teachers are differentiating instruction based on that feedback is a great strength. The emphasis on professional growth also was cited as a strength.

The respondents also recognize that there will be challenges with implementation. Two respondents cited ensuring higher interrater reliability as a major challenge. One respondent cited the fact that the new system will be more resource intensive in a fiscally austere environment as a major challenge to the implementation and longevity of the system. The respondent also mentioned that the new system will be more time intensive for principals.

On many topics, the responses provided by the district personnel interviewed were corroborated by evidence found in the document review. However, there were several cases in which it was evident that the respondents had put a great amount of time and effort into considering different aspects of the evaluation systems but that this had not been translated into a systematic plan of action.

Wentachee School District
Case Study Report

District Overview/Description

The Wenatchee School District, located in the central part of Washington state, serves approximately 7,700 students in its seven elementary schools, three middle schools, alternative high school, 4A high school, academy for home-school students, and skills center. Its students' demographic characteristics are as follows:

- 0.52 percent American Indian or Alaskan Native
- 1.3 percent Asian or Pacific Islander
- 0.48 percent African American
- 43.2 percent Hispanic
- 51.83 percent white
- 54.94 percent eligible for free or reduced-price lunch

The district employs approximately 500 teachers, 62.5 percent of whom have at least a master's degree.

Case Study Method

This study was undertaken to help the Washington Office of the Superintendent of Public Instruction (OSPI) better understand how to support districts' implementation of Senate Bill (S.B.) 6696, which establishes new criteria for evaluating educators. For this component of the study, four individuals affiliated with Wenatchee Public Schools were interviewed in an effort to collect information concerning the process of developing and implementing the district's new teacher and principal evaluation systems. Each of the respondents has played an active role in the development of the new systems. In an attempt to gain a diverse range of institutional perspectives, we interviewed one district-level administrator, one school-level administrator, and two teachers. All responses were coded based on topic and related through the following case study.

Moreover, the district was requested to submit pertinent documents concerning the development and implementation of its teacher and principal evaluation systems. These documents were then reviewed by AIR staff who documented evidence of the district's actions.

Teacher Evaluation

Setting the Evaluation Framework and Goals

The Wenatchee district did not have an instructional framework in place prior to participating in the Teacher-Principal Evaluation Pilot (TPEP). Therefore, the district has been developing an instructional framework in parallel with its efforts to design the evaluation system for the 2011–

American Institutes for Research®

12 pilot. One interview participant felt that developing both frameworks simultaneously has been a challenge, particularly because the design of the evaluation rubrics seems to be driving the development of the instructional framework, rather than the other way around. Although still not finalized, the district has decided to base its instructional framework on the Marzano model, but is also consulting with Charlotte Danielson, particularly regarding the seventh and eighth criteria mandated by the state.

The district is still working on adapting some of the Marzano language to the criteria provided by the state. The evaluation committee is particularly focused on ensuring that the rubrics are clear and consistent, so that principals and teachers will be able to understand the intent of the language. The district has not yet determined how the instructional standards will be linked to the evaluation process, but ultimately intends to align them. However, none of the respondents thought that the evaluation results would be factored into compensation decisions or used to determine "advancement or [career] differentiation between teachers."

One respondent mentioned that compensation decisions are currently dictated by union contracts, so even if the district decided to align the pay scale or promotions with the evaluation results, it would have to bargain with the local union to make changes to contract language. Another respondent thought that evaluation results might eventually influence whether a teacher should be counseled out of the profession, if it is the case that he or she consistently receives unsatisfactory evaluations.

All respondents stated that teacher professional growth was the primary focus of the new evaluation system. Two respondents mentioned that the criteria and evidence indicators will help encourage and guide professional development conversations between principals and teachers, allowing them to target specific areas for growth, identify appropriate next steps for improvement, and align professional development support with stated goals. A related outcome will be that in moving toward a more comprehensive system with multiple measures, evaluations will gather richer information on instructional practices that can be used as feedback for teachers to improve their craft. Another respondent stated that an important objective of the new system is to create a common language for evaluation throughout the district, helping ensure that expectations for teachers are the same and that rigor is consistent across schools.

Establishing Evaluation Criteria

The state mandates that teachers receive ratings on a four-point scale for each of eight criteria established in S.B. 6696. The new evaluation system in the Wenatchee district will rate teachers as Unsatisfactory, Basic, Proficient, and Distinguished on specific instructional skills. The *Unsatisfactory* rating indicates that a teacher is not competently performing a specific skill and that "failure to fix this problem could result in non-renewal." A *Basic* rating indicates that a teacher is exhibiting some proficiency but still has to make significant improvements to meet district expectations. A *Proficient* rating characterizes effective teaching, and *Distinguished* is reserved for exemplary practices. One respondent felt that the Basic rating—which alerts a teacher to an area in which he or she is not performing up to expectations but which does not subject the teacher to discipline or non-renewal—has great potential to contribute to teacher professional growth.

Respondents indicated that the Wenatchee district has developed draft rubrics for the eight criteria provided by the state but is still in the process of choosing evidence indicators and measures to evaluate teachers on four levels for each of the rubrics. The committee also is focusing on reducing any redundancies that have developed across different criteria. The district is struggling particularly with developing objective measures for the state's seventh criterion regarding school community.

The document review, which was conducted later in the district's planning process, indicates that the Wenatchee district developed a comprehensive and well-articulated scoring rubric for its teacher evaluation model. The rubrics consist of several evidence indicators with which evaluators could rate teachers at different levels. However, the document review found that the summary document created by the district to describe the rubrics includes vague descriptions of each level of the four-point scale. Most of the descriptions reflexively mention the level's name, such as requiring that teachers at the Basic level demonstrate "basic" competency at that skill.

S.B. 6696 mandates that all classroom teachers—but not support staff such as librarians and counselors—be evaluated using the new evaluation criteria. It was unclear to respondents whether the system will eventually be adapted to include additional staff as well. One respondent thought that "the ultimate goal would be to have the same language either applied to everyone or adopted for everyone."

The district currently conducts evaluations more frequently for provisional teachers than contract teachers. Provisional teachers are observed four times annually and receive evaluations of their performance 90 days into the school year and again at the end of the year. Contract teachers are observed twice annually and evaluated once. The committee has not yet determined whether the frequency of evaluations will change in the new evaluation system. Two respondents mentioned that state law actually dictates how frequently teachers are evaluated, so the committee is not likely to increase the number of evaluations unless the state mandates a change. Yet another respondent assumed that the frequency of evaluations would be driven by which measures the committee chooses to evaluate effectiveness. The document review did not find any information on how frequently teachers will be evaluated.

Measuring Performance

The committee still has not determined which measures will be used to evaluate teachers, both core and noncore. All of the respondents were certain that the new evaluation system would include multiple measures, especially because observations alone will not suffice to address all of the state-mandated criteria. One respondent indicated that members of the committee still had "differences of opinions" regarding which measures should be used and that it was a "hot topic," but several measures had been discussed as possibilities. Two respondents stated that the committee is concerned about whether value-added scores can be "tied solely to teacher-controllable performance." Another respondent echoed these concerns and also mentioned that the measure can produce inconsistent results from year to year, explaining:

> Hypothetically, I'm a teacher. … This year I had a great group of kids … we worked hard. … But the next group of students I get the next year, they're not the same kids. So how do you pay a teacher based on that from year to year to year? ... They might receive

> a stipend or a bonus amount because they performed so well last year, but then they get another group of kids that's not the same group of kids who doesn't show the same amount of growth that the class did last year.

The document review, which was conducted later in the district's planning process, found that the teacher evaluation model will incorporate several measures of effectiveness, including measures of student growth, classroom observations by an administrator or peer, instructional artifacts, teacher portfolios, and teacher self-report measures. These measures are supported by a solid research base, in addition to being valid, reliable, and linked to student achievement. The rubrics are described clearly and in understandable language. The rubrics allow evaluators to assess practitioners on all integral aspects of teaching, including teacher assignment and placement, working conditions, teacher participation in committees, student demographics, and teacher planning time.

Training for Evaluators

All of the respondents emphasized that evaluator training will play a critical role in ensuring the successful implementation of the new evaluation system, but the district has not yet determined what the training sessions will entail. One respondent stated that the district will likely train the teachers and administrators together for the pilot. Another respondent mentioned that the training sessions would probably require observers to rate videos of lessons so that facilitators could check for interrater reliability. There are also plans to use monthly professional growth meetings that the district currently holds with its administrators to continue training on assessments and interrater reliability throughout the pilot year, which would amount to two hours a month. The respondents emphasized the importance of encouraging consistency in the evaluations across schools and administrators. The document review did not surface any information on plans for evaluator training because this piece of the evaluation system is still in development.

Using Results for Professional Development

One of the district's main goals in designing the new evaluation system is to promote teacher professional growth, so the committee does intend to link the evaluation and professional development systems. However, the district has not specifically defined what this relationship will look like, in part because it has not yet finalized the rubrics and measures that will be used to evaluate teachers. Respondents indicated that the district plans to use the data and information gleaned from the evaluations to help plan for future professional development opportunities. One respondent explained that the professional development piece would likely continue to develop and change to meet any needs identified through the pilot. The document review reflected this stage of development and did not surface any specific information on using evaluation results to inform professional development.

The district has begun to train teachers on how to use student growth data to drive changes in instructional practices. The district implemented a system called Data Director one year ago, for which teachers already are receiving training. Teachers are instructed on how to analyze measures of student growth and adapt instructional practices to address any revealed gaps in achievement. Several respondents indicated that the new evaluation model's emphasis on student

academic growth will be supported by these kinds of data systems, along with other possible measures such as the state achievement index scores. When the document review was conducted, this component of the plan for the new teacher evaluation system was still in development.

Communicating Results

The district has prioritized efforts to keep stakeholders up-to-date on developments in the design of the evaluation pilot. The district developed a page on its website devoted solely to plans for the new evaluation system where minutes from committee meetings are posted. In addition, the association president has included updates on the evaluation pilot in newsletters to members. The district does the same for a monthly newsletter that is released to the public.

The document review confirms that newsletters are disseminating information to stakeholders about the evaluation design process. The newsletters include a list of the contact information for district representatives who are willing to answer questions and address concerns. However, it is not clear how or whether the feedback from stakeholders will be incorporated into the design or revision of the teacher evaluation plan in any formal way.

The district intends to pilot the evaluation system in the fall of the 2011–12 school year. This deadline will require that the committee spend the rest of the 2010–11 school year finalizing instruments, choosing evidence indicators, clearly defining rubrics, and settling on measures to evaluate teachers on each of the criteria. The district has not determined how many schools will pilot the new evaluation system in the 2011–12 school year, nor how many grade levels or staff members will be involved. The 2012–13 school year will allow for more fine-tuning based on lessons learned in the pilot year. At the time that the document review was conducted, the district was still working on finalizing the timeline for implementation. However, there was a reference in a newsletter that the pilot would be launched in the fall of 2011.

Principal Evaluation

Setting the Evaluation Framework and Goals

Respondents explained that the Wenatchee district has been examining the Association of Washington School Principals (AWSP) model for effective leaders to help construct its leadership framework. The document review also found evidence that developing rubrics were aligned with the Interstate School Leaders Licensure Consortium's *Standards for School Leaders* and the state professional standards.

The goals for the principal evaluation system include promoting consistency across schools in the expectations for principal effectiveness, as well as increasing accountability. Similar to the teacher evaluation system, respondents felt that evaluating principals using the four-tiered rubrics for the eight state criteria would help principals identify areas for growth and focus professional development efforts on unique individual needs.

Establishing Evaluation Criteria

To develop the rubrics under the eight criteria for the principal evaluation system, the committee did research on each criterion and examined best practices from around the country. Committee

members then brainstormed which aspects of the criterion they felt to be the most important and built rubrics off of that foundation. They examined the AWSP's standards for an effective principal or school leader and also looked at best practice evaluation models from around the country, including those developed in North Carolina and Jefferson County, Colorado. The committee used those examples to choose specific language for the rubrics.

As with the teacher evaluation system, the principal evaluation system will be based on a four-point rating scale with the Unsatisfactory, Basic, Proficient, and Distinguished categories. The expectations for proficiency in each of those categories are the same as in the teacher evaluation system. Although Wenatchee intends to evaluate all principals in the new evaluation system, the district has not yet determined how frequently principals will be evaluated throughout the year.

Measuring Performance

The principal evaluation committee has not settled on which measures it is going to use to evaluate principals, but the committee feels strongly that it will include multiple measures of effectiveness. There are currently no plans to use student value-added scores as a measure in the principal evaluation system, but there is a category of the criteria that addresses closing achievement gaps. The evaluation system is more focused on principal behavior than student behavior. The committee is hesitant to require principals to submit portfolios because of the potential burden it may place on their time. However, one respondent mentioned that the current evaluation system uses artifacts and that the new evaluation will likely use artifacts as a measure for several of the criteria.

The document review, conducted later in the district's planning process, found that the Wenatchee district had developed a clear format for the evaluation rubrics and had begun to identify multiple measures of rating principal effectiveness. Although the rubric generally includes observable and measurable language, the document reviewer found several instances f high-inference or ambiguous language.

Training Evaluators

Neither the interviews with stakeholders nor the document review found any references to a finalized evaluator training process. The project timeline indicates that it is currently under development.

Using Results for Professional Development

Interview respondents indicated that they were very excited about the possibility of using the results of the evaluation to help target professional development for administrators. One respondent explained:

> [Principals] come to the job with different skills and abilities. [The goal of the evaluation system will be to] enhance the skills that [principals] have and [to] focus on areas that [principals] can grow. And that might be different from principal A as opposed to principal B. So it would be more [sic] focused, more differentiated staff development opportunities.

The same respondent also discussed the possibility of developing internal training or professional development units specific to each evaluation rubric. In this case, there would be specific resources within the district for a principal who is struggling with a particular criterion. This part of the evaluation system is still in development, so the document review did not find any description of how the principal evaluation system will be linked to principal professional development planning or how the district will track data from evaluations.

Communicating Results

The document review surfaced several well-developed documents indicating that the district has engaged consultants, teachers, district-level administrators, union representatives, and school principals in formulating the new principal evaluation system. The documents are primarily informational. The district also created several presentations for stakeholders and the general public, which provide detailed information on the district's progress toward TPEP goals. The Wenatchee district has not summarized the communications plan for the principal evaluation system in a single document.

Summary and Conclusions

All four respondents in the Wenatchee district felt that the major strength of the new evaluation system is its focus on improved professional growth. Respondents mentioned that the more specific and meaningful feedback will make it easier for teachers and principals to receive constructive criticism, define professional goals, and focus professional development efforts. A related strength was the new system's use of multiple measures of effectiveness, rendering it more comprehensive than the current system. One respondent also felt positively about the fact that the new evaluation system is research based and focused on student growth. This respondent summarized as follows:

> What I really believe is that it's more focused professional development, it is for both administrators and teachers, it really forces conversation about what effective teaching is, and it also looks at multiple ways to do the measuring of that. So it's not just a check-off sheet.

However, respondents noted several logistical challenges in the planning process. Two respondents mentioned that designing the new evaluation system is incredibly time consuming; both worried about completing the instructional framework and evaluation tool during the current planning year (2010–11). Another respondent mentioned that some of the rubrics in development seemed to overlap, so the committee was still struggling to attain real clarity for the teacher and principal evaluation criteria. Finally, two respondents felt that the guidance the state had provided to districts seemed rather open-ended initially, and somewhat of a moving target recently. They were uncertain about whether the state had recently decided to change course and more clearly define rubrics and benchmarks for districts, which could render mostly moot the intensive work that the district had done so far in the process.

Respondents also stated several concerns about implementing both the pilot, and eventually the new evaluation system, districtwide. One of the most frequently cited challenges was ensuring interrater reliability among the district's 18 administrators and 500 certificated staff through

intensive training. In addition, the committee is still not sure how it will be able to guarantee that administrators have time to conduct effective and comprehensive evaluations of all staff members. Finally, one respondent also mentioned that there will likely be pushback within the system from faculty and administrators.

Overall, however, the respondents felt that, if implemented properly, the evaluation system would be "immensely valuable" in terms of tying professional growth directly to evaluation criteria and providing more specific feedback to teachers and principals. This is the main goal motivating the district's efforts to design and implement the new evaluation system pilot.

APPENDIX E

Washington TPEP Statewide Data Analysis Report

AMERICAN INSTITUTES FOR RESEARCH®

Education,
Human Development,
and the Workforce

Washington's Teacher/Principal Evaluation Statewide Data Analysis Report

Summary of Findings

Judy Stewart, Consultant to AIR

June 2011

Contents

Page

1. Executive Summary 1

2. Methodology 4

3. Make Improved Teaching and Learning the Primary Goal of the Educator Evaluation System 9

4. Develop Appropriate Rubrics and Multiple Rating Levels and Collect on Multiple Measures 12

5. Observe All Educators at Least Annually 14

6. Invest in Evaluator Training .....

7. Conclusion .....

Appendixes 34

- Appendix A. Survey Tables 34
- Appendix B. Educator Forum Full Report 21
- Appendix C. List of Educator Forums 34
- Appendix D. Telephone Interview Protocol 34
- Appendix E. Statewide Data Emerging Themes .....

1. Executive Summary

Introduction

The single most important in-school factor contributing to a student's success is his/her teacher.[1] Likewise, principal leadership contributes significantly to the learning environment in which students and teachers thrive.[2] These research-based conclusions have opened the floodgate for further research, particularly about the characteristics of an effective teacher and administrator, and how to evaluate for effectiveness. While the research is inconclusive on both scores, significant advances have been made to inform what constitutes an effective educator and how to design more rigorous and reliable educator evaluation systems. Crucial to these advances is the acknowledgment that teaching and leading are complex enterprises that require complex assessment systems to capture effectiveness in individuals, classroom practice, and leadership settings.[3]

This report summarizes Washington teachers' and administrators' perceptions about the educator evaluation systems currently being implemented in their districts, and the impact of Senate Bill 6696, which established new criteria for evaluating educators. Data were collected from three sources: (1) an online questionnaire, (2) focus group discussions at regional educator forums, and (3) one-on-one telephone interviews. Appendix E of this report highlights the overall themes that have emerged from all of the data collected. The findings inform a set of considerations. These considerations should help OSPI support districts' implementation of SB 6696.

Considerations

1. **OSPI should clearly articulate the primary goal for the educator evaluation system; that goal should be to improve teaching and learning.**[4] A well-articulated purpose statement should find fertile ground among educators. Teachers and administrators overwhelmingly indicated the need for an instructional framework to guide the evaluation system and tools for improvement and growth. Such tools should include an instructional framework to guide the evaluation system. The instructional framework should reflect best classroom (and leadership) practices. The framework should be routinely revisited to ensure that it represents the most up-to-date, evidence-based knowledge of good practice. In addition, clearer criteria are needed to guide evaluations, along with measures of effectiveness that are both credible and trustworthy, more defined rating systems, and better alignment between evaluation results and professional development.

---

[1] D. K. Cohen, S. W., Raudenbush, & D. L. Ball, Resources, instruction, and research. *Educational Evaluation & Policy Analysis, 25*,2 (2003), 119.

[2] Strategic Management of Human Capital, *Taking human capital seriously: Talented teachers in every classroom, talented principals in every school* (Madison, WI: Author, 2009). Accessible at http://www.smhc-cpre.org/resources/

[3] L. Goe, *Evaluating teacher and principal effectiveness* [PowerPoint presentation, 2001]. http://programs.ccsso.org/projects/Membership_Meetings/LEG/documents/Laura%20Goe%20CCSSO%20Leg%20Conf%202011.pdf

[4] ibid.

2. **OSPI should support the development of appropriate rubrics and multiple rating levels and collect data on multiple measures.** The rubrics should spell out levels of performance and progression within the framework. They may also include the kinds of evidence or artifacts needed to demonstrate performance at each level. The evaluation system should move beyond a two-level rating system to further differentiate between levels of educator effectiveness. Districts should have some flexibility in determining its mix of measures. The menu might include classroom observations, peer observations, student growth data, student work, lesson plans, attendance at committee meetings, and formal and informal demonstrations of leadership in such activities as professional learning communities or peer mentoring. Administrators' measures should give some attention to school improvement plans and parent and community outreach.
3. **OSPI should support the observation of all educators at least annually.** Many educators would like to see a system where all colleagues—teachers and administrators alike—experience classroom (or school) observations annually. This would allow all educators to identify areas of strength and for improvement, and consistently promote the message of professional growth to improve student learning. If OSPI elects this option, it will have to think carefully about how to move teachers from the Professional Growth Option to annual observation review. OSPI will also need to manage the additional time burden on principals and other evaluators, especially in more challenging schools.
4. **OSPI should invest in evaluator training across all levels—teachers, assistant principals, principals, and superintendents.** Common, high-quality training at the district and school levels should address how and how often to conduct meaningful evaluations. In addition, OSPI should give careful consideration to who should conduct evaluations.

In addition to these considerations, OSPI must plan for the union's role in carrying out SB 6696. Responses from the field reveal that the decision-making role of the local teachers union varies across the state. OSPI will need to develop a strategy to engage the local teachers unions in the educator evaluation reform process.

Conclusion

Washington, like many states, is recalibrating its educator evaluation system. The legislature has signaled that it wants change. It is encouraging that data collected for this report show that many Washington educators and administrators also support change. But they want it to be thoughtful, offering some degrees of freedom to local districts, and well supported with training. In many respects, research can help meet these expectations. Research has moved the needle on how to identify an effective teacher and leader. The good news, then, is that Washington's educator evaluation system can be improved. The not-so-good news is that research has yet to produce definitive evidence about the validity of certain measures to estimate teacher and administrator effects. As a result, the call for new educator evaluation criteria must be tempered with what we know about good classroom practice, instructional and managerial leadership, and professional development. Education stakeholders would be wise to see changes to Washington's educator evaluation system not as an endpoint but as a process toward developing a system that explicitly contributes to improvements in teaching and learning.

2. Methodology

This report is informed by three measures: (1) questionnaire data; (2) educator focus groups at regional forums; and (3) telephone interviews. A description of each measure follows.

Survey Data

An online questionnaire was sent to Washington state superintendents, principals, and teachers to gather information on the current state of teacher and principal evaluation in Washington districts. The questionnaire asked educators to provide basic information about their teacher and principal evaluation systems, including the evaluation purposes and criteria, evaluation processes, performance ratings, measures used to determine effectiveness, frequency of evaluations, training for evaluators, use of evaluation results, and connections between the evaluation and ongoing professional development.

Three separate questionnaires—one for each group of educators—were e-mailed to approximately 271 district administrators, 2172 principals, and 48,928 teachers. To encourage participation from educators, Washington OSPI posted the questionnaire link on their website and provided information describing state-level efforts to support implementation of Senate Bill 6696. In addition, other organizations from the WA TPEP Steering Committee representatives posted the link and/or sent it on to educators that they represent. Also, the questionnaire was kept brief so that it would not take more than 10 minutes to complete. In total, 111 superintendents and/or their designees, 263 principals and assistant principals, and 6,081 teachers representing 250 of 295[5] (84.7%) public school districts in Washington responded to the questionnaire. Response rates for each category of educator were 41.3 percent, 12.1 percent, and 12.4 percent for administrators, principals, and teachers, respectively. See Table 1 for an itemization of these results.

Table 1. Response Rates for the Questionnaire

Respondents	Responses	Response Rate
District leaders (superintendents and/or designee)	111 (out of 271)	41.3%
Principals and assistant principals	263 (out of 2,172)	12.1%
Teachers	6,081 (out of 48,928)	12.4%
Overall district representation	250 (out of 295)	84.7%

Feedback on the questionnaire from district administrators was used to document the current set of policies and practices in Washington districts. To promote a high response rate and help ensure accurate reporting, a prenotice e-mail was sent to district superintendents with instructions for completing the questionnaire. Specifically, the e-mail summarized the purpose of the questionnaire and requested that superintendents, upon receipt of the survey link, forward the questionnaire "to the appropriate administrator overseeing the district's educator evaluation

[5] Eight districts and one consortium of districts are currently developing new evaluation systems as part of a statewide pilot program. These districts were not included in this survey.

process." Several days later, a second letter was sent to superintendents. This letter included the questionnaire URL, along with a brief reminder to forward the questionnaire to the appropriate administrator. Within one week of sending the e-mail, a research assistant followed up with each district via telephone to verify that the superintendent received the e-mail and to remind him or her to either complete the survey or forward it to the appropriate administrator to complete. In total, 111 administrators from 102 districts responded to the questionnaire.

Among the 102 districts that responded to the questionnaire, 94 submitted feedback from 1 administrator, s6ix submitted feedback from 2 administrators and 2 submitted feedback from 3 administrators. Within the 8 districts in which more than 1 administrator provided feedback, a descriptive analysis by district showed that administrators' feedback within districts matched on nearly all the items.[6] In rare cases where administrators within the same district endorsed conflicting response options (e.g., one endorsed "yes" and one endorsed "no"), their responses were not included in the analysis. For "check all that apply" items, all options selected by each individual administrator were included in the full set of options endorsed by the district.

Table 2 presents the characteristics of districts that responded to the questionnaire and compares characteristics of district respondents to nonrespondents. This comparative analysis was completed to determine the extent to which district respondents reflect the population of districts statewide. For this analysis, "Not Responding" includes all districts from which no district administrator completed more than 50 percent of the survey questions. Results are presented for the following district-level characteristics: locale, minority or "nonwhite" enrollment, percentage qualifying for free or reduced-price lunch, adequate yearly progress (AYP) status, and four-year on-time graduation rate.

Of the 102 districts represented in responses, more than half are located in rural areas, followed by 19.6 percent from suburbs, 13.7 percent from towns, and just under 10 percent from cities. About three quarters of these districts serve a student body with less than 40 percent who are minorities, and slightly over half serve more than 40 percent who qualify for free or reduced-price lunch.[7] Two thirds of the districts responding met AYP in 2009, and about 83 percent graduate more than 60 percent of their students. Percentage differences between responding and nonresponding districts are relatively small, ranging between –7.0 percent to +3.7 percent.

[6] In three cases, one of the multiple respondents within a district left the majority of items unanswered.

[7] Schools with more than 40 percent of students eligible for free or reduced-price lunch qualify as "schoolwide program" schools, which have more flexibility in the use of Title I funds and the delivery of services.

Table 2. Differences in Characteristics Between District Respondents and Nonrespondents[8]

		Districts Responding (*N* = 102)		Districts Not Responding (*N* = 193)		Difference in Percentage
		Number	Percentage	Number	Percentage	
Locale	City	9	8.8%	10	5.2%	3.6%
	Suburb	20	19.6%	32	16.6%	3.0%
	Town	14	13.7%	40	20.7%	–7.0%
	Rural	59	57.8%	111	57.5%	0.3%
	TOTAL	102	100%	193	100%	
Minority enrollment	Less than 40% nonwhite minority	77	75.5%	140	72.5%	3.0%
	Greater than 40% nonwhite minority	25	24.5%	53	27.5%	–3.0%
	TOTAL	102	100%	193	100%	
Free/reduced-price lunch (FRPL)	Less than 40% eligible for FRPL	46	45.1%	88	45.6%	–0.5%
	Greater than 40% eligible for FRPL	56	54.9%	105	54.4%	0.5%
	TOTAL	102	100%	193	100%	
AYP status	Met AYP in 09	66	64.7%	132	68.4%	–3.7%
	Did not meet AYP in 09	36	35.3%	61	31.6%	3.7%
	TOTAL	102	100%	193	100%	
Graduation Rate	More than 60% graduation rate	85	83.3%	155	80.3%	3.0%
	Less than 60% graduation rate	2	2.0%	5	2.6%	-0.6%
	N/A (Only Elementary Grades)	15	14.7%	33	17.1%	-2.4%
	TOTAL	102	100%	193	100%	

A similar process for disseminating the questionnaire to district administrators (described above) was used for teachers and principals. First, a prenotice e-mail was sent describing the purpose of

[8] *Source*: Data obtained from the Washington OSPI website at http://www.k12.wa.us/DataAdmin/default.aspx. Data on locale, minority enrollment, and free or reduced-price lunch was reported in the 2010–11 school year. Graduation data was reported after the 2009–10 school year, and AYP status was reported after the 2008–09 school year.

the questionnaire and how results would be used, followed a few days later by a second e-mail with the questionnaire URL and further encouragement by a research assistant, who called each district superintendent's office to complete the questionnaire.

Feedback from teachers and principals was used to examine differences between perceptions of evaluation practice among teachers or principals and district-level staff overseeing the evaluation process (see Appendix A for survey tables that compare teacher and administrator responses). Disagreement may suggest that (1) evaluation policies and guidance are interpreted differently among district and school level staff; (2) the evaluation practices implemented in schools do not align with district policy and guidance; or (3) district policy and/or guidance about how evaluations should be carried out are unclear or unavailable. To determine the extent of agreement between teachers and district administrators, researchers counted the total number of teachers within a district who selected the same response as their district administrator. Next, the total number of teachers who agreed with their respective district administrator was divided by the total number of teachers who responded to the questionnaire item.

District feedback on teacher and principal evaluation policies and practices represents the group of districts that responded to the questionnaire and should not be generalized to all Washington districts. Similarly, results reporting the extent of agreement between teachers/principals and district administrators should be interpreted with caution. The percentage of teachers and principals who responded within a district varies widely and does not necessarily reflect the perspectives of all teachers or principals districtwide. Furthermore, because the overall rate of responses among teachers and principals are relatively low (approximately 12 percent), results do not necessarily reflect the percentage of agreement between teachers/principals and their administrators statewide.

Educator Forums

Educator forums were held between February and March 2011 (see Appendix B for the full report). The purpose of the forums was to collect participants' views on (1) the current teacher and principal evaluation system in Washington, (2) the state's recent education reforms and pilots, and (3) what features and characteristics should be included in the new evaluation system. Teachers, principals, superintendents and other district officials, education directors, school board members, and parents and community members participated in 10 forums hosted by ESDs across the state (see Appendix C). The approximate number of participants in total was teachers (98); principals (85); district administrators (71); community members and parents (24).

Participants were recruited through an invitation process to ensure that teachers, principals, and community members from a diversity of backgrounds and regional locations were included, while at the same time not excluding any educators in the state who wished to participate. The project team reached out to all ESDs to identify 20 to 40 invitees that represent the diversity of districts in their region in terms of performance levels, poverty levels, size and urbanicity, and the diversity of grade and subject levels taught in the schools. In addition, members of the Steering Committee supported the invitation process and ensuring a well-rounded distribution of participants at the forums that reflected the makeup of the Washington TPEP Steering Committee.

Each forum included a 20-minute overview and introduction, a brief presentation on current developments in Washington's educator evaluation process, a large-group question-and-answer session, and a facilitated, one-hour small-group discussion organized by stakeholder group. Participants were also given an exit survey.

Although facilitators prompted and asked participants about both teacher and principal evaluation, participants' knowledge and interest was greater on the topic of teacher evaluation, and therefore the preponderance of discussion in the focus groups centered on teacher evaluation. Although this report presents findings on both systems, the extensiveness and depth of data available on views of principal evaluation is somewhat diminished in comparison with the data on teacher evaluation.

Transcripts were coded and analyzed using NVivo 9, a software program designed for systematic analysis of qualitative data. Initial themes or codes were based on the focus group protocol and the 10 topics identified in the TPEP Technical Proposal. The initial themes were refined as new themes and ideas emerged upon further examination of the data.

Telephone Interviews

A protocol was developed to guide one-on-one telephone interviews with teachers and administrators (see Appendix D for the interview protocol). Prospective interviewees were drawn from the list of individuals who completed the electronic survey, indicated that they would be willing to participate in a follow-up telephone interview, and were not participating in any of the state's pilot educator evaluation projects. Interviewees were randomly selected. Every effort was made to match non–pilot site administrators and teachers by school district to ensure feedback was collected from an administrator and teacher in the same district. Where an exact district match could not be made, a match was made on geographic locale (e.g., mid-size suburban administrator and mid-size suburban teacher; distant rural administrator and distant rural teacher; large city administrator and large city teacher).

Twenty seven interviews were conducted in April and May 2011, representing 12 teachers and 15 administrators across all locale distinctions. Interviewees were asked about their perceptions of the evaluation system currently being implemented in their district. Interviews generally lasted between 30 and 45 minutes.

3. Make improved teaching and learning the primary goal of the educator evaluation system

When asked, "What is the primary goal of your district's educator evaluation system?" Washington teachers in telephone interviews offer a range of responses, as varied as complying with the teacher employment contract, identifying weak teachers and providing support, identifying especially strong teachers, and supporting professional growth. Seventy eight percent of district administrators who responded to the online survey said the primary purpose of their district's teacher evaluation system is compliance. In telephone interviews, principals also cite a range of responses for evaluating principals and other administrators, but with less variation. Principals tend to say the purpose of their administrator evaluation system is to set individual, professional goals that are aligned with ISLLC standards and district goals for student improvement and to address areas of concern.

Nevertheless, Washington educators want tools for improvement and growth. Interview and focus group data consistently point to educators' desire to engage in rich, substantive dialogue with their peers and evaluators to inform and improve their practice. The comments of one focus group participating principal capture this well:

> I think a system should cause conversations around teaching and learning and even help your teachers that are above average become better. I think the biggest challenge is to move a good teacher to great. To be able to have something specific, to have them set a goal and you can collect data for it. That's what we enjoy doing. I mean, a system that is a formatted structure for rich conversations around the learning is really what we want. We need a system that works for reflective teachers, the teachers that say I want to get better. And we're their coach and we're supposed to help them get better. Whatever the system is, the purpose should be to increase the teaching and learning in the classroom. (Appendix B, p. 24)

OSPI should make clear the purpose of the educator evaluation system; and that purpose should be to improve teaching and learning. Teachers, principals, administrators, parents and community members should hear this message consistently, especially during this period of transition. Stemming from this single point of purpose, OSPI will be able to make more strategic decisions about the framework(s) it will consider, criteria, ratings, measures, training, and how to link the evaluation results to meet professional development needs.

The data suggest that some districts already do tie their evaluation results to professional development needs. Teachers, especially those in larger districts, are given course offerings, peer mentoring, and whole-school professional development and other opportunities. Smaller districts need greater support to match individual or clusters of teachers' needs to professional development opportunities. Educators, as evidenced in interviews and focus group findings, support aligning evaluation results to professional development. However, time for professional development remains an obstacle. As one focus group participant pointed out:

> [U]nless you figure out a way to adjust time in the school system, not much, frankly, can change. It just can't. …[T]he key factor, as it seems to me, is the collegial time away from students. Our system is built to put us in front of students as many hours and minutes as possible and get better on our own time. … [a]nd that's not really very effective. For the kind of practice we're talking about, we need to figure out a way…to give us that kind of time to build good, solid lessons that mean when we are in front of kids they're getting the very best we can give them. Somebody who used to work here used to say what gets in the way is the "dailyness" of teaching. "I can't do that. I'm here today. If I wasn't here today I could probably adjust that but right now I've got the next class period coming in. So I hope I can get to it tomorrow." That's the reality of our job. It's unfortunate. And I think if we can change that or find a way to manipulate that, I think we have a way to really make some serious progress. (Appendix B, p. 29)

Few districts explicitly tie their teacher or principal evaluation system to assignment decisions, promotion/advancement, or hiring/firing. No district reported a link between its teacher or principal evaluation system and compensation. Implicitly, however, both educator evaluation systems are occasionally tied to identifying leaders. For example, principals do recognize the shining stars in their teacher ranks and make mental note of their efforts. In some cases, stand-out teachers demonstrate their leadership by their own professional growth goals and reflection narratives. Repeatedly, principals said that their strongest teachers take the professional growth plan option very seriously and apply it to improve their craft.

The effort to design an effective educator evaluation system is, in many ways, hindered by the absence of a clear definition of effective teaching. From the federal level to researchers, one agreed-upon definitive description of a highly effective teacher has not been put forth, although broad definitions do exist in isolation. For example, the National Comprehensive Center for Teacher Quality offers a five-point definition (http://www.tqsource.org/publications/RestoPractice_EvaluatingTeacherEffectiveness.pdf).

Without one definition and a common language, states, districts, and individual teachers and leaders often operate without a common vocabulary to describe their practice and the evaluation of that practice. This might help account for the mismatch between teachers' and administrators' response to the survey question: Does your district have a clear definition of effective teaching? Fifty-two percent of teachers and administrators agree that their district has a clear definition of effective teaching. Forty-eight percent of district administrators believe their district has a clear definition of effective teaching. Only 35 percent of principals and district administrators agree that their district has a clear definition of effective principal leadership (see Appendix A for results).

Moreover, only 24 percent of district administrators surveyed said their district has adopted an instructional framework or set of standards for assessing teachers. Among teachers and administrators, there is a 34 percent agreement rate with this statement. Principals and district administrators agree at a rate of 42 percent that their district has adopted an instructional framework or set of standards for assessing principals.

Some educators participating in telephone interviews suggested a process should be put in place to ensure the continuous review and updating of instructional frameworks so that they reflect the most current knowledge base about good teaching and leading. Regrettably, several educators in telephone interviews reported a five-year gap or longer since the district engaged in any review of its instructional framework for teaching.

In telephone interviews, Charlotte Danielson and Bob Marzano were the most frequently referenced influencers on district instructional frameworks for teachers, and ISLLC standards were the most frequently referenced for assessing principal practices. OSPI might want to consider building upon the work of Danielson and Marzano and the cross-state awareness of these models in refining its educator evaluation system. At the very least, OSPI should carefully consider how districts, steeped in Danielson and Marzano, might respond to a radically new framework for assessing instructional or leadership practices.

4. Support the development of appropriate rubrics and multiple rating levels, and collect data on multiple measures

Rubrics

The data suggest spotty use of rubrics to evaluate teachers and principals. In fact, less than 25 percent of administrators responding to the online questionnaire report the use of rubrics to evaluate teachers. Developing appropriate rubrics should help align the instructional framework to demonstrated practice in order to inform educator evaluations. They can also help spell out levels of performance and progression within the framework. In addition, rubrics can reference the kinds of evidence or artifacts needed to demonstrate performance at each level.

Currently, and in accordance with level of experience and time in district, many teachers have the option to be evaluated using a long form, short form, or professional growth plan. The details of these options are well known to OSPI. According to both telephone interview and forum data, however, few teachers experience these options with the support of rubrics. Both interview and forum data suggest that rubrics that help identify areas of strength and areas for improvement might foster the kinds of conversations teachers and principals say they value. Rubrics can provide a common vocabulary for best practices. Hence, well-designed rubrics that are fitted to the instructional framework can lead to more meaningful conversations about evidence-based practices for teaching and leading. Principals interviewed by telephone agree rubrics can help flesh out the ISLLC standards for day-to-day practice. They can also be used to inform the narrative report written by principal evaluators, as well as principals' own reflections on their performance.

Because so few districts appear to currently use rubrics, OSPI might consider explaining to districts the importance of using rubrics. As noted earlier, many districts appear to be especially familiar with Charlotte Danielson and Bob Marzano's work, respectively. OSPI might choose to support districts in aligning their instructional framework and rubric development

Rating Levels

Telephone and focus group data strongly suggest the need for multiple rating levels. The current practice in many districts of a "satisfactory" or "unsatisfactory" rating finds little support in the field. As one focus group participant said,

> If anyone tells you they're getting better at teaching from hearing satisfactory or unsatisfactory, they're lying. Really, how do you get better from that? I really feel like currently with the walk-throughs we have going on, the principal is in the building more often. Informally, I feel it's more of those conversations that lead toward growth, more than "satisfactory." At any time, the formal process is just the way it's set up in our contract. It's not a growth tool. (Appendix B, p. 34)

Another focus group participant added,

> When I moved up to the tenure track it was once a year and same kind of checked yes "you did it, no you didn't," and of course I've gotten all satisfactory, which is very unsatisfying because I happen to think I'm exceptional, and I would like to get credit for what I do exceptionally well. Just being told I'm satisfactory is not satisfying at all and it doesn't tell me where I need to grow and that's another concern that I have about this process. (Appendix B, p. 34)

Measures

Telephone interviewed administrators and teachers reported wanting some flexibility in determining the mix of measures used in district's evaluation systems. The menu might include classroom observations, peer observations, teacher value-add scores, student growth data, student work, lesson plans, attendance at committee meetings, and formal and informal demonstrations of leadership in such activities as professional learning communities or peer mentoring. For principals, focus group participants emphasized measures that capture adherence to school improvement plans, and perceptions data on communication and outreach to parents and community members.

Telephone interview and focus group data suggest OSPI plan to maintain the reflection component currently in its professional growth option. Teachers and administrators benefit from the opportunity to reflect upon their own practice and to benchmark their progress over time.

Telephone interview data cautiously suggest OSPI consider including student growth data as one measure for educator effectiveness. OSPI will need to work hard to assure teachers and administrators that student data is being used for goal-setting to promote professional growth and not as a punitive, "gotcha" measure. Furthermore, teachers and administrators will need assurances that any use of student growth data controls for nonschool factors (e.g., student motivation, low-income household, student mobility) that may influence student test scores. The validity of student growth data to help determine teacher and administrator effectiveness is one area in which the research is currently still in development and requires more time, support, and resources to produce final conclusions.

Telephone interview and focus group participants overwhelmingly suggest OSPI wait to include teacher value-added scores, unless in pilot educator evaluation settings. Research shows teacher value-added scores tend to capture classroom and not individual teacher effects.[9] Moreover, they do not provide data that will help a teacher improve his or her classroom practice.[10] Add to that, that principals and teachers in telephone and focus group interviews expressed deep concern that value-added scores might close the door on professional collaboration. If the goal of the educator evaluation system is to improve teaching and learning, then, once again, current research on teacher value-added measures is inadequate to meet the expectation of a single source of data.

---

[9] Goe (2011), *Evaluating teacher and principal effectiveness.*

[10] Ibid.

5. Support the observation of all educators at least annually

The mainstay of the teacher evaluation system is classroom observation. Observations can take the form of brief principal walk-throughs (usually not attached to the formal evaluation but part of that rich discussion and feedback that teachers value) or official observations that can include a scheduled pre- and postobservation discussion. As a result of teacher, principal, and union decisions, teachers can be observed multiple times a year, once a year, or not at all. As one telephone interviewee put it, "There is a belief that after a certain number of years teaching, you know what you're doing." Principals, then, tend to focus on those teachers who need additional support—often those new to teaching. OSPI may want to push for moving all teachers onto an annual cycle of classroom observation. This would dovetail well with OSPI's stated mission to improve teaching and learning. It would require OSPI to plan carefully, however, to move teachers on the professional growth option to annual observation review. It would also require OSPI to consider how principals in more challenging schools will manage the time burden associated with more evaluations. As one principal put it:

> I think I hear a couple of us saying that [the observation] needs to be more than just the principal. I also wonder that to do it every year… principals working in my high-needs schools are working their tails off and they never give instruction what it's due and they know that and they feel terrible about it. But we haven't yet figured out what we can take off their plate in order to make that happen. A kid gets busted with a weapon, there goes your whole day. So what can you take off their plate so they can be part of it? (Appendix B, p. 60)

Data collected for this report suggest that classroom- and school-based observations are inconsistent in timing, value, and person conducting the evaluation. As noted, some teachers report frequent observations that result in timely, constructive feedback and ongoing dialogue. Others report frequent walk-throughs that result in dialogue. Still others report sporadic—maybe once or twice a year—observations that yield limited, less useful feedback and infrequent dialogue. This is especially the case for administrators. Teachers and administrators are hard-pressed to adjust and monitor their instructional and leadership practices when the feedback is out of date or loosely tied to instructional or leadership challenges.

Similarly, focus group and teachers and administrators interviewed by telephone often report that they find it difficult to improve their practice when feedback comes from different individuals—at times offering conflicting comments. The data reveal no consensus about who should conduct the evaluation. In some cases, teachers receive feedback from principals, assistant principals, and instructional coaches. Principals, by and large, are evaluated by their superintendents and/or their district leaders (e.g., director of finance, director of human resources).OSPI should provide guidelines about who should conduct evaluations and follow through with appropriate training.

6. Invest in evaluator training

Interview data, in particular, suggest that the state has largely been successful in building capacity at the district and school levels to use data to improve student achievement. In contrast, administrators show marked variation in their evaluation training: some administrators reporting their last training in "principal school" and others reporting regular training on evaluation content, compliance, interrater reliability, and instrument use. The telephone interviews suggest that school principals have an especially weak understanding of how much and what kinds of evaluation training those evaluating them have received.

Educators in telephone interviews and forums expressed a common concern about subjectivity. They said the observations, upon which evaluations are largely based, are often led by individuals poorly trained to conduct them. Ten of the 15 administrator focus groups included discussion of subjectivity in the evaluation process, particularly because the instruments being used were vague and not connected to an instructional framework. As one principal stated, "[B]asically [my evaluation is] what I think should take place in your classroom" (Appendix B, p. 38. Another principal noted, "The criteria that we have is very vague. So it's just a very, very subjective process" (Appendix B, p. 38). This concern is echoed even among educators who have received favorable evaluations. They, too, recognize the shortcomings of a subjective instrument in the hands of a poorly trained evaluator. OSPI may consider integrating into its educator evaluation system a training strategy for building school and district capacity to conduct rigorous, valid, and responsive educator evaluations.

7. Conclusion

The research simply has not evolved to design a truly exemplary educator evaluation system. Nevertheless, states and school districts can design more fitted evaluation instruments that articulate evidence-based practices, identify educator strengths and areas for improvement, and provide the kinds of feedback to support professional growth that leads to improved student achievement.

The considerations offered in this report are informed by Washington's own educators. If OSPI adopts these considerations, a blueprint for educator evaluation might look something like the following:

- Squarely focused on improving teaching and learning
- Guided by instructional frameworks that reflect the most up-to-date, evidence-based practices for teaching and leading
- Supported by rubrics that include clear descriptions of practice and performance; multiple rating levels; and multiple measures to stimulate conversation and reflection that support improved practice for student learning;
- Implementing at least annual observation review of all educators
- Providing intensive evaluator training

Such a model will need to leave some degrees of freedom so that local districts and educators can customize their evaluation systems to align with district goals for student achievement and individual goals for professional growth. The Washington stakeholders who informed this report, including teachers, principals, district administrators, parents and community members, welcome thoughtful change to the current educator evaluation system that respects the complexity of teaching and leading, and aims to improve professional practice.

Appendix A:
Survey Results

Exploring Differences in the Perceptions of Teacher/Principal Evaluation Policy and Practice

Washington teachers and principals received an online questionnaire similar to the one received by administrators. Using these common questionnaires, we explored the extent to which teachers' perceptions of district evaluation policies and practice reflected reports from district administrators. Table 1 reports the overall percentage of within-district agreement between teachers and their district administrators about their perceptions of teacher evaluation policies and practices.[11] Similarly, Table 2 reports the overall percentage of within-district agreement between principals and their district administrators about their perceptions of principal evaluation.

Table 1: Percentage of Within-District Agreement Between Teachers and Administrators

Teacher Evaluation Questions	Percentage Agreement Total	Percentage Agreement by Locale			
		City	Suburb	Town	Rural
Does your district have a clear definition of effective teaching?	52.4% (*N* = 77)	36.4% (*N* = 6)	57.2% (*N* = 18)	52.7% (*N* = 11)	52.6% (*N* = 42)
Has your district adopted a particular instructional framework or set of standards for assessing teachers?	34.3% (*N* = 76)	18.0% (*N* = 6)	35.2% (*N* = 18)	30.4% (*N* = 11)	37.3% (*N* = 41)
How many levels are included in your district's teacher evaluation rating system?	44.3% (*N* = 77)	21.8% (*N* = 6)	47.2% (*N* = 18)	42.9% (*N* = 11)	46.5% (*N* = 42)
When evaluating teachers, does your district use rubrics that describe what different levels of effective teaching looks like?	56.6% (*N* = 77)	50.0% (*N* = 6)	51.6% (*N* = 18)	49.7% (*N* = 11)	61.6% (*N* = 42)
What is the purpose(s) of your district's teacher evaluation system?	12.9% (*N* = 71)	16.3% (*N* = 6)	13.5% (*N* = 17)	11.0% (*N* = 8)	12.5% (*N* = 40)
What measures does your district use to evaluate teachers?	2.5% (*N* = 71)	2.5% (*N* = 7)	3.8% (*N* = 17)	3.0% (*N* = 9)	1.8% (*N* = 38)

[11] Percentages reflect the overall percentage of teacher–administrator agreement among districts in which at least one teacher and one district administrator responded to the questionnaire. Within-district teacher response rates were often low. Results should be interpreted with caution and do not necessarily reflect the true statewide within-district percentage of agreement.

Teacher Evaluation Questions	Percentage Agreement Total	Percentage Agreement by Locale			
		City	Suburb	Town	Rural
Is student growth data used to evaluate certain teachers in your district?	72.9% (*N* = 74)	71.9% (*N* = 6)	75.0% (*N* = 18)	67.9% (*N* = 11)	73.4% (*N* = 39)
Does your district provide training for teachers to use data to inform student improvement?	61.8% (*N* = 71)	57.0% (*N* = 6)	63.2% (*N* = 18)	65.0% (*N* = 10)	61.1% (*N* = 38)

N = Number of districts in which at least one teacher and one district administrator responded.

Table 1 shows that between 71 and 77 districts had at least one teacher and one district administrator respond to the questionnaire, which represents between 24 and 26 percent of all districts in Washington. The table suggests high levels of disagreement between administrator and teacher perceptions about their district's evaluation system. The highest rate of agreement pertained to the use of student growth data to evaluate teachers, on which almost 73 percent of administrators and teachers provided the same response. Agreement rates were above 50 percent on only half the questions completed by both groups. Of particular note, only about one in eight teachers agreed on the purposes of the district's teacher evaluation system (12.9%), and only 2.5 percent agreed on the measures used to evaluate their teaching.

Table 2: Percentage of Within-District Agreement Between Principals and Administrators

Principal Evaluation Questions	Percentage Agreement Total	Percentage Agreement by Locale			
		City	Suburb	Town	Rural
Does your district have a clear definition of effective principal leadership?	34.6% (*N* = 40)	52.2% (*N* = 5)	31.1% (*N* = 15)	15.0% (*N* = 6)	40.5% (*N* = 14)
Has your district adopted a particular instructional framework or set of standards for assessing principals?	42.0% (*N* = 41)	51.3% (*N* = 6)	45.0% (*N* = 15)	40.0% (*N* = 6)	35.7% (*N* = 14)
How many levels are included in your district's principal evaluation rating system?	40.7% (*N* = 39)	59.3% (*N* = 6)	46.6% (*N* = 14)	36.0% (*N* = 5)	28.6% (*N* = 14)
When evaluating principals, does your district use rubrics that describe what different levels of principal leadership looks like?	25.4% (*N* = 40)	21.7% (*N* = 5)	37.0% (*N* = 15)	41.7% (*N* = 6)	7.1% (*N* = 14)
What is the purpose(s) of your district's principal evaluation system?	16.0% (*N* = 39)	22.1% (*N* = 6)	14.7% (*N* = 15)	12.9% (*N* = 5)	15.9% (*N* = 13)

Principal Evaluation Questions	Percentage Agreement Total	Percentage Agreement by Locale			
		City	Suburb	Town	Rural
What measures does your district use to evaluate principals?	67.8% (*N* = 37)	50.5% (*N* = 5)	60.2% (*N* = 14)	60.2% (*N* = 14)	78.1% (*N* = 37)
Does the principal evaluation process include the use of student growth data?	57.8% (*N* = 40)	35.0% (*N* = 6)	49.0% (*N* = 15)	75.0% (*N* = 6)	70.5% (*N* = 14)
Does your district provide training for principals to evaluate teachers?	69.9% (*N* = 39)	92.4% (*N* = 6)	77.7% (*N* = 15)	62.5% (*N* = 6)	52.8% (*N* = 12)
Does your district provide training for principals to use data to inform student improvement?	55.4% (*N* = 39)	78.5% (*N* = 6)	67.1% (*N* = 15)	41.7% (*N* = 6)	36.1% (*N* = 12)

N = Number of districts in which at least one principal and one district administrator responded.

About 40 of 295 districts were included in the analysis of agreement between principals and their district administrators, which represent about 13.6 percent of district in Washington. As with teachers, principals' perceptions about their district's principal evaluation system was often incongruent with their district administrator's perceptions within each district. Close to 70 percent agreed on whether evaluation training was provided for principals (69.9%) and on the measures used to evaluate principals (67.8%). Overall, however less than half of principals agreed with their administrator(s) on five of nine common questions about principal evaluation. Although agreement rates appear low for principals and teachers who responded to the questionnaire, findings should be interpreted with caution and should not be generalized to represent the percentage of agreement statewide.

District Administrator Survey Results[12]

Teacher Evaluation Practices

Does your district have a clear definition of effective teaching?

		Locale			
	Total (*N* = 97)	City (*N* = 8)	Suburb (*N* = 20)	Town (*N* = 11)	Rural (*N* = 58)
Yes	47.9%	55.1%	60.0%	45.5%	43.1%
No	41.9%	44.9%	20.0%	54.5%	46.6%
Uncertain	10.3%	.0%	20.0%	.0%	10.3%

[12] Weights were applied to survey responses so that districts were given equal weight. For this reason, percentages within certain cells may appear lower than expected.

Has your district adopted a particular instructional framework or set of standards for assessing teachers?

		Locale			
	Total (*N* = 97)	City (*N* = 8)	Suburb (*N* = 20)	Town (*N* = 11)	Rural (*N* = 58)
Yes	23.8%	58.7%	35.0%	45.5%	11.2%
No	74.1%	41.3%	60.0%	54.5%	87.1%
Uncertain	2.1%	.0%	5.0%	.0%	1.7%

How many levels are included in your district's teacher evaluation rating system?

		Locale			
	Total (*N* = 97	City (*N* = 8)	Suburb (*N* = 20)	Town (*N* = 11)	Rural (*N* = 58)
Two levels	50.4%	36.7%	50.0%	45.5%	53.4%
Three levels	31.6%	32.6%	25.0%	27.3%	34.5%
Four levels	18.0%	30.6%	25.0%	27.3%	12.1%

When evaluating teachers, does your district use rubrics that describe what different levels of effective teaching looks like?

		Locale			
	Total (*N* = 97)	City (*N* = 8)	Suburb (*N* = 20)	Town (*N* = 11)	Rural (*N* = 58)
Yes	20.4%	42.9%	36.7%	27.3%	10.3%
No	77.5%	57.1%	58.3%	72.7%	87.9%
Uncertain	2.1%	.0%	5.0%	.0%	1.7%

What is the purpose(s) of your district's teacher evaluation system?

		Locale			
	Total (*N* = 102)	City (*N* = 10)	Suburb (*N* = 20)	Town (*N* = 13)	Rural (*N* = 59)
Compliance	78.1%	36.7%	85.0%	69.2%	84.7%
Professional learning	63.1%	46.7%	73.3%	53.8%	64.4%
Recommendations for state licensure advancement or renewal	10.3%	15.0%	10.0%	.0%	11.9%
Continuing contract decisions	56.9%	40.0%	70.0%	30.8%	61.0%

Compensation	2.5%	15.0%	.0%	.0%	1.7%
Dismissal/selective retention decisions	64.4%	46.7%	90.0%	46.2%	62.7%
Assignment decisions	20.4%	11.7%	28.3%	.0%	23.7%
Promotion/advancement decisions	7.4%	15.0%	5.0%	.0%	8.5%
Determination of and reporting on the distribution of effective teachers	9.3%	5.0%	20.0%	.0%	8.5%
Identification of teacher leaders	14.2%	5.0%	20.0%	.0%	16.9%
Advance the district's strategic goals	24.7%	25.0%	23.3%	7.7%	28.8%
Other	4.9%	.0%	10.0%	.0%	5.1%

Appendix B

Washington's Teacher/Principal Evaluation Pilot (TPEP) Educator Forums: Final Report

Introduction and Background

The following report is a preliminary analysis of qualitative data from the Teacher/Principal Evaluation Pilot (TPEP) Educator Forums held in the state of Washington between February and March 2011. The purpose of the forums was to collect participants' views on (1) the current teacher and principal evaluation system in the state of Washington, (2) the state's recent education reforms and pilots, and (3) what features and characteristics should be included in the new evaluation system. Teachers, principals, superintendents and other district officials, education directors, school board members, parents, and community members participated in 10 forums hosted by educational service districts (ESDs) across the state. The approximate number of participants in total was the following: 98 teachers, 85 principals, 71 district administrators, and 24 community members and parents.

Participants were recruited through an invitation process to ensure that teachers, principals, and community members from a diversity of backgrounds and regional locations were included but at the same time ensuring that any educators in the state who wished to partake were not excluded. The project team reached out to all ESDs to identify 20–40 invitees that represented the diversity of districts in their region in terms of performance levels, poverty levels, size and urbanicity, and the diversity of grade and subject levels taught in the schools.

Each forum included a 20-minute overview and introduction, a brief presentation on current developments in Washington's educator evaluation process, a large group question-and-answer session, and a facilitated, one-hour, small group discussion organized by the stakeholder group. Participants were also given an exit survey.

Methodology

The qualitative data analysis presented in this report is based on the facilitated, small group discussions, which were audio-recorded and transcribed. Teachers, principals, superintendents, and community members were organized into separate discussion groups; however, depending on the distribution of participants at each forum, occasionally there was some overlap between the groups.

Although facilitators prompted and asked participants about *both* teacher and principal evaluation, participants' knowledge and interest was greater on the topic of teacher evaluation, and, therefore, the preponderance of discussion in the focus groups centered on teacher evaluation. Although this report presents findings on both systems, the extensiveness and depth of data available on views of principal evaluation is somewhat diminished in comparison to the data on teacher evaluation.

Transcripts were coded and analyzed using NVivo 9, a software program designed for systematic analysis of qualitative data. Initial themes or codes were based on the focus group protocol and the 10 topics identified in the TPEP Technical Proposal. The initial themes were refined as new themes and ideas emerged upon further examination of the data. The preliminary report includes limited representative quotes; additional quotes and comment counts will be included in the final iteration of the report.

Teacher Evaluation Systems

Who currently evaluates teachers and how frequently are they evaluated?

Teachers largely report being evaluated either by a principal or an assistant/vice principal. One focus group reported that their principals shared teacher evaluation responsibilities with either the superintendent or instructional coaches.

Teachers report variation in the frequency of evaluation. Most teachers indicate that their school uses three different types of forms: a long form, a short form, and a professional growth plan (PGP), and that the frequency of their evaluations depends upon which of these they use. The long form generally entails more frequent observations and often includes both pre- and postobservation conferences and unannounced, drop-in observations. The short form generally entails 30 minutes to 1 hour of observation annually. Under the PGP, teachers usually engage in goal setting at the beginning of the year and have a follow-up meeting with their principals later in the year to discuss whether their goals have been met.

Teachers indicated that evaluation implementation can be spotty. Teachers regularly reported being observed in the fall and receiving no feedback until late spring or even summer. One teacher noted that the pre- and postobservation meetings are often skipped, due to the principal's tight schedule, *"*[T]he pre- and postobservations are done about half the time, I would say, in my school. The administrators are so darn busy dealing with so many things they frankly don't get to them. And they kind of say, 'Mind if we skip that? It's okay with me, it's okay with you'. *"*

In contrast, other teachers experience regular and frequent informal "walk-through" observations from instructional leaders and principals, which when done effectively, can provide valuable feedback; however, most teachers stated these observations were often unconnected to their formal evaluations.

Principals and superintendents concurred with teachers on the role of the principals or assistant principals as the primary evaluators for teachers. A few did report relying on instructional coaches or professional learning communities (PLCs) to conduct classroom observations targeted more directly at professional growth and learning; again, however, these observations were not connected to formal, summative evaluations. One principal stated that he was required to carry out three, 30-minute observations annually for each teacher. Principals in two focus groups noted they also found consistent implementation of teacher evaluation to be difficult because of poor instruments or time constraints.

Community members believed that most of what they knew about teacher evaluation was second-hand information from relatives or friends in the education profession. Members in three focus groups stated that they thought teachers were evaluated a maximum of once a year and for only 15 to 30 minutes by the school principal.

What do participants want out of a teacher evaluation system?

Teachers overwhelmingly want an evaluation system that will help them grow professionally. Although they frequently acknowledge the tension between an emphasis on professional growth and the practical need to determine when a teacher is unable to demonstrate improvement and should be terminated, teachers reiterated their strong desire for a system that would provide helpful, productive feedback. One teacher recalled a principal who set a collaborative tone for evaluations that she felt represented the purpose of a good evaluation system:

> "One of my first teaching jobs…he said to me, "You know what, take a look at the base word of evaluation, it's to value, and when I come into your classroom and I see [you] interacting with kids all day, I'm looking for what I am valuing about you as a staff member and what you contribute. Nobody is perfect, everyone has strengths and weaknesses." After that I learned so much during that time with that person because when I walked in I knew I was going to get not only positive feedback, but I was going to get feedback to help me be more valuable, instead of just saying what I'm doing wrong and what I'm not doing wrong in June and not being able to do anything about it."

Principals and superintendants echoed teachers' comments and resoundingly endorse creating an evaluation system that helps teachers constantly grow and improve.

> "I think a system should cause conversations around teaching and learning and even help your teachers that are above average become better….I think the biggest challenge is to move a good teacher to great. To be able to have something specific, to have them set a goal and you can collect data for it. That's what we enjoy doing. I mean, a system that is a formatted structure for rich conversations around the learning is really what we want. We need a system that works for reflective teachers, the teachers that say I want to get better. And we're their coach and we're supposed to help them get better. Whatever the system is, the purpose should be to increase the teaching and learning in the classroom."

Four administrator focus groups believed that terminating teachers should not be the ultimate purpose behind an evaluation system; however, two administrator focus groups noted that, although growth was the goal, accountability still needed to be part of an evaluation. If a system is oriented towards constant growth and achievement, mediocre teachers who fail to improve will stand out. As one principal said, *"*If it's…a growth model and it's perceived as a way of helping everybody be more effective, the bottom five or ten percent are going to stick out*."* One principal believed that the evaluation process should promote creativity and experimentation among teachers: *"*They have this safe person to go to, to try something new and bounce something off. It's not being evaluated. It's safe for them. So this won't be as threatening…The evaluation system should be to improve teaching…." Two of the six community focus groups addressed the broader question of the goal behind teacher evaluation, and both groups expressed support for the view that teacher improvement and growth were the ultimate goals of the system. As one member stated, *"*I think sometimes the evaluation system…they want to use it as [a] hammer to get rid of teachers…It's not what it's supposed to be….[T]he way I think it should be used is as a growth model…It should be a tool to use to help teachers get better*."*

What do participants believe are appropriate uses of teacher evaluation results?

Compensation

Teachers overwhelmingly reject the idea that evaluation results should be tied to compensation. In nine out of the 11 teacher focus groups, participants expressed disagreement with or raised concerns about merit-based pay. The most prominent concern centered on the potential threat it might pose to teacher collaboration. Participants in five teacher focus groups believed that it would establish a competitive environment that would undermine hard fought efforts to build collaboration among teachers and "open doors" in schools.

"It does seem like sometimes it almost flies in the face of the whole PLC [professional learning community] mentality. I love my team, but if I'm going to get more money than them I'm not telling them what I'm doing. [Laughs] We worked so hard to get that collaborative sense. And I sometimes have that same fear. I think healthy competition is a good thing. Competition with yourself is really good. But if it becomes this competitive nature, are we just going to close up and not share what we know? That makes me really nervous."

Participants in four teacher focus groups were particularly concerned about the potential inclusion of student achievement data in the evaluation should it be used for compensation decisions. Many teachers reiterate that students are not randomly assigned to classrooms, and it is not yet clear that the evaluation system will be sensitive enough to account for demographic or context-specific differences between classrooms and schools; therefore, using test scores as a basis for high-stakes decisions such as compensation would be unfair.

> "I've got the special and the Title kids, all of them. And I've got lots of apathetic learners this year—larger than usual. I'm like, gosh, I've got twenty F's out of my 120 kids. I talked to the other language arts teacher, and he's like, 'I've got nine.' What? [Laughs] This is totally why I'm against merit pay….He would get all the pay and I'd get none."

> "I teach special education and I teach the lowest kids, so I do the WAZ portfolio, the alternative assessment, and what we're asked to do with those portfolios is not reflective of what we're teaching every day in the classroom…how can you compare a kid with an IQ of 69 who's not going to make a year's growth in a year to the gifted and talented kid who's growing two or three years in a single school year? How do you balance that out so you don't say the special ed teacher is not doing a good job because Suzy didn't make a year's growth, and oh look, the TAG teacher is doing a fabulous job because Johnny made three year's growth? How do you balance that out to be fair if you're considering it as part of a teacher evaluation?"

Participants in three teacher focus groups believed that compensation could be considered if it were tied to a career-ladder or building-based approach to counteract the potentially negative effects of a competitive environment.

"We have people that work their butt off in their building and work extra hard and then you have slackers that still meet satisfactory. It would be neat to see merit pay based on your whole building's performance. Then you're forced to collaborate. I'm going to care what the English

teachers do, and I'm going to care what the math teacher is doing. Dangle a carrot and say, 'If you reach X percentage, this benchmark with your students, you all get this $3,000 or $4,000 bonus' instead of individualizing it from teacher to teacher. It would be a buildingwide approach."

Superintendents and principals raised similar concerns about undermining collaboration and the possibility of teachers "cherry-picking" the better students. Two of the administrator focus groups believed that a merit-based pay system would be in direct conflict with a professional learning community culture. One participant noted, "[I]f you tie money to it…what I really worry about is that it could lead to a culture where the door starts shutting again…In my district, we're just getting the doors open. We're developing professional learning communities and I don't want those doors [to close]. I don't want this to become a competitive thing where people are closing up…." Echoing the teachers concerns about the inclusion of test scores, one administrator focus group discussed the problem of how students are distributed, with one participant asking fellow participants, "On the merit pay situation, how often is it that the teacher that's been around the longest, that has seniority, has the cream of the crop of the kids at their school?"

One principal stated that his view on compensation depended heavily on the fairness, validity, and acceptance of the evaluation system. Unlike teachers, however, principals and superintendents in two focus groups were more willing to consider using evaluation results to reward exceptional teachers. Although administrators who mentioned this solution preferred a career-ladder approach rather than a straightforward merit-based pay system: "If we're going to keep those teacher leaders teaching, then I think you do have to explore some compensatory measure as well…I think there's a need for some kind of career ladder for people, but when it becomes competitive, I think that's where there's a real danger there, that you lose the benefit of it."

Community members were more divided on the question of including evaluation results in determining teachers' compensation. Participants in two community member focus groups believed very strongly that rewarding teachers equally for different levels of effort and results was unfair and that using evaluation results to determine teacher compensation would be an important shift away from the current emphasis on seniority.

> "I think it should be related to your pay and raises. Your performance should have to do with that. It seems to me it could improve the morale in a building because our current model is years of education plus years of experience. We've all seen where second grade, there are two different teachers, and…. [a]ll things being equal, they have the same of those two parameters, and one is engaged and busting their hump to do it and the other is phoning it in. Yet, they get paid the exact same amount. How is that the correct thing in any universe?"

Participants in one community focus group expressed qualified support for including evaluation results in compensation decisions, with one member noting: "I think some sort of pay. I don't think like direct merit pay is helpful, but I think some sort of career progress should be based on that."

In another community focus group, one participant believed that there seemed to be little evidence that merit-based pay had succeeded in improving student learning, although another expressed agreement with educators that current student testing instruments were not of a high enough quality and should not be part of determining an individual teacher's pay.

Termination

Teachers are predominantly opposed to the idea that evaluation results should be used in making teacher termination decisions. Participants in seven out of the 11 teacher focus groups believed that termination decisions were not an appropriate use of evaluation results. These teachers believed that using the evaluation for termination decisions would undermine the formative purpose behind the evaluation and defeat the ultimate goal of creating a culture of improvement and growth. One teacher noted, "You have a tendency to put people on the defensive posture if they think their job is at stake. And I don't think they're as open to some real creativity and testing out different theories and data." Another teacher asked, "The reason for an evaluation is to grow professionally. Do we want an evaluation to be the only tool that a principal can use to terminate our employment, based on one evaluation?"

However, the focus groups were not unanimous in opposing linking evaluations and termination decisions. Participants in two focus groups believed that if a teacher were receiving poor scores and not demonstrating improvement overtime, then termination might be appropriate. One teacher stated quite forcefully that struggling teachers should be supported; however, everyone needs to be meeting a minimum standard to continue teaching:

> "I see only two things that evaluation should be used for: professional growth and I will say it, teacher elimination. I'm sorry, if we have crappy teachers out there that aren't doing their job, and they're at 1's all across the board, I don't want to follow up. I don't think it's fair to the children that we're serving….Let's support our 2's and give them the extra [help]… all the support we can….But if you're on 1's all across the board… I don't care if you're right out of college, you should be at least at 2's. If you've been teaching for 5, 10, 15, 20 years, you should be at least at 2's. If you're not, maybe you're in the wrong profession. That's my opinion."

District administrators are more divided on the issue of whether evaluation results should be used in terminating teachers. In five of the administrator focus groups, participants opposed their inclusion in termination decisions; however, participants in four administrator focus groups also expressed the opposite opinion. Those participants opposed to using evaluations in termination decisions believe that inclusion would continue to create an adversarial relationship between teachers and administrators and that it would provide no motivation to teachers for continued improvement.

> "It amazes me that when you put somebody that's unsatisfactory [on probation, even though half the staff tells you that they know they're unsatisfactory, how that rift becomes between teachers and administration, and then that's what makes the evaluation criteria not a collaborative approach, it becomes [an] "us versus them" type of approach. And so when you go down that road, you have to be real sure you're going to do a good job, because it's an ugly scenario for two years when you get rid of them. [Y]ou have to be pretty ruthless and it's isolating."

> "When I got my evaluation at the end of the year as a teacher, I'd read through it…and just file it. But when the [student achievement] results came out, I was first in line to see how my kids did. [T]he real results of your children are extremely powerful. You want to see how they did. You want to get better as a teacher, but you want to get better because you care about the kids and their progress, rather than being tied to, 'you're going to get canned if your kids don't perform better.' "

Administrators in two focus groups stated that they would only endorse the inclusion of evaluation data if it shortened the time and lowered the expense required to dismiss a teacher. As one participant put it, "If this gets us back to where we still gotta spend two years or $200,000 to work a teacher out of a system, it hasn't done us a bit of good."

In contrast, administrators who believed that evaluation results should be included in termination decisions argued that termination decisions are part of what administrators are tasked to carry out, and evaluations are necessarily a part of that responsibility.

> "I still think you have got to do your job as the building principal and make sure that you're giving support to those teachers that are struggling and then the reality is, every once in a while, there's a teacher that maybe should not be a teacher. I think it behooves us to do our job."

> "They've got to step up to the plate. They were saying, 'Give them a 2 because they're a first year teacher.' I'm not going to give anybody a 2….They get a 2 because they deserve a 2. They don't get a 2 and then I'm going to keep them if they're a first year teacher. They better show me some 3's if they're a first year teacher. If they're all 2's, they're gone. In my school I will recommend that to my superintendent."

Community members are heavily in favor of using evaluation results for termination decisions, particularly if it enables districts undergoing budget cuts to keep teachers who are very talented but who lack seniority. All six community focus groups included participants who expressly supported the view that evaluation results and termination decisions should be linked.

> "The state has had many opportunities to take action, in terms of giving us and empowering us with local control to use the evaluative process as part of the RIF criteria. But right now, in this state, we're required by law to simply use seniority. If the legislature, who is so keen on this legislation, wants to improve student learning, then they should put on their big boy pants. They should step up and change the law so that we can use evaluation, in terms of seniority for [reduction in force] RIF."

"[T]wo years ago, for the first time in the history of [our] school district, we had to RIF teachers. That was hard to take…we had teachers out picketing in front of our school board meeting and…[one of the teachers said] "Look what you're doing, you're getting rid of our youngest, brightest, most creative, energetic young teachers," and I just had to go back at her and say, "I'm not doing it, you are with your contract." I am, as a school board member, looking for the tools for us to be able to keep those young, bright, [and] energetic teachers….That's all I'm interested in here. So, you know, I'm very excited about this program, I hope it does enable us and give people the tools to do that. So, that's what I'm expecting out of this."

However, one participant in a community focus group did express reservations about the validity of the evaluations, noting that such a high-stakes outcome would require that all principals be properly trained to evaluate teachers in a fair manner: "In my district, what I experience is that some politics get into that. You might have managers that are micromanagers. So I would maybe not want to use it that way. We might have some building principals who never had the professional development on how to evaluate….They don't have the experience to do a good evaluation of their staff because they've never been allowed to make decisions."

Professional Development

Teachers expressed substantial support for using evaluation results to create targeted, valuable professional development (PD), either at the individual or district levels. In the six teacher focus groups that discussed PD, the participants stated that they would like to see PD planning and evaluation results linked; however, participants in two teacher focus group emphasized that funding needs to be provided to do so, especially in smaller districts.

> "If you're getting targeted, valuable feedback that could help channel the professional development you choose. Or even the professional development that's offered if they're seeing a lot of trends districtwide. You get to a point where, you've got to have your 150 clock hours or 15 credits every five year[s]. Well, it would be nice to not just take the random tour through Washington just to get it. It'd be nice to have something that says, 'This would be a good professional growth plan for you.' Then you could be making valuable choices so you're spending your money on something that truly furthers your practice."
>
> "I think what's frustrating about the evaluation system, as a teacher for years, I've struggled with that oftentimes there's no real specific feedback. But then when there is specific feedback, if you get it from a variety of different ways, there's nothing that takes us further. I would love to see something tied [in] like, you were really struggling in this area and here is the professional development that we will pay for, that we will give you, and find a way to really remediate what you need. I just see that's a huge gap that we have."

One teacher suggested that teachers be provided regularly scheduled collegial time to engage in PD as colleagues.

> " [U]nless you figure out a way to adjust time in the school system, not much, frankly, can change. It just can't. The irony of the research I've done on other schools in other countries, the key factor, as it seems to me, is the collegial time away from students. Our system is built to put us in front of students as many hours and minutes as possible and get better on our own time…[a]nd that's not really very effective. For the kind of practice we're talking about, we need to figure out a way…to give us that kind of time to build good, solid lessons that mean when we are in front of kids they're getting the very best we can give them. Somebody who used to work here used to say what gets in the way is the "dailyness" of teaching. 'I can't do that. I'm here today. If I wasn't here today I could probably adjust that but right now I've got the next class period coming in. So I hope I can get to it tomorrow.' That's the reality of our job. It's unfortunate. And I think if we

> can change that or find a way to manipulate that, I think we have a way to really make some serious progress."

In the four administrator focus groups that discussed PD, principals and superintendents agreed that PD should be designed in response to teacher evaluation results. One administrator noted that linking the two would "allow us to align professional development to better meet the needs of the district and really define what we're going to be offering teachers throughout the year [for what] they can sign up for through the district." Several principals echoed teachers' ideas about targeting PD to specific, focused areas and addressing the problem of making the process manageable for teachers. One superintendent noted, "It would be nice to have more professional growth opportunities, but our professional growth plan is so cumbersome for the teachers that they don't even want to try it. And so you know, it's just basically the forms and checking the box instead of really working together to be the best we can be."

Community members did not discuss linking PD and evaluation results at any length in their focus groups; however, one participant did state that she felt evaluation results should be used, "[f]or professional development. It should be used as a tool so you can get your teachers help. Some of them may not improve, or may not want to improve as they become more senior. They want to do it the way they want to do it. Those days are out the door."

Promotion

Tying evaluation results to promotion decisions was not discussed extensively in the focus groups. Participants in one teacher focus group discussed promotion; however, one teacher believed it would be appropriate and might make the process of promotion less subjective: "I like that 'promotion to leadership positions' because sometimes it seems people are selected just because they get along with someone really well." Another teacher noted, "I think there needs to be an application process. I think that could be part of the application process for a leadership position…So I can see that as being part of their personnel file screening if you apply for that. But I can't see it being the sole purpose for why you'd get a promotion." One additional teacher believed that evaluation results should only be included as a tiebreaker between two, equally qualified candidates.

One principal noted that the evaluation results could help principals recognize untapped talent and potential teacher-leaders.

> "A rubric may lend itself to the identification of [teacher] leaders in another way, in that [you can say], "You have not stepped forward to lead, an expectation of leading is that you are exceeding expectations, and you clearly are. My expectation now is that you are now sharing this with other people." In our current model, there would be no conversation that would look like that, because we're all meeting level expectations. But if you have this differential between level three and level four, couldn't you see yourself saying to a teacher, "If you want to stay here, I need your help with…" and maybe that could lead to a mentoring conversation, maybe that could lead to a departmental kind of leadership task of some kind….I think there's a lot of untapped resources and I think we go to [current teacher] leaders…but if we had some data…that we would have on someone that we could say, 'Your data suggests that you'd be good at something.'"

Public Availability of Evaluation Results

Teachers in three focus groups expressed concern over the public availability of evaluation results. In the words of one teacher, "What's the purpose of making them public anyway and why do they have that right to do that? I would never make my student's test score public to anybody. My husband works for the city, his evaluation is not public." Another teacher raised the point of how students might feel knowing that their teacher was not among the district's strongest.

District administrators expressed conflicting views about publicizing evaluation results, mentioning the balance that must be struck between accountability and creating a positive school atmosphere: "The challenge I think is striking the balance because I think that one of the drivers is that there's a public perception that there's a lack of accountability for teachers."

Several district leaders agreed that there is a public perception that a lack of accountability exists, and that it must be demonstrated to the public that it is no longer the case that 99 percent of teachers are rated proficient or distinguished. On the other hand, they noted that ramping up accountability in this way can harm relationships, create negative parental reactions, and opens up the possibility of misuse of the information: "I think that's very, very fair, but the problem is when you start going down that road of trying to make sure that you're dealing with those ineffective teachers, how is it that you're striking the balance so that you're helping people to grow, or the same thing with principals because you're absolutely right. As soon as you start ramping up that accountability, it does things to relationships."

Parents and community members largely agreed that evaluation results should not be made public, with a primary concern being that parents will all request not to have their child taught by the lower-performing teachers. In five of the six community focus groups, participants voiced strong support for maintaining teachers' rights to confidentiality.

> "I don't think it serves any purpose to have teacher evaluations public. But I fully expect the school board to be accountable to the public to ensure that those evaluations are being taken seriously by staff.…And if that isn't happening, it should be brought to the school board members' attention and taken up with the superintendent. What sort of good would it do a teacher to have every child in that classroom know that they got a D- last year for their performance rating? That's not going to help them. It should be well known to the principal. It should be well known to the superintendent… And if parents want to hold a school board accountable for ensuring that this process is moved forward, that's where I think that should be."

There was some disagreement, however, by those who preferred to at least share evaluation results among teachers to promote healthy competition. One community member commented: "I think there should be almost a rating system, A, B, C, D, and I think it should be pretty public to at least the other teachers. I wouldn't put community, but when I'm competing it's other people in my own company, I'm going to do the best so that I'm getting the good grade." Another participant believed that evaluations should be used to at least publicly acknowledge the good work of those teachers with the highest ratings: "[Y]ou could do accolades. Accolades so you would know, these teachers are doing great and just leave it at that. But they're like the top tier,

kind of like honor roll. It doesn't mean the other kids are failing because they're not on the honor roll. It just means, 'great job….So it's a pat on the back to the teachers who are doing really well."

One community member did believe that evaluation results should be made public if it were necessary to address a consistently low-scoring teacher*:* "I don't know how much information should be shared with the public….But I also think that if there is somebody that's consistently getting 1 out of 4 and it takes the community knowing that in order to not allow that to happen anymore, then that's what it takes. And if there's teachers consistently getting 4's, then I don't see why the community would not want to know that."

What do participants cite as best practices and flaws in the existing teacher evaluation system?

Best Practices

Teachers repeatedly talked about the importance of good, productive conversations as the most helpful aspects of evaluation processes. In seven of the 11 teacher focus groups, many teachers spoke about the benefit they had received from rich conversations—either with principals or with fellow teachers through peer review. One teacher said, "From my classroom experience…the value in it for me was not the evaluation form, but the time to sit down with my principal and the conversations we had then were very productive. They weren't always just on the evaluation. You had time to focus on a number of things. So that relationship was really important. That's where the value was."

A few teachers specifically mentioned unannounced, drop-in observations and peer-based feedback as a "best practice" that they would like to see retained or expanded in the new evaluation system, in part because these strategies create opportunities for conversations and collaboration to occur. One teacher mentioned a practice in her school of using "learning walks." Several times throughout the year, four or five teachers at a time visit a fellow teacher's classroom to learn by observing his or her approach to teaching—a practice that helped build trust among teachers and increase teacher learning.

Teacher focus groups also emphasized systematic opportunities for self-reflection, particularly when connected to goal setting, should be continued. In three teacher focus groups, multiple participants noted that they liked the self-reflection and goal setting that are specific aspects of the professional growth plan option for evaluation.

Videotaping came up briefly in both teacher and principal focus groups as effective in promoting critical self-reflection. One teacher recounted, "I was marked down in one area and I thought, 'that can't be.' So I asked the person to videotape me, like you said. And I looked at it. I also had the principal demonstrate it to me and I actually did some notes myself. That was a powerful thing. There were several things we could put together." In one focus group, an administrator recalled, "When I taught that was one of the most powerful things I did, videotape myself. You learn a lot about your classroom. You see the things that the principal will see that you may not

see up front teaching. I think that's very powerful. We've got all kinds of technology tools. I don't think we use them enough to help our teachers get better and help ourselves get better."

In line with teachers' views, principals and administrators cited rich, honest conversations as the most important 'best practice' that they wanted to see in an evaluation system. Participants in 10 of the 15 administrator focus groups discussed this topic. Some administrators use an informal approach to generating conversations with staff: "One of the things I most love is to close the door and clean out the clutter and just talk to some of people you work with all year long that you see on the fly. You just get to sit with them for as long as you can afford the time to talk about what they're doing and listen to them. Just talk about things. I love doing that. It [evaluations] almost forces you to that opportunity. I like that." Other administrators report that using pre- and postobservation meetings are effective in cultivating constructive conversations:

> "I like that in the preobservation I will ask them what they want me to watch for. I think that's a really powerful way when you come back to be addressing that. I'm in there, I'll do research. If they want me to watch student engagement, I'll pull out some good suggestions for them and then follow up to see if they're doing it. And they'll often come tell me [and say], " Hey, I tried that." So I think those conversations are much more powerful than the written documents. I feel like they're often just a rubber stamp unless there's a very serious situation. But pretty much, I feel like the conversations are the power."

One superintendent whose school was trying out a new evaluation system highlighted the change in conversations and collaboration that has occurred.

> "[T]he richness of the conversations is incredible. In the time that I was principal...the conferences I would have with my teachers went something like…[a] litany of what I observed in the classroom. The conversations now, it's almost taken a reverse…Yes, we have a legal requirement of going in to observe, but the amazing thing is, our teachers are coming to our principals saying: "Could you come in today? I want to show you this that I'm doing and tell me what you think. Could I do this different or this different?" They're [also] doing that with colleagues….So, I think our current model really lends itself to…enhancing the instruction, but it's from a grassroots instruction, it's not from an accountability instruction."

Principals and district administrators noted that self-assessment and PD for teachers work best when done collaboratively in teams, when focused on student growth, and when providing opportunities for teacher self-reflection and goal setting.

Flaw 1: Lack of Useful Feedback

Teachers offered a wide variety of information on what they viewed as significant flaws in the current teacher evaluation system. According to teachers, the most common flaw in the current system is its inability to generate good, fruitful conversations that provide helpful, productive feedback. In total, all 11 teacher focus groups discussed the quality and timing of the feedback they received as a significant flaw in their current evaluation systems. In one teacher focus group, when the interviewer asked if the feedback they received was helpful, the participants unanimously responded with spontaneous laughter. One teacher described the evaluation process

in the following way: "[O]ur administrators do evaluations at the very last minute. They're often times rushed. They're often times not the full time that's allotted. The conversations often times are pretty meaningless and very quick and you get the sense that they just want to get through teachers as quickly as possible."

Teachers noted that the most helpful feedback they received came through informal conversations with principals and rarely through the formal, written evaluation process. Most participants believed that the observation protocol promotes a compliance model and offered no incentive or direction for professional growth. A common theme in all teacher focus groups was frustration at simply being told they were "satisfactory" and "doing fine." Many teachers expressed a desire to have critical, constructive feedback. As one teacher recounted, "I mean, they stop by all the time, but in fact, when we did our CPDP we have to meet in their offices, I was asking specific questions on needing some help with my goal and all I heard about was 'You're a great teacher, it's really great.' I'm still kind of stuck with this problem." Teachers took particular issue with the use of a binary "satisfactory or unsatisfactory" model, which provided no real, actionable feedback and did not recognize exemplary teaching.

> "If anyone tells you they're getting better at teaching from hearing satisfactory or unsatisfactory, they're lying. Really, how do you get better from that? I really feel like currently with the walkthroughs we have going on, the principal is in the building more often. Informally, I feel it's more of those conversations that lead toward growth, more than 'satisfactory.' At any time, the formal process is just the way it's set up in our contract. It's not a growth tool."
>
> "When I moved up to the tenure track it was once a year and same kind of checked yes 'you did it, no you didn't', and of course I've gotten all satisfactory, which is very unsatisfying because I happen to think I'm exceptional, and I would like to get credit for what I do exceptionally well. Just being told I'm satisfactory is not satisfying at all and it doesn't tell me where I need to grow and that's another concern that I have about this process."

In six of the teacher focus groups, participants were frustrated with the timing of feedback. Teachers reported being observed in the fall and provided feedback in April or May, often when it was much too late to make any adjustments or changes based on whatever feedback was offered.

Principals and superintendents shared teachers' views on the unhelpful and untimely feedback that the current system provides. Eleven of the 15 administrator focus groups mentioned poor feedback that provides little direction as a major flaw. Participants in eight administrator focus groups stated that they found informal conversations with teachers to be much more constructive and helpful; whereas, they described the formal evaluation as "a checklist," "rubber stamp," "constrictive," and meaningless. As one former teacher (now a principal), said, "Twenty-seven years of teaching with 27 "satisfactories" is nothing, really. [There's] no feedback on how to get better at what you do." A human resources (HR) person recounted a particularly striking example: "…we consciously got rid of, in bargaining language, the short form, because we had teachers that actually said, which scared me as a human resource person, 'But I haven't been evaluated in twenty years.' Well, they had, but it was on the short form, so therefore it had no meaning."

Several principals noted that the finality of the document is particularly constraining to how both principals and teachers respond to feedback: "It was a formal document and final. Knowing that's not going to change, why listen to the post-observation, what can I do to improve? You've already made your decision. As a principal on the other side, it's kind of the same way. With the contract negotiations, that is the final document."

Community members had less to say, but in three community focus groups, participants shared the view that the current system did not provide good feedback and that the evaluation "[D]oesn't sound like it's a tool for improvement. It's a tool to get the checkmark done."

Flaw 2: Ineffective at Improving or Removing Teachers

Related to their concerns over poor feedback, teachers were also quick to note that the current evaluation system does not work well in removing or improving ineffective teachers. Although teachers in two focus groups believed that ineffective teachers were being summarily terminated without being first offered support or feedback, in four focus groups, teachers indicated that they believed many ineffective teachers were simply given "satisfactory" marks and passed on through the system when they should have been terminated. As one teacher noted, "Frequently the evaluations just don't give the place to be able to say to the teacher what needs to be said about what's happening in their classroom, either positively or negatively. It's sort of a very vanilla, hands-off evaluation, and I don't think it's doing our teachers, at any level, any credit."

In one focus group, the teachers discussed the lack of trust the current system creates between administrators and teachers—largely because everyone is rated as satisfactory, and more attention from a principal (in terms of feedback) is often viewed as threatening, as illustrated in the following exchange:

> "Not too many people get lousy evaluations."
>
> "Some teachers feel it's [goal setting and gathering evidence] not valued in the outcome of the evaluation, it's simply a check mark. So nobody comes back at the end of the year and says, Cindy, your goal was this. What did you do to meet that goal? How are you doing towards it? What are your plans next? So it'd be nice to link that."
>
> I think the reality of our current system would be that the principal couldn't do that. That would mean they were targeting you. The system is built to protect teachers, currently. That's kind of what it was built for. So that's where you get a lot of grievances. 'Why are you in my room again?' That culture was built long ago. That's what the major change seems to be."

Principals and superintendents were significantly more vocal about termination, in part because the principals view the termination of a teacher in the present system as an adversarial, time-consuming, and expensive process (13 administrator focus groups agreed). One principal noted, "We move from viewing it as, 'okay, well, we have a problem we want to move to correct or facilitate some change here' and then instantly it's adversarial after that. They don't like you anymore."

Principals expressed concern over the legalistic nature of teacher evaluations because of their connection to termination processes and union requirements (three focus groups).

> "It really becomes a very legalistic process now. It's not the process about whether the teacher is effective or ineffective. It becomes did you meet your timelines, did you cross the T, did you dot the I, did you do it on the right day and if you didn't we're done, it's over. So, it really becomes, it goes beyond just trying to help a teacher or give them instruction to make them better. It's very formal and it really distances you from that person and the union becomes involved, our attorneys become involved, and again, that's when you come back to language, every word you say to them, everything that's written down, better be precise. In fact, our attorneys, we've had the come in a few times, they tell us not to use any verbiage that isn't on your evaluation form. You just transfer whatever it says there, as to write down here, write a recommendation on that and another piece that may say what you'd like to see happening, but use that same verbiage, otherwise you're going to get yourself in trouble."

Administrators brought up their own unwillingness to rate teachers as unsatisfactory or to put struggling teachers on growth plans because teachers would immediately bring in a union representative (four focus groups). Instead of using the evaluation system, a number of principals report having to use an entirely different, time-consuming documentation process in order to dismiss a teacher (three focus groups). Several principals noted that they felt frustrated that it can often take 1½ to two years to remove a teacher because they are simultaneously trying to offer the struggling teacher support through probation and improvement plans during which time they are also carefully documenting everything if it becomes necessary to terminate the contract (eight focus groups).

Other principals pointed out that they disliked connecting termination decisions to the evaluation process, in large part because the line between "satisfactory / unsatisfactory" or "meets / does not meet expectations" is thin, vague, and difficult to defend (two focus groups). One principal notes that it was a struggle to get teachers to use videotaping to engage in self-reflection because the teachers expected it to be used against them later and felt intimidated by the process.

Flaw 3: Rater and Instrument Subjectivity

In six teacher focus groups, participants were largely in agreement that the effectiveness of the existing system depends heavily on *who* is conducting the evaluation. Teachers report a deep sense of subjectivity in the current system. As one teacher noted, *"You have multiple people that are evaluating you. Like you said, you might get in the sights of one person and all of a sudden your evaluations are looking one direction and it's going, hey, what's happening?...It can happen. You can have someone whose rating is way over here and then…."*

In three focus groups, teachers stated that they believed their principals did not pay attention during the observation, did not follow the protocol, or simply did not care and just wanted to "check it off the list." One teacher said, "To be honest, some of them are written up in July. Some of them aren't even done during the school year. I see it as a tool you use to go after some teachers that they deem not effective versus those that are effective they kind of just let go and move on quickly and get them over it." Another teacher commented, "[W]e've had principals in

our building who literally copy and paste the same thing over and over for every teacher and it's absolutely not helpful at all."

The observation protocol itself came up for criticism in two teacher focus groups, with several teachers noting that their protocol seemed subjective and that the lack of consistency and common language across schools was problematic: **"**We use the same form but I think that has something to do with [the] perspective of the principal that's looking at the form and going, 'okay, I view this as this. And I view this as this.' So they do it differently. I think you've got some generalizations inside of the evaluation so a principal can take it one way or the other. I think that's something we need to work on, whether we want to make it a little more by a common language, so it's understandable all the way across the board." One teacher recounted her frustration that her rankings would suddenly change from year to year with no explanation.

Principal and superintendent conversations largely agree with teachers' perceptions. Ten of the 15 administrator focus groups included discussion of subjectivity in the evaluation process, particularly because the instruments being used were vague and not connected to an instructional framework. As one principal stated, "[B]asically it's what I think should take place in your classroom is how it's basically just determined. [I]f you had a rubric like what's shown and so forth and gives a more open idea of what should take place, instead of just John Doe thinks, period. Another principal noted, "The criteria that we have is very vague. So it's just a very, very subjective process."

Two principals reported having no criteria or rubrics, but simply being given blank forms with slots for the teachers' names and subject area to be filled in. Several administrators noted that the tool's usefulness depended heavily on the person using it. As one superintendent put it,

> "I think the instrument itself is as good as the administrator wants it to be. So we've got one administrator at X high school who told his teachers that he would be in their rooms at least six times each semester and is doing that and is very involved on a weekly basis with instruction and evaluation with each teacher. We have other teachers in the district or other principals in the district that I suspect don't even get their two mandatory observations done and there's everywhere in between. So I don't see that it's holding anybody back from being highly effective. But I don't see that it's requiring anybody to be highly effective at it either."

A common refrain from principals was that the current system lacked any common language or vocabulary to ensure consistency on what teachers were being evaluated on.

Flaw 4: Unrepresentative Observations

In addition to subjectivity, teachers raised concerns about the quality and representativeness of preplanned observations. Seven teacher focus groups raised the problem of "putting on a show," or becoming a "super teacher one day a year." Most felt preplanned observations were a completely unrepresentative measure of teachers' true instructional practices, particularly because students act differently when the principal is in the room. The following exchange represents this view well and illustrates teachers' open discussion of the problem of unrepresentative observations:

> "Can the teacher manage the kids if the kids look like they weren't paying attention? That's all it's about. You bribe the kids. "Be good, the principal is coming in." [Laughs] It's dumb."
>
> "Sometimes you don't even have to bribe them. The principal walks in and they go…." [Laughs]
>
> "In our district, our admin team has started to do walkthroughs. Now you have six of them walk in."
>
> "I was videotaping my entry in December…Chad said to me, "Don't worry Miss Miller, I won't fall asleep like I did in Miss [Inaudible]'s observation." [Laughs] I was like, oh my God, Chad. You don't know what I do for the rest of my world. [Laughs] It was really funny."

Principals and superintendents shared teacher's views on the "dog and pony show" aspects of pre-planned observations. In five of the administrator focus groups, participants stated that they were frustrated that they were not allowed to include in evaluations the information from their informal walkthroughs or information about how well teachers worked with colleagues or acted as leaders, but instead are limited to including only what they see in the 30-minute observation. Similarly, a repeated comment from administrators (six focus groups) was the overemphasis on teacher behavior and a lack of focus on student achievement in teacher evaluations. Several principals explicitly wanted student achievement data to be part of reflective conversations during the observation meetings.

Community members shared similar views about the inadequacies of planned observations and noted that a single observation once a year was simply not enough to help teachers grow. One community member recounted her own memory of being on "best behavior" on teacher observation day. Another community member, although sympathetic to the need for better and more frequent observations, noted that principals are already overtaxed and do not have time under current staffing reductions to do much more.

What do participants believe an "ideal" teacher evaluation system should look like?

Involve Collaborative Process

Teachers overwhelmingly want greater ownership over their evaluation process by engaging in goal setting, regular dialogue, and greater team-oriented, peer review (six focus groups). Several teachers reiterated that it was important for them to have input into their goals for growth rather than being told through a one-way dialogue by somebody else where they need to focus their improvement. Regular dialogue between the teacher and evaluator was desired, including using the preobservation meeting to decide together what the evaluation will focus on. Several teachers also valued working with teams. This strategy was suggested both in terms of teams of evaluators, to prevent a biased evaluator having too strong a say in a teacher's score, and in terms of teacher teams serving as resources for one another to achieve mutual goals because teachers value each other's opinions and recognize the constraints on principals' time. As one teacher put it, "The thing I really like about what you were talking about with working as a team of teachers and giving each other feedback and working together on specific goals. I think that sometimes we overlook the fact that we are a resource for each other and these poor principals, God love them, they only have so many hours in a day."

The principals and district administrators in the focus groups stressed over and over again the importance of collaboration (11 focus groups). One principal noted the importance of tying in peer coaching and professional learning communities, although another discussed the importance of ongoing professional conversations among both colleagues and evaluators, including those difficult conversations that some principals are afraid to have:

> "I think one of the problems is that you can make any tool effective or ineffective. But if the leader, whether it's the superintendent or the principal, is afraid to have the conversations that need to take place, it really doesn't matter what tool you have. We've done this to ourselves as principals because we're afraid to say what needs to be said. And we've sold kids out because we're afraid of that. Just creating a new evaluation tool will not change that. You'll still just put everybody at proficient if you're afraid to have the conversations."

At the school and district administrator level, this collaboration was sometimes seen as a need for everyone to be on the same page, looking for the same things, using common assessments to drive reflection and dialogue, and speaking a common language. Several administrators drew attention to the need for districtwide collaboration, with district administrators modeling with their principals the types of conversations principals should be having with teachers.

Emphasize Professional Development

With regard to using evaluations to inform PD, a range of ideas were presented by teacher focus group participants. Teachers in six focus groups noted that the PD that comes out of evaluations should be specific and focused. One said, "When there's something that's found that I need to improve in my classroom, I would like a specific option for professional development. I want there to be support for me in order to get better so it doesn't just sit at a Level 2 or a Level 1 or

whatever. There's a way for me to get better. That's inherent in the evaluation. This is where we all agree that I need to improve and these are the places where you can get the help for that." Also reiterated more than once by teachers was the notion that principals need to be given the time to provide effective PD, as well as capacity to provide effective PD because not all principals were necessarily the most effective instructors.

Administrators, too, believed that PD was an important component of an ideal evaluation system (five focus groups), with one claiming that it was a necessity to provide teachers with the skills they need if they are to be held accountable for high performance. Among administrators, there was some concern expressed about the difficulty in combining open, honest PD with high-stakes evaluations where teachers' energy is focused on looking their best. Several administrators noted that having peer academic coaches in every building to focus on the PD side would help. Other administrators believed that it is their job to show teachers exactly how they should be doing things differently to be more effective. Administrators brought up the usefulness of professional growth plans, professional learning communities, and using data as things that should be tied to evaluations. Several administrators believed that creating career ladders for teacher leaders would be an effective way to incorporate professional growth into the evaluation system.

The community members who participated in focus groups agreed that continuing education should be part of the evaluation system (two focus groups), but they had less to contribute about the nature of this relationship. One community member noted that it could be difficult to promote PD through evaluations, since the evaluation process is not very relaxed. Another parent/community member participant noted the importance of trust between teachers and their evaluators in order to have reflective conversations.

Provide Sufficient Frequency of Evaluation

Among the teacher participants, there was agreement that frequent, brief (10-15 minute) observations should take place throughout the year, taking the form of principals "wandering in" (five focus groups). Some participants believed that the more the principal can visit teachers' classrooms the better, although one participant stated that struggling teachers ideally should have more observations and evaluations than teachers who have shown themselves to be effective. One teacher pointed out that, by definition, an evaluation occurs once per year but that multiple observations can and should take place throughout the year; another teacher posited that, in the same way that students receive a report card three times per year, so too should teachers.

Principals and district administrators in four focus groups believed that more frequent observations would be an important part of an ideal evaluation system. Several principals believed that ideally they would spend more time in the classroom than they do at present. In theory, they appeared to agree that observations and conversations should be ongoing throughout the year to get a fuller picture of a teacher's performance (even if formal evaluations are only once per year), although in practice they recognized that time was a barrier. A couple of administrators supported the idea of allowing veteran teachers on a professional growth option to be evaluated every two or three years so that principals could concentrate their time on conducting more frequent evaluations for struggling staff.

Community members in two focus groups believed that evaluations should take place several times per year. They were less sympathetic than teachers to principals' time constraints; one noted that, although it takes time, the principals are being paid to do this work, and another community member noted that she didn't see why it's so hard.

Determine the Appropriate Person to Evaluate

Teacher participants were split about whether the best evaluator is the principal or another teacher in their subject area. Teachers in five focus groups saw principals as instructional leaders who could provide useful resources to help teachers grow, but concern was reiterated on several occasions about principals playing favorites or having personality conflicts with some teachers. Two teachers also were concerned that many principals have not been in the classroom for a long time. Teachers also recognized how limited a principal's time was to conduct effective evaluations for his or her staff. Participants in seven teacher focus groups viewed peers in their subject areas (either teachers or curriculum/instructional coordinators or coaches) as the better evaluators because they could more accurately assess a teacher's instructional effectiveness and provide more useful feedback. However, several teachers did acknowledge that it might take some work to build the trust needed in a school for a peer-evaluator system to work. A couple of teachers suggested that an evaluation team including both peers and the principal would be ideal. A couple of teachers also believed that the evaluator needed to be somebody highly trained in conducting evaluations. A couple of teachers highlighted how helpful they found their high school students' honest feedback, although surely they would not support students as the sole evaluators. One teacher said the best evaluator would be "myself. I'm there every day. I know what my weaknesses are and what my strengths are and what I need to improve."

The administrators mirrored the teachers in their views of the ideal evaluator. In two administrator focus groups, the principal was identified as the person who must officially evaluate, even if others are involved in the PD associated with evaluations. Three administrator focus groups, however, spoke of the need to involve more than just the principal. In some cases this suggestion was because it was deemed unrealistic for a principal to do a quality job according to the legislation on his or her own; in other cases it was deemed that peer or joint evaluations would provide more teachers with leadership roles. Administrators in two focus groups were intrigued by the idea of also including student, colleague, and parental feedback into evaluations, although they did not know how that strategy could be done in practice. One administrator stated that the ideal evaluator depends on what your goals are—PD or job status.

Community members (two focus groups) generally agreed that the appropriate evaluators are principals, department chairs, coaches, or lead teachers. One community member believed that the appropriate ratio is 5–7 staff per evaluator, and that the number of evaluators in a building should be adjusted so that this ratio can be met.

Incorporate Appropriate Measures in an Evaluation System

Teachers in nine focus groups expressed support for the use of multiple measures. They did not question whether classroom observations should be included. Rather, they debated whether or not walkthroughs should be part of the evaluation or solely for PD and what the appropriate

amount of time is for an observer to stay in the classroom. At least one teacher viewed the rubric as making evaluations more objective.

With regard to test-score growth, teachers had mixed opinions. In total, teachers in nine focus groups expressed support for including evidence of student growth in teacher evaluations. Of the nine focus groups, participants in five supported the inclusion of growth in test scores, but a very large number of teachers (seven focus groups) had a negative reaction to including test scores at all, and it took some discussion before the concept of growth in test scores seemed palatable. Test scores were seen as inappropriate because teachers believed there were so many factors out of their control that determined a student's performance on a test, and they wouldn't want teachers to avoid having difficult students in their class because they knew they'd be evaluated on their performance. Even for teachers who supported using test score growth, concerns were expressed about the other factors that affect test performance, teaching to the test, looking at trends, and taking into account circumstances around test score growth or lack thereof. Teachers were also frustrated that students who often show strong levels of growth but narrowly miss proficiency are still considered "failing."

In seven teacher focus groups, participants supported the idea of including a self-evaluation piece in an ideal system as a means of aiding teacher reflection. However, the idea of creating a portfolio of evidence to include in an evaluation did not resonate as much with teachers, who viewed it as not worth the time that goes into creating it. Although teachers were open to feedback from parents (four focus groups), they universally did not think that parents were sufficiently informed about their classroom or professionally trained in what constitutes good instruction to evaluate teachers. In contrast, participants in the four teacher focus groups that discussed student surveys being taken into account were evenly divided, with both supportive and opposing views expressed in each focus group. Teachers largely supported the idea as long as it only accounted for a small percent of their evaluation. Other teachers were skeptical that students had the cognitive ability to meaningfully evaluate their teachers or that the evaluations by outlier children who disliked a teacher could be misused by administrators.

Principals and district administrators also unanimously supported using multiple measures in teacher evaluations and did not express any concerns about doing so (six focus groups). With regard to using student test-score data, principals and district administrators had mixed views. In the eight administrator focus groups that discussed the issue, there was a strong consensus that including various measures of student growth was preferable to a single benchmark based on student achievement scores. If asked to choose, administrators far preferred student test-score growth over simple achievement data, believing that it could serve as a good kickoff point for a thoughtful discussion and could motivate teachers to improve. However, participants were split over whether testing data, even if used as a measure of student growth, should be included in evaluations. Participants in three administrator focus groups believed that some test scores do provide useful information; however, an equal number of focus groups had participants stating that they did not think they should be included. These participants were concerned about the objectivity of test score data, but this concern was based less on which students a teacher was assigned and more on student migration issues. They were also concerned about teachers' opinions on including test scores. Unlike teachers or community members, the administrator focus groups included a fair amount of discussion about what the law did or did not require

around student test-score data and what the implications were for local collective bargaining and union politics.

Some administrators did, like teachers, have concerns over the dog-and-pony show that scheduled observations can become. They would prefer to supplement observations with conversation with a teacher based on that teacher's reflection about his or her performance and some evidence that the teacher can bring to the meeting to support the self-evaluation. This approach to evaluation was especially supported by school principals, one of whom stated, *"I think the best learning comes when it comes from the teacher…and they have that readiness to grow. So if we could start with a self assessment, and then as we do our observations have conversations with them…."* The administrator focus groups discussed their desire for concrete and evidence-based evaluations rather than relying on their subjective opinions, although a couple of administrators were skeptical that evaluations of instructional effectiveness could ever be completely objective.

Principals and district administrators welcomed the idea of including parent and student feedback, as long as it was meaningful and targeted and "*not a rate your teacher.com situation."* The school and district administrators also unanimously agreed that the extent to which teachers performed well outside of the classroom, as team members supporting the school's mission, was important to include in the evaluation system. They wholeheartedly supported peer involvement, although whether peers should evaluate or only serve as coaches was questioned. Several administrators supported the idea of using videotaped lessons for evaluations.

Parent and community members tended to relate teacher evaluation measures to those used in evaluation systems they have experienced or heard of in other professions. They agreed that peer evaluations were valuable because peers understand the context and unique circumstances affecting an individual's performance. Several parents thought that it couldn't hurt to have student feedback incorporated into teachers' evaluations. Several parents also believed that parent volunteers in the classroom have a good sense of whether teachers connect with all of their students. With regard to test scores, parents did express concerns over weighing them too highly, given that students come to school with such different abilities, but they tended to be of the opinion that, one way or another, test scores did have to be included.

Principal Evaluation Systems

Who currently evaluates principals, and how frequently are they evaluated?

In seven administrator focus groups, participants reported that the superintendent or assistant superintendent evaluates principals; however, a few reported that school directors or other administrators carried out this task. One principal indicated that the district had assigned each principal to a member of the district cabinet for evaluation. Participants in three administrator focus groups reported being evaluated once in the fall, usually with goal setting involved, and once in the spring, to discuss how successful they had been in meeting those goals. However, an equal number of participants in the administrator focus groups mentioned only being evaluated once during the year. Principals in two focus groups indicated that they did not know who evaluated their performance.

Echoing the administrators, teachers in four focus groups reported that either the superintendents or their assistants evaluate the principal in their schools; however, teachers in two focus groups indicated that they did not know who evaluates their principals or how often.

Community members commented less on the details of principal evaluation. In three of the six community focus groups, participants stated that they knew little to nothing about how principals were evaluated; however, one community member indicated that he thought it was probably a "pass/fail" system in which principals who did not meet expectations were removed. Another community member thought it probably had not been done in the district for several years. One participant in a community-member focus group definitively stated that the superintendent evaluated both the principals and curriculum directors in her district.

What do participants want from a principal evaluation system?

Principals and superintendents report that they want a more dialogue-based, collaborative process that focuses on continual learning. In 11 administrator focus groups, participants expressed either appreciation of or a desire for an evaluation process that would involve ongoing, periodic engagement with their evaluators (superintendent, and so forth) As one administrator stated, "I like…what you were talking about, about having some conversations periodically built in. I think what we want to do with our teachers is what I would like to have done with me. I'd like the superintendent to say 'hey, show me this, you've been working on it, let's go see it in action.'" Another stated, "I like the idea, but you usually get your evaluation at the end of the year, it would be nice to have more dialogue throughout the year, so if you're in need to work on an area and you knew that during the school year, rather than at the end… So, I think once again, if it's more collaborative."

One superintendent emphasized the importance of dialogue as the lynchpin in true, systemic change: "I think a great key too, would be when the system change[s] both for superintendents or whoever is kind of looking at principals, we have to go into it thinking this is a learning process for both of us. And somehow have that open dialogue; otherwise there'll be a new system and the same outcome. I'm here to get you in some way." In addition, in eight of the administrator

focus groups, participants expressed support for linking ongoing dialogue and conversation with goal setting and evidence gathering.

One principal mentioned that he believed that the evaluation system should help *assistant principals* prepare to step up into the roles of principals through PD.

> "One of the things that I wouldn't mind seeing with this and I've thought about it quite a bit since downloading the AWSP manual was application of this for evaluation of assistant principals and those roles because dependent on where you are at, because the budget more specified in areas than trying to create a well-rounded administrator who's ready to take the next step, versus someone who focuses simply on attendance and discipline or other things."

Another principal mentioned that he believed a good principal evaluation system would help ensure collective accountability but without being threatening.

> "[H]aving worked in three districts over twenty-eight years, we get a lot of talk time, if you will, to context of accountability to outcomes….So, I think that's a key piece of a future system is that, how do we assure that people aren't threatened if you will, but collectively accountable and hold each other accountable to at least surfacing the conversation around results, achievement results."

One teacher focus group also emphasized the importance of facilitating productive, ongoing dialogue between teachers and administrators as an important role that a principal evaluation system should play.

In three community member focus groups, participants were quite adamant that principal evaluation ultimately needed to be a tool for accountability that would allow poorly performing principals to be ushered out of the profession.

What do participants believe is the appropriate use of principal evaluation results?

In five administrator focus groups, participants expressed support for using evaluation results to guide PD. One principal expressed opposition to the idea that evaluation results should be tied to merit pay, noting,

> "When you're in the trenches you know how hard the job is and how busy we are doing everything. I just hope that they stay realistic and they don't start going down some path of a merit pay system where your kid's 90 percent pass so you get a bonus check or something like for teachers or principals. I think that would be disastrous. It's been tried before, it's never worked. But this does have that feel to it."

In one administrator focus group, three or four principals indicated that they did believe evaluation results should be considered when making promotion decisions. Administrators spoke less about their views on how evaluations should be tied to termination decisions; however, two administrator focus groups expressed concern over recent decisions to move or fire principals who, while not meeting adequate yearly progress (AYP) goals, have made strong, continuous progress in improving school performance. As one principal stated:

> "[O[ne thing that I would hope would be in any criteria would be similar to what's in the teaching one, is that as long as you're growing and as long as you're demonstrating growth for your building, that regardless of the SIG rules if you're under performing in AYP, all four of those criteria, remove the principal, remove the principal, remove the principal, remove the principal. We've had principals removed across this state that have taken graduation rates from 20 percent to 60 percent in[a] three-year period of time, but it's still not high enough, the principal is gone."

Community members believed strongly that evaluation results should be used to make termination decisions for principals. As noted earlier, participants in three community focus groups explicitly referenced terminating poorly performing principals as the purpose behind principal evaluation. One community member stated that principals should be held accountable, and termination decisions should not be made based on length of service but by results. Another community member believed that poorly performing principals were very difficult to remove from their positions, and the new system should address the problem.

What do participants cite as best practices and flaws in the existing principal evaluation system?

Best Practices

Participants in 11 administrator focus groups overwhelming described best practices in the current system as instances where they were able to have rich, ongoing dialogues and conversations with both their evaluator and with stakeholders affected by their work. In four administrator focus groups, principals spoke favorably about their superintendent asking them to tie their goals to Interstate School Leaders Licensure Consortium (ISLCC) standards or expressly using ISLCC standards in the evaluation process.

Principals uniformly appreciated feedback and productive conversation around setting and achieving goals. Several principals indicated that they found gathering data and evidence to assess their progress in meeting school goals to be helpful, particularly when conducted through a collaborative, group-oriented process.

> "[O]n the principal side, we are using these criteria, but we really base it on an initial goal setting conference with the principal and quite a discussion of what their achievement goals for the school are for the current year and how they're going to get there through professional development. Then we're in the schools and in the classrooms a lot. A lot of that being with principals and a conversation usually in April or May, that's also fairly lengthy and in depth, show us some data, show us what you've done since you set these goals at the beginning of the year. So on paper it's these, but it's a deep and ongoing conversation."

> "I've had good evaluations at other places. You do some goal setting, planning. You're graded on performance. What you're doing. How you're moving the school forward, how you're working with the staff. People actually come and observe, see you do it. You have conversations about what you're doing, where you're going and how you're going to get there. You work as a team, as a group, to do that sort of thing administratively. And you're all graded individually but you have real benchmarks to shoot for and hit. I've been doing this long enough that I do it anyway now on my own. But those are really beneficial because they can also play things out that you hadn't thought of, more brains working are better than one. So I need more brains and that helps."

One principal specifically mentioned that he had found using the Washington Achievement Index particularly helpful:

> "So I appreciated my supervisor bringing that to me and in the goal setting arena it's very open and I feel like we're empowered to come up with the solutions to address something that's clearly measurable, and probably similar to what a teacher would feel with a principal if they're on professional growth. I'm going in this direction and needing support on that. So I liked the goal point of that and also this year we're using, you're choosing your data points to [Inaudible]. So this year using that Washington Achievement Index has been good for us."

Another principal described his district's evaluation system in detail and noted that he really appreciated how evaluations were well aligned with district initiatives, which provided clarity and direction. The principals are evaluated by members of the district cabinet, who meet with each principal at least once a month for a focused dialogue related to a specific aspect of their evaluation. The meeting allowed both the principal and evaluator to "check in" about the way they were progressing in an area specifically related to a district initiative and to address specific struggles or challenges.

Although only a few teachers commented on this topic, one teacher did mention that her school had created a school climate survey and was repeating the survey for a second year to look at trends. Another teacher commented that her principal met three to four times a year with the superintendent for evaluation of specific goals, a process that had been successful from her perspective.

One community focus group was quite vocal in its support for the use of school improvement plans in creating accountability and dialogue between the school board and the principals. As one member of the group stated:

> "I think one of the best legal requirements we have in our district, and one of the best state mandates that we have, is a requirement that we use school improvement plans. That keeps us in touch. And that gives us the power to ask the question, is it working as well in South [Inaudible] as it is in Manchester? Are those kids receiving as good of care, education and services as they are in other places in our district?"

Other participants in the focus group agreed and added that they particularly appreciated the emphasis on parental involvement and feedback that is incorporated into most school improvement plans. One community member commented on this topic, noting that his area had "chief academic officers" who were each responsible for a particular region, which helped by having an extra "ear to the ground" for evaluations.

Flaw 1: Insufficiently Frequent Evaluation and Feedback

Principals raised concerns over the frequency and type of feedback they receive in the current evaluation system. Participants in six administrator focus groups indicated that they often received an evaluation only once a year, although participants in seven administrator focus groups indicated that they found the feedback they received relatively meaningless and unhelpful.

Several principals were frustrated at being evaluated only once per year and expressed a desire for a process of ongoing dialogue or conversation throughout the year. One principal mentioned she was concerned that the evaluation would be based on a very limited, occasional observation of her duties, although another principal indicated that his evaluation lacked observational content altogether.

> "When you talk about district personnel coming in to evaluate me, how often are they in the building? You want to talk about snapshots of a classroom, the snapshot of my building is even further out there as to when they come in. So my concern is, again we come back to that concrete piece of, is that conclusion being drawn based on a board

> presentation, is that being drawn based on how many angry parents bother the district office during the day, and if it is test scores, are they aware of the school improvement plan that we've conducted with all the work that we're doing in there, are they truly up to date on that? So I see it as more of a stretch for our evaluations to be meaningful than what we can do with the classroom teachers, since we're there a significant amount of time."
>
> "Ours are based on the Interstate School Leaders Licensure Consortium (ISSLC) standards. Then there are comments underneath based on the standards. We have goals that we have at the beginning of the year. Have we met those goals? And then just, they're doing this or they're doing that. But to sit down and formally watch me doing something, no, that doesn't happen. Or to ask me 'how it's going?' That doesn't necessarily happen either. It's usually at the end. 'Here you go, it's June 1st. You better have it by now.' "

Flaw 2: Burdensome Self-Assessment

Although a number of principals and superintendents expressed support for self-assessment and reflection, other principals found this approach less helpful or more problematic when not tied to detailed feedback and actual dialogue. Two principal focus groups expressed frustration with having to spend extensive time creating portfolios through a self-assessment process in which they collect data and present it to the superintendent who merely signs off that it has been completed. Principals spoke of the process as a "show" and referred to it as "ridiculous." One participant described it as a waste of time unless it is tied into a clear, effective, collaborative dialogue that provides feedback. One superintendent in another focus group shared these views, stating: "I'm really not into giving principals a lot more work. If I'm going to have them do this big portfolio, is it really that good? There's a lot of work. It's like, why do that?...You could spend hours and hours putting together this book. Or you can do your work as a principal and we'll come in and do it with you. As long as you're really clear. In the end there's got to be a form. It's much more effective and collaborative." Another principal focus group reiterated this view and suggested that their superintendents were not in the building often enough to really know enough to evaluate their work and were relying too much on self-assessment alone.

Principals also raised issues with consistency across evaluators and time. In four principal focus groups, participants believed that the evaluation process was inconsistent, both across districts and across time—the process often changed from one year to the next depending on staffing changes. One principal indicated he did not know how he was evaluated or by what standards: "I'm not sure how our superintendent evaluates. To evaluate, you have to be in the building and see people working. I don't see them in the building. I don't know where he's getting his information. My evaluations show up on my desk." In another principal focus group, participants discussed the lack of helpful feedback they received when self-assessment tools were not tied to a meaningful evaluation process:

> "I think something that isn't working is, we have a great superintendent, but how much does he really know firsthand? I think it's a lot of second hand or just very general perceptions, but I'm being held accountable for this and he's in my building periodically, but not for very long. I see a lot of heads nodding, so you all understand. Then he asks

> me,' How is it going, how is it going, how is it going?' which I appreciate the self evaluation…."
>
> "You say, 'good.' "
>
> "It's going great, best year ever."
>
> "Four."
>
> "Yeah."

Another principal stated that he believed that there was a lack of consistency in how principals were being evaluated on implementing districtwide initiatives. Still another had seen his evaluation rates rise and fall from year to year but was unable to receive an explanation for the changes from his superintendent.

Teachers regularly commented in their focus groups that they believed that they did not know enough about how principals are evaluated. In all six teacher focus groups, participants wanted to have their feedback as teachers solicited and included in some way, although they were unsure how to do so formally. One teacher focus group indicated that they would like their input included but feared possible negative repercussions or being reprimanded, especially in small districts. In three teacher focus groups, participants believed that the current process was too narrow because it seemed to include only brief walkthroughs or short meetings with the superintendent. In general, most teacher focus groups emphasized the need for greater transparency in how principals are evaluated.

> "[T]he principal should be the instructional leader, and therefore, they're the ones who should do this first. They should say, this is how I'm measured, it's completely transparent, I want to be measured this way. I want your input. I want the parents' too. I think that's the way it really needs to be. If a principal did that, I think the inspiration to the staff that they're going to lead would be phenomenal. You stepped out there on that limb first, wow! That's new and different. It's been my experience that when those things come along, usually it's circle the wagons and keep from change, rather than embracing it and saying we can get better and we should get better."
>
> "[I]f that could be transparent and principals would be able to say, these are my goals, this is what I'm going to do….Then it feels like we're all in it together to improve."

What do participants believe an "ideal" principal evaluation system should look like?

Incorporate Appropriate Measures

Largely in line with their comments about flaws in the current system, principals and superintendents believed that an ideal principal evaluation system should have clear expectations and standards that are linked into a communication loop organized around goal setting, evidence gathering, and regular feedback. Six administrator focus groups believed that growth in student achievement was an important and useful measure that should be included. Two administrators believed that including growth in student test scores was an important measure to include because changes in scores provide clear motivation and direction to principals and give principals support in confronting teachers who need improvement. Another principal mentioned he strongly supported using the new ISLCC standards for principals because it allowed principals to be evaluated on multiple criteria and it matched up with the new teacher standards.

In slight contrast to some principals' characterization of too much self-assessment and self-reflection as flaws in the current system, five principal focus groups mentioned that they wanted more opportunities to do self-assessment or to set some of their own goals. Three principals wanted to be able to engage in a "360 degree" assessment that gathered feedback from multiple viewpoints; however, one principal qualified this suggestion by emphasizing that the information should be used only for professional growth and not be part of his formal evaluation.

Participants in six administrator focus groups mentioned that community, teacher, and student input should be included as a measure, but again, they were unsure how to do so. One principal emphasized the importance of student feedback but wanted the feedback collected informally and from student leaders. Other principals noted some concerns with using surveys to collect feedback from students and teachers. One principal believed that the adversarial nature of his job often made him unpopular with some of his staff on any given day, and the survey results were unlikely to be stable or representative. Another principal was concerned that, with a small number of respondents, one or two people with very high or very low scores could easily skew a survey.

Teachers were more specific than superintendents and principals about what they believed should be included in evaluating principals. Communication was a frequent mention, with five teacher focus groups highlighting this measure as an important aspect that should be included. Teachers wanted principals to be evaluated on how well they communicate with staff and the community, how well they build and maintain relationships with them, and more broadly, how successful a principal is at creating an open environment where teachers can speak without fear, be heard, and listened to without judgment.

> "[A] principal who fosters respect for him and his position and for the administration at the school… Teachers need to have respect.…The principal I have now is the best principal I've ever had. And I can talk to him about virtually anything. I can say anything to him and I'm not in fear of any kind of a [Inaudible] or anything. I know he will listen to what I say. And I know that he will respect it, coming from my opinion. He may say no to me. He may say, I can't look at it that way. But he will listen and he will not judge

> me for it. And that's something I do not see in this principal. Teachers should feel comfortable talking to their principal about almost anything. Almost anything can happen."
>
> "I've been a teacher for 33 years. The most effective administrators I've worked with were great communicators. You knew their expectations for the building, you knew their expectations in the classroom, you knew their expectations for working in the community and your parents. Open, clear, constant communication. A good evaluation for a principal should give them feedback on how well they're communicating with their staff, with their community, with other buildings in the district, especially in this society of [response to intervention] RTI and [professional learning communities] PLC. How are you getting that across? How are you making known the needs of your staff? We had a colleague over here mention, same building three years later, still the same things going on, there's been two regime changes. Well, that's a communication issue and that should be part of the process of evaluating principals. How effectively do they communicate? Just like we're evaluated on how we communicate with parents."

Seven teacher focus groups referenced either instructional leadership or knowledge and wanted principals to give evidence of how they supported teachers in achieving their instructional goals. Two teacher focus groups mentioned that principals should be evaluated on how well they understand and utilize the professional training and background of their staff.

Teachers also emphasized school climate in six focus groups. Although teachers wanted their own feedback included (six focus groups), their openness to a survey format depended heavily on the careful design of the survey and the usefulness of the feedback it would provide. One teacher noted that the survey data should be for the principals to use as formative assessment and should not be part of their summative evaluation. Another teacher expressed support for teacher and student survey data but only if it was carefully weighted against other data.

Community members were also very specific about what they wanted included in principal evaluations. Community members echoed the ideas of many teachers by emphasizing including measures of how well the principal communicates and builds relationships with teachers, parents, and the community. A few expressed support for doing this evaluation through surveys (four community focus groups). Two community-member focus groups mentioned the safety of the school or the broader school climate. One community member noted,

> "The principal is the one that kind of sets that example when you walk in that door. Like I said, you feel it. You went from one school where there was a terrible principal, she's now gone and oh my gosh, these teachers that were going to retire or leave were staying because the administration has changed and there's that feeling that the community is welcome that they want our input, they want parents there. They want what's best for the child. And you feel it."

One community member suggested the use of "smart" goals—or very targeted, specific goals that the principal sets based on his or her school's own needs. Community members' views on including student achievement data in principal evaluations are somewhat difficult to discern. Explicit references to student growth or student achievement data related to principal evaluation

are infrequent. One community member agreed that student achievement data should be included. When asked whether student growth scores should be included in principal evaluation, another community member noted, “It does. I think in our district, the test scores…if a school is constantly a school in improvement…and [has] the same principal for maybe two or three years, I would say that that principal would be out of that building in a heartbeat.” It is possible that community members largely assumed that student achievement scores were already included in principal evaluations and thus believed it more important to focus on things they wanted included *in addition* to general student achievement data. The paucity of data does not allow us to make a general conclusion on this point; however, the infrequency of discussion of student achievement data in principal evaluation should not be interpreted as a lack of support for its inclusion.

Determine the Appropriate Person to Evaluate

Administrator focus groups did not discuss to any great degree the topic of appropriate evaluators. Although a number of principals raised the problem of disconnected or unengaged superintendents as a flaw, participants did not follow up with suggestions regarding other possible persons to evaluate principals. In general, principals and superintendents largely agreed that principals should be evaluated by the superintendents or one of the assistant superintendants; however, one principal believed quite strongly that the evaluation process should be conducted by an external, professional team, particularly if student data were being included because most superintendents are only novices at interpreting such data correctly.

One teacher commented that she believed that principals should be evaluated by more than one person because she would not want her own evaluation to come only from a single evaluator. Community members largely concurred with principals that the superintendent or assistant superintendant should conduct principal evaluations.

The State of Washington's Recent Education Reforms

What do participants believe are the best aspects of recent education reforms?

When asked about the best aspects of the recent education reforms in the state of Washington, participants in six administrator focus groups cited the cultural shift to a more flexible, collaborative, and growth-oriented model of educator evaluation. One administrator expressed the importance of this shift in his view as follows: "I think this is a huge cultural shift for us…[W]e're going to have to take the time to do this because it's so important. And we do have a moral obligation to make sure kids are achieving. This is one way to get there." One principal believed that that the cultural shift would help bridge the gap of distrust that often exists between teachers and administrators:

> "The one thing that I thought, since I've been in my role, is that I came [from] sixteen years in a classroom into an administrative role…I'm very much more of an administrative coach. But the history and the climate in the school—there's been such distrust of administrators that it's very difficult for them to believe that I am authentic. [Laughs] And not evaluative in what I'm saying. So I think that changing the structure that allows that relationship to change is going to be huge in the field."

In addition, administrators were quick to mention that they liked the shift to a wider focus and using multiple measures in both teacher and principal evaluation (five administrator focus groups). As one administrator stated: "I think the new evaluation is going to open up that can and give us the opportunity to look at teachers out of the classroom, how they work with their colleagues. How they perform in collegial groups and with parents and students. Before that was taboo. It'll be helpful to be able to open it up to all their performance, not just one." Compared with teachers, administrators were somewhat more likely to mention the inclusion of student data as an aspect of the reforms that they liked; however, they often qualified their statements by emphasizing that the data's most valuable role was as a point of reflection and collaboration for teachers and administrators in strategizing how to adjust teaching in the classroom to best suit students' needs.

Administrators also frequently cited moving to a four-tiered system with clear criteria specifying how an educator can move from one level to the next as a significant improvement:

> "I think part of what will happen is, principals if nothing else tend to be kind of evaluation motivated. So if they're on a one through four scale too, so that principal really wasn't putting much energy in evaluations and they get a two, is going to be motivated to get after it a little bit more and especially when they know what the other folks in the district are getting. If you want to get a three or four, this is what you've got to do."

Administrators in three focus groups also mentioned the high level of stakeholder input that the state has sought out. One administrator commented that, "At least with this process, which is very dissimilar to other things the legislature does with education, they're seeking professional input. They're coming to us and asking for some ideas on what it should look like."

Teachers echoed administrators in voicing their appreciation for the level of professional input being incorporated into developing the new evaluation system. Participants in five of the teacher focus groups explicitly discussed stakeholder involvement as an aspect of the reforms that they appreciated.

> "I think more conversations like this need to be happening across the state with the very people that are working in a classroom. And whether it's the ESD's or the people from [Office of Superintendent of Public Instruction] OSPI, those people making the decisions need to be asking us these questions. I feel this is the first point in time that I myself as a teacher have been actually asked questions and listened to and my input has been taken into account or seriously."
>
> "…the fact that our district put teachers in charge of developing the tool. Teachers, in conjunction with the EA president and having administrator support….But teachers are the ones developing the tool that will be used to evaluate their work."
>
> "I think for me, it's the fact that it's everybody together working on this, it's not somebody from the outside trying to impose their own point of view, it's the educators coming forward and saying this is how we can improve our profession."

Several teachers referred to the state's participatory approach to developing the new system as "empowering." One teacher felt the process would have a long-term effect in changing the way teachers and nonteachers view the profession:

> "This is going to empower teachers. I really think, because we hear sometimes, "Why would anybody want to be encouraged to work in public education?" I think the opposite. This is a really good opportunity for teachers or people that are thinking of getting into it because it empowers teachers to be a force behind a change. Like you said, if you're really enthused and supporting the whole concept, you're sitting in PLCs or whatever and we're seeing what we have in our building isn't working….As a group we're working together and we're willing to change what we're doing to a different process for the betterment of the kids, making the whole school better. I think this is a very powerful thing right now."

In a related vein, participants in four teacher focus groups also cited the cultural shift to a more collaborative and growth-oriented model of evaluation. Teachers expected the new system would be a tool for improvement and growth rather than a punitive "gotcha" for principals and teachers. One teacher said: "I can get real feedback on my instruction in the classroom and leadership within my building and be able to do something about it. Just real feedback." In one teacher focus group, multiple participants said the emphasis on collaboration was what they liked best.

Like administrators, teachers in one focus group explicitly mentioned the move to a multitiered rubric with clear criteria for advancement. As one teacher stated, "I like the idea, when you're looking at that four-tiered system of criteria. If they say I'm at a basic or a proficient, what do I need to do to get to that next level? As we're doing with our students right now, what do I need to be changing to get to that next level? To become proficient or distinguished? What would that look like?"

In another teacher focus group, the conversation centered on the chance to use multiple measures of student growth and achievement rather than a single test score. Several teachers were appreciative of being able to include multiple forms of student data. In addition, two teacher focus groups specifically mentioned that they liked the fact that principal evaluations were being overhauled at the same time, and that the new system would include an emphasis on instructional leadership when evaluating principals.

Community members shared the views of both teachers and administrators on the benefits of making the evaluation system oriented towards improvement and using multiple measures to evaluate teachers' performance. One participant stated, "It sounds like it is a tool for improvement, it's not a punitive catch you doing badly. I think that's really what we want…That's the whole thing of education is to improve how you are, not to catch you doing badly. So I like that it's got that vibe to it." Another participant stated, "Increasing the quality of instruction and I also like that it includes communicating and collaborating with parents and partnering with the school community." One community member explicitly linked the shift towards growth and improvement to a wider evaluation of teacher's jobs, including parental engagement and student growth:

> "[T[he changing in the tone and tenor of what is teaching and learning. You actually have to use evaluation of parent engagement, family engagement, how well you're working and playing with others in your building. The concept of growth models as opposed to strictly, did you do a good job this year? It doesn't matter how your students performed. Well it does matter how your students performed. And if you have kids that are on the low end of the scale and you, on average, brought every kid up 30 points, you're a great teacher. But if you, on average, brought each kid up two points, you're not a great teacher and we need to have you on a plan of improvement."

What are participants' major concerns about the recent education reforms?

Clarity Over Next Steps

Administrators and teachers expressed a number of concerns about the lack of clarity surrounding how the final evaluation system would finally be selected, how it would be successfully implemented, and how the evaluation results would ultimately be used. In general, the lack of clarity and uncertainty on these points came up frequently in all the focus groups.

District administrators in two focus groups were heavily concerned with how the new evaluation system for teachers would be negotiated between the districts and the various levels of teachers' unions. Administrators were particularly concerned over whether or not the new evaluation system would be bargained locally in each district or whether leadership would be provided at the state level.

Use of the Evaluation Data

Administrators in six focus groups were specifically concerned about the lack of information regarding the way that the state ultimately intended to use the evaluation results. A frequently cited concern was how the state would protect teachers' and principals' evaluation data from being misused. A frequent reference was the recent misuse of student achievement data by the Washington Policy Center. As one administrator expressed after the Washington Policy Center was raised in the discussion, "That's the thing that worries me most about the state and they haven't thought this through very well and when you ask them what they plan to do with this information, you just don't get a sense of confidence that they really know what they plan to do with this stuff."

Several administrator participants were concerned about how a summative score used to rank teachers might affect teacher buy-in. As one superintendent stated:

> "[O]nce we have this system where a teacher can evaluated, that was a 3.2 teacher, this is a 3.6 teacher, that's a 2.9 teacher by looking at scores from the eight different criteria, my concern is again, what is the state going to do with that information because they're already introducing bills which can go anywhere now, because of our current system, but there's a mindset that we can now use this new system for writs and other kinds of things, which isn't why [our district] got involved for sure, we got involved for professional growth, not to identify some weaker teachers in the event writs. You have that seniority versus competence debate in another setting, but not around this, I don't think, or that's going to kill it from the teacher perspective, they're going to go, the hell with that, we didn't sign up for that."

Administrators were especially concerned that principal evaluation results would not take demographic or district-specific challenges into account. Two principal focus groups cited the recent removal of a number of colleagues from low-performing districts without adequate consideration being given to the specific challenges their colleagues were addressing. Many participants expressed the desire for further clarification on how student data, specifically test scores, would be incorporated into the evaluation system.

In three teacher focus groups, participants had extended discussions that raised concerns that echoed the administrator focus groups. First, some participants were concerned that evaluation data would be used to rank teachers and undermine the cultural shift toward a growth-oriented model by reducing teachers to a single, summative number. As one teacher expressed:

> "I am concerned that ultimately through all of this work that it's not going to be a professional growth model. It's not going to be about true growth for the teacher. It's going to end up being punitive. It's going to be somehow manipulated down to being about the numbers. Some little negative information that can be easily put in the news, on the newspaper, and then it really won't be the growth model it should be. It should be a growth model."

> "You were talking about the idea of averages. If you take these eight categories, you put sub items in them because it should be rich and dynamic. Everyone gets a number and you just average it. Well now everything holds the same weight and maybe your classroom management might be more important than whether you're participating in school committees or something like that. Yet somehow that all gets equalized and we want to create some massive, mathematical model? That's not practical and it's not realistic. And it does belittle the rich conversations that the rest of it is all about."

Other teachers expressed concerns about how student achievement data would be incorporated into the final evaluation score: "Student evaluations being connected to it. That's my only concern, student progress, because each student progresses differently and how are you going to use that as part of the evaluation tool?" Another teacher was opposed to having a larger percentage of the evaluation be based on student achievement: "Way out there, originally what concerned me was discussions that all you have to do is have 51 percent of the evaluation be based on scores or data, which was discussed by some people. That makes the rest not really worth anything if all of a sudden everything is on that data. They need to be really careful how they use the data from students."

Participants in one community member focus groups also raised similar concerns about the use and protection of evaluation results. Like administrators, several people were concerned about misuse of data. As one member put it, "[I]t needs to be protected, just like the Washington Policy Center got a hold of the state boards accountability index and twisted it around and made it look like some schools are just failing, terrible use of statistics."

Implementation

District administrators were uniquely concerned with how the state plans to ensure interrater reliability and fidelity in the implementation process across districts. In nine administrator focus groups, participants had extended discussions about the difficulty of ensuring consistency and comparability across districts although still allowing enough flexibility for local modification.

> "I think the challenge is going to be, and the importance of this whole model, is going to be what components of this need to be uniform and what components of this need to have the flexibility for local input and use. We know the four tiers are going to be the same. We know the criteria is the same, what we really have to come up with is operational

> definitions of those criteria so there's not a randomness around the state about what each criteria means. I think the rubrics, the indicators of what is successful in that area and how you pick your framework is going to be important to have some district control in it. Otherwise you're not going to have ownership."

Of particular concern was the way different districts and principals may approach the four-tier summative scoring for teachers. A number of administrators indicated that they feared that if results were made public, administrators would be incentivized to continue to rate the majority of teachers as 4's; on the other hand, if the districts only allowed principals to rate a small number of teachers as 4's, several administrator feared it would undermine the evaluation process as source of motivation for improvement.

Multiple participants in three administrator focus groups were deeply concerned that the state would not provide adequate funding for the PD needed to train principals and teachers in using the evaluation tools. Other administrators wanted more information on what process there might be for calibrating the various evaluation instruments. In five administrator focus groups, participants raised concerns about the way student achievement data would be incorporated in teacher evaluations. Participants raised the substantial problems of migration of students between schools and districts and the need to account for differences in school context when considering student achievement and what constitutes good instructional practices.

Instability of the Legislative Process

Administrators were significantly concerned that changes to the state legislature over the next two years might reverse the grassroots, collaborative approach that has been taken thus far. Participants in six administrator focus groups raised this issue as one of their concerns with the new education reforms. One participant expressed concern that the state already seemed to be signaling that it was backing away from the growth and improvement approach and moving towards a more top-down, punitive approach.

> "All of that [Inaudible] what will legislature look like in 2013? We have no student evaluation data in the evaluation plan there, they could look and say, 'Excuse me, what was the intent?' That's the thing that's going back and forth. Some [Washington Education Association] WEA folks in maybe a weak moment will amend that."

> "I just think that OSPI has been so decimated with reduced staffing and reduced resources, so when we state involvement, are we talking about the legislature?. The governor's new education, if that comes to fruition….I mean, it sort of reads counter to what's the [Inaudible] shared as the most valuable part of all of this being the process, the engagement, the ownership and now some outside entity is going to come in and overlay something."

> "Even amongst some of the other districts though, last year, I remember when I was at Kennewick when we were going through this process, there was still a lot of buy-in from other districts about the process itself and I think Dave hit on something is, when some of the new legislation is being proposed, I see people backing out now from where this was a very processed based, to now it's starting to get twisted by a change in state government. And some of the things that are being said are just… they're flying [in] the

> face of what a lot of people felt was a very, not good process to start out down the road. And it's already….It's just taken a year and a slumping economy to start whacking it, for lack of a better term."

One teacher was particularly concerned that the state would end up changing important aspects of the new evaluation system after the pilots have concluded their work: "What concerns me is that somehow, government level administrators will change something after we've put this much time in it. They have and they've done it before."

One community member also noted similar concerns that after, "[h]aving seen legislators kind of back away from promises before, I'm still concerned that this is going to be another long battle about very finite details that some will be important, some will not be, but it will end up derailing the whole thing. It's still a concern that I have."

Timelines

All stakeholder groups repeatedly expressed genuine concern over the potential time burden that the new evaluation system will pose to educators, especially the principals tasked largely with carrying it out. In multiple focus groups, administrators repeatedly expressed their support for a new system and their concerns about how they would manage to carry it out effectively, given their already full plate of responsibilities:

> "I'm in total support and agree with everything, but where in the heck am I going to find the time? I mean, there are so many things, when you're the only principal or when you've got another principal there too. There are so many things that we have to do that finding the time on doing this new system worries me a little bit. I'm stretched."

> "This is going to take more work. The state has not accommodated that. This is a tool that we're looking at right now, but the system, what about mentoring, what about peer assistance, what about peer review, what about freeing up principals to be trained and to keep current? It will be more work. It will be more costly. What can we take off the plates of our administrators to free them up from some of that? These questions the state has no response to."

> "The unfortunate part is, this whole thing, whole conversation, whole process isn't going to change the fact that this is supposed to be 80 or 90 percent of our job and the reality is we have 5 percent of our time to spend on it. Nothing is going to take away the day to day management, providing a safe environment, all the different things we have to do."

Teachers in five focus groups were concerned that the principals are already overburdened and that the new evaluation system may require changing principal responsibilities to enable the evaluations to actually succeed. As one teacher described his concerns:

> "I think I hear a couple of us saying that it needs to be more than just the principal. I also wonder that to do it every year… principals working in my high-needs schools are working their tails off and they never give instruction what it's due and they know that and they feel terrible about it. But we haven't yet figured out what we can take off their plate in order to make that happen. A kid gets busted with a weapon, there goes your whole day. So what can you take off their plate so they can be part of it?"

Teachers in two focus groups were particularly concerned that the short timeline for completing the pilots, training evaluators, and rolling out the new system was much too tight: "I'm most concerned about the timeline….We're still in the process of developing the tool. We are under great pressure to figure out the implementation system. How do we get teachers trained on what they need to be able to show and do? More importantly, how do we get those principals trained so that they know what kind of evidence is supporting evidence in each of those criteria? And it's next year." Some expressed the view that the state should honor the work the pilots have already done by giving them adequate time to carry out the work: "I think the state needs to back off and let the groups that are working on this, a very hard focused thing, and honor that work. These are educators and people in the field that are doing…I appreciate the programs that are being piloted for us."

Community members also expressed the concern over the extra workload the new system was likely to create for principals: "I don't think it can be all administration, because there's only three, one to a building, four maybe five, I don't know, but that's because there's that many more teachers. You've got to have something that is not all on the shoulders of the principals to make it happen." Interestingly, community members did not raise the shortened time frame as a concern. In fact, one participant felt the short time frame was a good aspect of the reform process.

What do participants think the state government's role should be in education reforms?

Development of Models

District administrators held four distinct viewpoints on the state's role in developing educator evaluation models. In three administrator focus groups, participants believed that the state should mandate a single mode across all districts. Respondents supporting this view believed that it would simplify the bargaining process with unions and ensure greater fairness across districts (especially if the state resorts to ranking teachers and principals); in seven administrator focus groups, participants stated that they believed a single model could be acceptable but only if a waiver process were in place to allow districts to modify the mandated model to fix local needs and contexts. In contrast, other administrators were opposed to a single, state-mandated model and believed it would make the evaluation system a compliance model rather than a growth model; moreover, it would create problems at the implementation stage by lowering stakeholder buy-in and diminishing local autonomy. Instead, participants believed that the state should offer a menu of model options or allow districts to choose from among the piloted models (seven focus groups). A third view, expressed in six administrator focus groups, suggested that the state provide a single framework with clear criteria and common language but allow districts to choose multiple ways to satisfy that framework in terms of instrumentation and evidence.

Teachers were much more unified in their preference for greater local control over the evaluation system design. No teacher focus groups expressed a preference for a single, state-mandated model; however, participants in three teacher focus groups indicated they would support a single model if a waiver process were in place to allow for flexibility and local modification. In line

with the administrators, some teachers believed that the best role for the state was to offer a menu of options, drawn from the work of the pilots, from which districts could select. Participants in three teacher focus groups expressed a preference for this view. Finally, participants in two teacher focus groups preferred that the state provide a common framework, criteria, and common language but allow districts local control over selecting instruments and evidence to satisfy the framework.

Similar to administrators, community members' views generally fell into one of four categories. Participants in two community-member focus groups expressed a preference for a single, state-mandated model, although participants in one community group believed that a single model would be appropriate only if there was a waiver to allow local modification. Participants in four community focus groups believed that the state should offer a menu of evaluation models and allow districts to choose from among them. And finally, in all six community-member focus groups, at least one participant expressed support for the state providing a single framework with criteria and common language but leaving the selection of instruments and evidence open for determination at the local level.

Implementation Support

Beyond the development of models, principals and superintendents believed that the state's role should be to provide the funding and resources that principals and teachers need to implement the new system with a high level of fidelity. In six administrator focus groups, participants specifically cited the provision of PD targeted at understanding and implementing the new system as the most important role of the state.

> "For me, I haven't had a lot of experience where they've done a great job of the professional development and coming up with the money that really needs to be given to the districts to provide that opportunity. I think that they shortchange it, then people are going to piecemeal it in and it's just going to be the same old crap… sorry, same old thing in a different way."
>
> "[L]ooking at growth and supporting professional development, you're right on about that too, because we can't financially within our own districts give the kind of professional development they talk about cutting the year down to 178, and I'm thinking we need to bump it up to 185, 84 of those 185 days [Inaudible] teacher development days."

Principals and superintendents in five focus groups believed that the state should play a role in ensuring consistency across districts in the criteria and calibration of evaluation tools. In addition, participants in two administrator focus groups indicated that they wanted the state to take a larger role in dealing with teacher unions to help simplify the amount and complexity of bargaining at the local level. Finally, principals and administrators in two focus groups stated that the state needed to assist in getting further information out to teachers about the changes and in generating teacher buy-in. As one administrator put it, "[A]nother role the state should play in this whole piece is to get the teachers to understand the value of it for their growth and for student instruction, that it's not punitive, and to get their buy in, not just as a hoop to be jumped through."

Teachers echoed administrators in calling for the state to provide funding for resources and training surrounding the new evaluation process. In six teacher focus groups, participants expressly stated that they believed the state's role should be to fund implementation support.

> "I think the state needs to step up to the plate when it comes to the funding. If they want us to show this kind of success in our students, if they want us to change, they have to pay for us to change. The state can't expect high expectations without the funding to make it happen."

> "PD around every single piece of this change. Principals have to have some kind of support. They just have to. Clarity on what exactly these criteria are for everybody in your [school] is going to take time and PD…There's just a ton of PD that's going to be needed about, how do we truly assess what that means? Like I said, even in a single district, two different buildings might have a different concept of what a professional decision is, on the part of a teacher. You've got to have clarity around those things and all that takes somebody, even not have the answer, to come in and facilitate that process."

Teachers in three focus groups emphasized the need for the state to engage in further outreach efforts to "get the word out" about the recent reforms, the pilots, and the steps going forward.

Community members shared administrators' views that the state's role should be limited to ensuring consistency across districts and fidelity of implementation. Participants in four community member focus groups spoke of general guidance and oversight as an appropriate state role. One community member stated that it was important that the state provide funding and resources for PD related to the new evaluation system.

Monitoring, Rewarding, and Sanctioning

All stakeholders expressed similar views in favor of a limited state role with respect to monitoring evaluation system results. Participants in five administrator focus groups shared the view that more expansive roles like monitoring individual teacher's scores at the state level would be impractical, and the state lacks the resources to do so effectively. District administrators believed that ensuring consistency across districts, assisting in the calibration of raters and tools, and assessing fidelity of implementation should be the limits of state oversight.

Teachers consistently believed that the state should not play a role in rewarding and sanctioning individual educators on the basis of evaluation results. Many believed that doing so would undermine the collaborative and growth-oriented culture shift the evaluations were aimed at achieving. Some teachers believed that rewards and sanctions that were based on an entire building's performance or that identified excellent teachers for mentoring roles would be appropriate because this system would contribute to a collaborative culture rather than undermine it.

Community members in two focus groups were more open to rewards and sanctions for individuals, but only under certain conditions: (1) if the emphasis was heavily on rewarding excellent teachers, (2) if it was carefully balanced with local control, and (3) if it actually contributed to improving schools. In general, community members (four focus groups) spoke about the state's role as one of holding the districts accountable—however, the participants did

not explicitly specify the nature of that accountability. Community members did not endorse the monitoring of individual teachers evaluation scores. Participants in three community focus groups believed that doing so was not a practical or a desirable option for the state.

Legislating Education Reforms

Teachers and administrators expressed concern and frustration over the legislature's lack of expertise in public education and its lack of knowledge about educators' day-to-day work. They repeatedly stated their concern over legislators' inappropriate reform models, especially corporate- or business-based models. They both emphasized the *human* dimensions of education and noted that educators are not producing a material product:

> "It's almost as if they're using a manufacturing model and applying it to education. We're not building a car. We're not building a bigger house, we're working with kids, we're working with people. One size does not fit all." (teacher)

> "I think the key is, and politicians probably need to realize this….For a long time, I was one of those, as a superintendent and as a principal, that kind of believed that maybe some of these business models do have some opportunities. But the fact is, and I think we need to come to a realization and politicians need to understand, we are in a business that we cannot choose our raw material. If I could control from conception to graduation, and please, bring it on…I can't do that. I can't control where they come from. I can't control what they go home to." (administrator)

> "And I think at the end of the day, the mistake that's being made, we are not producing a product. This is not an assembly line. We are nurturing individuals in getting through life and getting through the day. You know what, the people who want to write evaluations and make evaluations and change all these laws, I get it. We all want to do better, but they're beating up the wrong people." (administrator)

Teachers consistently expressed a desire for legislators to visit and spend time in schools and classrooms before engaging in major reforms (five focus groups). In addition, many teachers argued that experts—most importantly, teachers themselves—need to drive reforms in education.

> "They need to get into the schools, get into classrooms. There are people making decisions that have not stepped foot in a school or a classroom and they are so disconnected from the day in, day out dealings of what we as educators, principals are doing in the buildings and they're making these changes based on what?"

> "This is something I feel incredibly strongly about. I get increasingly frustrated with people who have never been formally trained in education making decisions for us....I never think that I could go in and be a physician because I've had a physical so many times in my life. So to me, we are the professionals in this profession, we have been trained, we know what we're doing, we're on the front lines every single day…I want people making these decisions who are familiar with our profession, who are well versed in our profession, who are trained professionally, who are educated in being teachers and principals and I would love for that discussion and those things to happen within our school districts, within our buildings, and to send the information to the legislators and give them options to choose on, but to have them make the decision for us, that we drive what's happening and they check it off."

Appendix C:
List of Educator Forums

February 7, 4 p.m.–6 p.m.
Educational Service District 121
800 Oakesdale Ave. SW
Renton, WA 98057

February 8, 4 p.m.–6 p.m.
Educational Service District 121
Professional Development Center
6501 N. 23rd St.
Tacoma, WA 98406

February 9, 4 p.m.–6 p.m.
Educational Service District 189
Skagit Valley College
2405 E. College Way
Mount Vernon, WA 98273

February 28, 4 p.m.–6 p.m.
Educational Service District 123
3918 W. Court St.
Pasco, WA 99301

March 1, 4 p.m.–6 p.m.
Educational School District 171
430 Old Station Road
Wenatchee, WA 98801

March 2, 4 p.m.–6 p.m.
Educational School District 101
4202 S. Regal
Spokane, WA 99223

March 3, 4 p.m.–6 p.m.
Educational Service District 105
33 S. Second Ave.
Yakima, WA 98902

March 7, 4 p.m.–6 p.m.
Educational Service District 114
105 National Ave. N.
Bremerton, WA 98502

March 8, 4 p.m.–6 p.m.
Educational Service District 113
601 McPhee Road SW
Olympia, WA 98502

March 9, 4 p.m.–6 p.m.
Educational Service District 112
2500 NE 65th Ave.
Vancouver, WA 98661

Appendix D:
Nonpilot Interview Protocol

Thank you for agreeing to participate in a telephone interview about your district's teacher [administrator] evaluation system. The purpose of this interview is to gather information about the educator evaluation system that is currently being implemented in your district. We would also like to know what you consider to be the perceived strengths of the educator evaluation system. The information you provide will be used to help the Washington Office of the Superintendent of Public Instruction understand how to better support districts' implementation of Senate Bill 6696, which established new criteria for evaluating educators.

The discussion today is voluntary and confidential. Neither you nor your school will be identified in any reports.

I would like to tape-record our discussion. I will use the tapes solely for the purpose of ensuring data-collection accuracy. Do I have your permission to record?

Framework/Goals

1. Has your district developed or adopted an instructional framework to guide teachers? If so, please describe the development process.
   a. Was the framework's design influenced by models that have already been established?

      Examples: Teach for America, University of Washington Center for Education leadership–CEL, Charlotte Danielson, Marzano model

   b. What instructional standards or elements has your district developed?
   c. Has your district linked its instructional standards/framework to its teacher evaluation system?

2. What are the major goals of your district's teacher evaluation system?

   Examples: Continuing contract decisions, hiring/firing, PD, compensation, assignment decisions, promotion/advancement, identifying teacher leader

3. What are the goals of your principal evaluation system?

   Examples: Continuing contract decisions, hiring/firing, PD, compensation, assignment decisions, promotion/advancement, identifying teacher leader

Criteria/Evaluation Format

4. What type of criteria do you use to measure performance? For teachers? For principals? *(Specific criteria for evaluating teachers/principals should be embedded in the rubrics being developed in pilot districts). (Probe using question T14 from actual teacher survey.)*

5. Who is evaluated under this new system and set of criteria? *(e.g., general education teachers, core teachers, noncore teachers)*

6. Does the frequency of the evaluations differ for provisional and continuing contract teachers? If so, how does it differ?

7. Is your district's evaluation system based on ratings? If so, please describe the rating system.
   a. How many levels does the rating system include?
   b. Are there specific benchmarks that determine rating levels? For teachers? For principals?
   c. Are measures included that will base a teacher's performance on how well they perform compared to their peers?

      Example: A teacher's value-added score might be used to rank a teacher against other teachers in the school, district, or state.

Measures (Probe using question T16 from actual survey)

8. What measures does your district use to evaluate performance?

   Examples: Teacher value-added scores, observations, portfolios, peer observations, lesson plans, attendance at committee meetings

   a. Do you use student growth measures in the evaluation? If so, for whom?

b. Describe the measure, how it is calculated, and how it is used to measure performance.

9. Do the measures differ for core and noncore teachers? If so, how?

Training

10. Describe the types of training evaluators receive.

11. Do your evaluators receive any training on the following?
    - Ensuring compliance
    - Using instruments reliably
    - Ensuring reliability of scores across raters and evaluators
    - Evaluation of content

12. Does the district offer training on data use to inform student improvement? If so, please describe the training.

Professional Development

13. Are evaluation results used to inform professional development needs? If so, please describe.

Communication

14. What strategies are in place to communicate the evaluation process to new teachers?

15. Are opportunities available for teachers or school administrators to provide feedback about the evaluation system, in order to improve the process? If so, please describe.

General Strengths and Weaknesses

16. What are the biggest strengths of your evaluation system?

17. In what ways could your evaluation system be improved?

Open

18. Is there anything I did not ask concerning your district's current educator evaluation system that you feel is important to share with OSPI?

Thank you very much for your helpful information.

Appendix E

Statewide Data Emerging Themes

Theme	Examples
No uniform purpose for evaluation	Seventy-eight percent of district administrators who responded to the online questionnaire said the primary purpose of their district's teacher evaluation system is compliance. In telephone interviews, principals also cite a range of responses for evaluating principals and other administrators, but with less variation. Principals tend to say the purpose of their administrator evaluation system is to set individual, professional goals that are aligned with ISLLC standards and district goals for student improvement and to address areas of concern. Nevertheless, Washington educators want tools for improvement and growth.
Spotty adoption of instructional framework to assess teachers	Only 24 percent of district administrators surveyed said their district has adopted an instructional framework or set of standards for assessing teachers. Among teachers and administrators, there is a 34 percent agreement rate with this statement. Principals and district administrators agree at a rate of 42 percent that their district has adopted an instructional framework or set of standards for assessing principals. Some educators participating in telephone interviews suggested a process should be put in place to ensure the continuous review and updating of instructional frameworks so that they reflect the most current knowledge base about good teaching and leading. Several educators in telephone interviews reported a five-year or longer gap since the district engaged in any review of its instructional framework for teaching.
Vacuum of appropriate rubrics and relevant measures.	Less than 25 percent of administrators responding to the online questionnaire report the use of rubrics to evaluate teachers. Telephone and focus-group data strongly suggest the need for multiple rating levels. The current practice in many districts of a "satisfactory" or "unsatisfactory" rating finds little support in the field. Administrators and teachers interviewed by telephone reported wanting some flexibility in determining the mix of measures used in district's evaluation systems.
Some educators can go for extended periods without a classroom observation	Classroom- and school-based observations are inconsistent in timing, value, and person conducting the evaluation. Some teachers report frequent observations that result in timely, constructive feedback and ongoing dialogue. Others report frequent walkthroughs that result in dialogue. Still others report sporadic—maybe once or twice a year—observations that yield limited, less useful feedback and infrequent dialogue. Many educators would like to see a system in which all colleagues—teachers and administrators alike—experience classroom (or school) observations annually. This would allow all educators to identify areas of strength and for improvement and consistently promote the message of professional growth to improve student learning.

Need for better, more consistent evaluator training	Educators in telephone interviews and forums expressed a common concern about subjectivity. They said the observations, upon which evaluations are largely based, are often led by individuals poorly trained to conduct them. Ten of the 15 administrator focus groups included discussion of subjectivity in the evaluation process, particularly because the instruments being used were vague and not connected to an instructional framework.
Plan for the union's role	Responses from the field reveal that the decision-making role of the local teachers union varies across the state. OSPI will need to develop a strategy to engage the local teachers unions in the educator evaluation reform process.

APPENDIX F

OSPI Practitioner Workgroups Proposal

Compiled TPEP Model Review Templates

Practitioner Panel Participants

Practitioner Panels System & State Feedback

TPEP Practitioner Panels
Educator & Technical Advisory Panels

Purpose

The TPEP pilot sites have developed evaluation models in the 2010-11 school year that will be piloted across their districts in 2011-12. There are practitioners and technical experts outside of those pilot sites who can provide insight and direction around best practice in teacher/principal evaluation. In the spirit of collaboration and viewing this work as a statewide endeavor, the practitioners will be chosen by the steering committee organizations. The teacher and principal evaluation development will be much stronger and more viable state-side using an outside group of experts representing a wide range of districts. The legislation stipulates consulting with and prominently representing experts in the project including: *"Human resources specialists, professional development experts, and assessment experts"* as well as: *"classroom teachers and principals due to the diversity of teaching assignments and the many developmental levels of students"* (E2SSB 6696, 7((a)).

Outcome:

- Provide expert advice and feedback regarding implementation of the tools, rubrics and models used by the pilot districts for teacher/principal evaluation (including: professional development, data, finance, human resources and legal).
- Provide information and research regarding best practice in educator professional growth to all stakeholders including pilot groups, steering committee, OSPI and the legislature.

Skills/Qualifications

1. Practitioner panel members should be practicing teachers, principals, administrators and experts in the fields of human resources, professional development, assessment and finance.

2. Practitioner workgroups should be appointed based on the following criteria:
   - Rich experience and scholarship around teacher and principal best practice and technical expertise. This includes an ability to see the larger impact of this state-wide implementation.
   - Accomplishments including, but not limited to: National Board Certification, scholarly achievement, or other documentation of school-based achievement and leadership
   - Strong ability to collaborate and discuss from various points of view best practice around teacher/principal evaluation
   - Strong ability to work with adults and facilitate discussions

3. Other panel considerations:
   - Region
   - District Size
   - High –needs schools
   - ELL and other diversity factors

TPEP Model Review Templates

These two templates (teacher/principal) are intended to give reviewers an opportunity to provide feedback on the Washington State TPEP Evaluation Models developed during the 2010-11 school year. Some of the elements are required under **RCW 28A.150.230** and some of the elements are consistent with best practice around implementing new initiatives (communication, stakeholder engagement etc.).

The objectives of this tool for a district are:
- Determine where a district is in the development and implementation plan of an evaluation system according to the required elements in **RCW 28A.450.100**
- Provide feedback and monitoring tool of the implementation process on new initiative best practices.
- Provide a mechanism to communicate the quality of the evaluation system back to the TPEP district in a constructive and supportive manner.

TPEP Model Review Template
Teacher

Anacortes

Required Model Elements (Per RCW 28A.450.100)	Evidence (Use the plan to find evidence that demonstrates development in this area)	Questions (Please ask any questions you have about the model)	Commendations/Suggestions (Please provide any positive points and suggestions for improvement)
Rubrics map back to the revised evaluation criteria	Yes Yes – map from criteria to related dimension is present. However, "evidence" of criteria is not defined. Only definition is on the list of observable behavior. Yes – crosswalk sheet 5D + criteria – sorted samples of behavior	• Define – few, some, most, all/rarely, occasionally, frequently, always • What evidence is there beyond observable behaviors? • Is this observed during walkthroughs or formal observations? • Frequency and intensity of observation • Will there be more specific behavior? • Will you define more specifically what see – is it clear?	What forms of evidence are there beyond observation? Please create key definitions; observable behaviors are not enough to base evaluation on. For example, "substantive intellectual work" is not defined and observable behaviors are unclear or is not defined and observable behaviors are unclear or apply to only a small subset of students. Seems to be helpful for consistency (statewide) to map it out from the 8 criteria. However, would it be more meaningful if the 5Ds are how teachers and principals are talking about practice, to have these as the organizer? • Wow – looks not specific enough. • Time in observation – will there be a check list? • Will you look at all 8 criteria in an observation; should you see all 8 in one?

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Agreement of the local associations representing classroom teachers and principals to collaborate with the district in the development of the evaluation systems.	Community and stakeholder engagement – mentions teachers, principals, variety of organizations (? List please), parent focus group, school board This list needs more examples or more explanation for an example • "Listening campaign" p. 6 • Survey • Description of core team • Stakeholders • Associations?	• Was your EA involved? In what capacity? • Which organizations provided input? • Is there a timeline of the project to see past work and what is planned for involving local associations? (what was/will done, who is involved) • How might you get input from 76% of staff? How will each be measured – should you make it clear for the teacher?	• How did EA participating group provide input? • How was that input used to adjust? Wider participation/feedback Forms of evidence
Participation in the full range of development and implementation activities including: development of rubrics for the evaluation criteria and ratings.	• Development of rubrics – need clarifying quantitative language • Match with evidence or possible forms of evidence Unclear how these were developed – but they exist	→ How were these developed? Did they come directly from the model?	• Cross reference rubrics and types of evidence, i.e. p. 8 state criteria (see collectibles on rubric) • Clarify measures and evidence – link to rubrics • Rubrics look very good • Need to ensure that terms are clear to all
Identification of or development of appropriate multiple measures of student growth (This can include the districts use of measures and evidence)	Needs clarification of how students use data to self-assess - what types of data, and how that relates to instructional strategy p. 8 Pg. 12 – different methods; will be tried and revised during the year	→ Don't understand what is written under "other measures and evidence" (p. 8) • Still developing? • How will you decide? • How will you score student growth?	The forms of evidence to rubrics to help differentiate between levels in order to help calibrate scoring between administrator and teacher, teacher to teacher, and admin to admin Lots of excellent examples here – p. 8 – "student growth data"

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Development of appropriate evaluation system forms	✓ Rubric ✓ Observable behavior for criterion 1-5 ✓ Raw score conv./criterion for 1-4 • Observation/evaluation form • 5 pathways • Rubrics – 5D	• Might it be possible to combine the observation/evaluation form with the complete rubric? (lots of paper, but worth it!) • Do not understand the "5 pathways"	Documents need to be compiled – rubric + observable behavior + raw score hybrid or cross referenced better for ease use • With just a couple of exceptions of terms that need definition ("quality talk", "intellectual work"), the rubrics look very good • Would be great to link descriptors in full rubrics to observable behaviors/artifacts that might show it
Professional development plan for principals and classroom teachers regarding the content of the new evaluation system	Pg. 11 – In development, early communication Still being developed P. 11 – Still in development Still in development plan	P. 11 – Still in development Will your training be around the framework? Do you think you have too… *(did not finish sentence)*	P. 11 – Still in development
Evaluator Training Plan	N/A Could not find evidence of an evaluator training plan P. 11 – Still in development		

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
4- Tiered Summative Evaluation System	• Only determined an overall rating for criterion 1-4 • Overall 1-4 statements are written, but no statement of how to get that rating Some criteria clearly have a 4-step differentiation that is easy to evaluate (1-4). Others use descriptors that are vague and would improve with definition (2.1 for example). P. 9 – Description of possible ways to determine summative score P. 10 – Descriptors, levels 1-4	Standards for a #4 seem very high. "Always" is a very high standard. Do you really mean "always"? Will the district be piloting all of these systems (p. 9)?	Smart idea to pilot a few systems for determining summative scores
Clearly articulated Vision and Goals for District Project	N/A		Plan is not clearly specified in document
Clear and Consistent Communication Plan	N/A Description of communication plan seems to be very comprehensive		DVD is a great idea!
Ongoing Stakeholder Engagement	N/A Yes, involving stakeholders		

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Other Areas	**Plus** • Framework is solid • Clear components under criteria • Qualitative rubric • Good basis for communication • Building a strong foundation **Minus** • Clarify criteria • Add definitions/quantitative explanation • Flush out each criterion • Cross reference documents (rubrics and observable behavior + raw score for conversion) • Needs PD/training plan for future/vision • Have to implement both at the same time • Really work on framework and how to gather evidence re: why good		Last item re: knitting – that is a meeting norm, not evidence of professional learning, under observable behavior

TPEP Model Review Template *Teacher*			
Consortium			
Required Model Elements (Per RCW 28A.450.100)	**Evidence** (Use the plan to find evidence that demonstrates development in this area)	**Questions** (Please ask any questions you have about the model)	**Commendations/Suggestions** (Please provide any positive points and suggestions for improvement)
Rubrics map back to the revised evaluation criteria	• Chose Danielson model – her framework link is included • "Crosswalks" included *some* of Danielson criteria; it appears the Consortium chose 12 top components to focus on – is there intent to include more later?	• Are there overlaps of rubrics in the different criteria that may cause numerous areas to be adversely affected by same? • Why are students not mentioned until the "Distinguished" column?	The works in progress appears incomplete. If you intend to expand, we suggest including all of Danielson's framework for further inclusion.
Agreement of the local associations representing classroom teachers and principals to collaborate with the district in the development of the evaluation systems.	Not clear evidence defining the membership of this consortium ("various")	• Was this a consistent group? • What was the "mix factor"?	
Participation in the full range of development and implementation activities including: development of rubrics for the evaluation criteria and ratings.	• Pre-conversation review of all resources; finally decided on Danielson - "The model is easier to understand than others we reviewed or tried to create" • Rubrics included – are they complete?	• Is rubric too general (not enough specificity)? • What is the opposition to a teacher portfolio?	On your scale, "Basic" does not seem to accurately define that this is a minimal or emerging category; may want to revisit
Identification of or development of appropriate multiple measures of student growth (This can include the districts use of measures and evidence)	• No evidence • All teacher/no student	This rubric document seems to not give enough detail to reliably rate between scales	

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Development of appropriate evaluation system forms	Most forms do not appear to match requirements, or 4 scale summary score	Have your forms been revised to reflect the new direction/thinking?	
Professional development plan for principals and classroom teachers regarding the content of the new evaluation system	Stated they wished to start training in summer 2011, but did not include a plan for that	• How will this roll out? • To whole district at once?	
Evaluator Training Plan	Not addressed other than in the intent to train them with 2 fold training – how to use tool, about the tool	• What? • When? • Who? • How?	
4- Tiered Summative Evaluation System	Summative Evaluation Form reflects "holistic score" – there is no evidence of how the overall score is determined	Annual Evaluation Form does not address 4 tier system – why?	
Clearly articulated Vision and Goals for District Project	No specificity, just a general statement that they have met and discussed	Has the fact that you are numerous districts with different levels of knowledge, philosophies, etc. made this a difficult process, or do you feel the greater input actually benefitted the process?	We are hoping to learn from your experiences of doing this through a consortium, rather than as individual districts. Thank you for taking this on! Looking forward to your pros and cons from working as a consortium.
Clear and Consistent Communication Plan	No specificity, just a general statement that they have met and discussed		

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Ongoing Stakeholder Engagement	No specificity, just a general statement that they have met and discussed		
Other Areas			

TPEP Model Review Template
Teacher

Central Valley

Required Model Elements (Per RCW 28A.450.100)	Evidence (Use the plan to find evidence that demonstrates development in this area)	Questions (Please ask any questions you have about the model)	Commendations/Suggestions (Please provide any positive points and suggestions for improvement)
Rubrics map back to the revised evaluation criteria	Yes, both self-assessment and evaluator rubric align to the criteria		Good examples of evidence for each criteria
Agreement of the local associations representing classroom teachers and principals to collaborate with the district in the development of the evaluation systems.	Yes, saw reporting evidence – newsletters	In our brief review we didn't see evidence of district/teacher survey to collect feedback – vote/consensus from association?	Did not see MOU or LOA
Participation in the full range of development and implementation activities including: development of rubrics for the evaluation criteria and ratings.	Overview letter with number of participation and partnership (how did they choose McRel?)	How much influence did McRel have over the process? Is there local ownership and buy-in?	Did not see a calendar of development activities or who/how many participated at each stage and level of implementation
Identification of or development of appropriate multiple measures of student growth	Although addressed in criteria 6, there doesn't seem to be a tool to address student growth over the school year in all content areas.		

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Development of appropriate evaluation system forms	Yes!	Why is the teacher summary evaluation form different in order of 4 tiers than on the actual evaluator report? This seems confusing.	It seems to us a better practice to have the same form for self-assessment, evaluator use, and final summative input.
Professional development plan for principals and classroom teachers regarding the content of the new evaluation system	PowerPoint mentions training for evaluator/principal but we did not find evidence of professional development activities for teachers or principals		
Evaluator Training Plan	No – missing (see above)		
4- Tiered Summative Evaluation System	Inconsistent with model chosen for principals	We question the equal rating of all criteria - for example, #5 (safe environment), we noticed the raw score/conditions hybrid addresses this.	Raw score model – how is the raw score differentiated based on experience, time in specific position? For example, is it ok to teach in the same position 20 years and score overall a 2?
Clearly articulated Vision and Goals for District Project	• Overview states one vision statement • No evidence of goals		
Clear and Consistent Communication Plan	No evidence of plan, but some evidence of communication		
Ongoing Stakeholder Engagement			

TPEP Model Review Template *Teacher*			
Kennewick			
Required Model Elements (Per RCW 28A.450.100)	**Evidence** (Use the plan to find evidence that demonstrates development in this area)	**Questions** (Please ask any questions you have about the model)	**Commendations/Suggestions** (Please provide any positive points and suggestions for improvement)
Rubrics map back to the revised evaluation criteria	• Used Charlotte Danielson's hybrid of 4 domains – embedded eight criteria • Not clear link	• Would your evaluation be based on four domains rather than state eight? • How will that work in HR if a non-renewal?	Could you organize a guide to quickly see where the eight criteria live on the four domains? (we see that one page has it bolded)
Agreement of the local associations representing classroom teachers and principals to collaborate with the district in the development of the evaluation systems.	• First page intro identifies, in 5th paragraph • Community Stakeholder Engagement section describes steering committee • Principals and teachers with parents to be added	• See Timelines in Communications • See Power Points and mention of working with associations • See Survey and results • See Next steps as inclusive	• How are you planning to gather student input? • See Evidence of Agreement
Participation in the full range of development and implementation activities including: development of rubrics for the evaluation criteria and ratings.	• Evidence of 4 tiered rubrics, criteria 1 • Summative evaluation system around 4 domains	See conditions: • 1 in fostering and management = 1 • Two ratings of 1 result is 1 See cutline is between 1 and 2 Growth model	See clear work
Identification of or development of appropriate multiple measures of student growth (This can include the districts use of measures and evidence)	Under Measures and Evidence many things listed • Observations • Artifacts • Impact on Learning	• How is the evaluator going to differentiate the multiple measures into rubrics? • At the point of implementation will the process be clear for evaluators on how and when to access multiple measures?	Can see evidence of measures being used

Teacher/Principal Evaluation Pilot

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Development of appropriate evaluation system forms	There are many forms	• How will self-assessment and goal setting work on final form? • Are there plans to create a manageable system that is streamlined?	Less is more • Consider reducing number forms – limiting background documentation when implementing will be helpful for clarity • Long form – summary • Short form - summary
Professional development plan for principals and classroom teachers regarding the content of the new evaluation system	• Pilot participants in August – PD training and follow up • Communication indicates may invite to participants	Where are details on who provides training/costs/long term timeline?	Basic plan
Evaluator Training Plan	• See mention of one-day training to familiarize participants • How will evaluators select, describe, and defend scores?	We wonder about: • Calibration • Transitioning principals • Training on how to defend a one/two/three/four How will teachers know alignment and clarity of performance?	• Need seamless connection between rubrics and state criteria • Will staff be trained to clearly understand objective comments that the scoring is based on?
4- Tiered Summative Evaluation System	We see the range of numbers on summative score	How will you calibrate?	
Clearly articulated Vision and Goals for District Project	• Know it is a hybrid of Danielson • The vision and goals are starting to emerge from a complicated and collaborative process	Do you hope to have clarity of vision and goal by August? Or June or July?	It seems that clarity is needed to implement next year to create confidence and communicate in a proactive way
Clear and Consistent Communication Plan	There, but sketchy	We are sure the depth of the work is there…it will be good to see next draft presentation!	It is important to have clarity to establish trust and to avoid conflict

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Ongoing Stakeholder Engagement	Listed but not clear		
Other Areas	How do we differentiate between: • Expectations • Recommendations • Suggestions	We are not asking people to be different at the core, we are asking to become more intentional and to clearly, consistently and objectively become transparent in using common language in how to measure expectations	How will you manage the student growth element?

TPEP Model Review Template
Teacher

North Mason

Required Model Elements (Per RCW 28A.450.100)	Evidence (Use the plan to find evidence that demonstrates development in this area)	Questions (Please ask any questions you have about the model)	Commendations/Suggestions (Please provide any positive points and suggestions for improvement)
Rubrics map back to the revised evaluation criteria	Used the criteria in rubrics	• Latitude in the "3" – 6A • Can you bring clarity to the rubrics by adding examples?	
Agreement of the local associations representing classroom teachers and principals to collaborate with the district in the development of the evaluation systems.			
Participation in the full range of development and implementation activities including: development of rubrics for the evaluation criteria and ratings.			
Identification of or development of appropriate multiple measures of student growth			

Teacher/Principal Evaluation Pilot

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Development of appropriate evaluation system forms		How do you intend to use the final classroom employee performance evaluation?	
Professional development plan for principals and classroom teachers regarding the content of the new evaluation system	Have a plan for PD	How do you attend to inter-rater reliability?	
Evaluator Training Plan			
4- Tiered Summative Evaluation System	Summative statements – need to have more		
Clearly articulated Vision and Goals for District Project			
Clear and Consistent Communication Plan			
Ongoing Stakeholder Engagement			

TPEP Model Review Template
Teacher

North Thurston

Required Model Elements (Per RCW 28A.450.100)	Evidence (Use the plan to find evidence that demonstrates development in this area)	Questions (Please ask any questions you have about the model)	Commendations/Suggestions (Please provide any positive points and suggestions for improvement)
Rubrics map back to the revised evaluation criteria	It's unclear as to how the four domains and their components map on to the revised evaluation criteria	The classroom observation report (template 3) describes the domains and components differently than appears starting on page 41	• It would be good to provide a key as to how the four domains map onto the eight criteria • There are quite a few forms and it is a bit difficult to sort through how the process works
Agreement of the local associations representing classroom teachers and principals to collaborate with the district in the development of the evaluation systems.	Yes, on page 13 there is evidence of collaboration. They started with a small group, expanded the group to 40, then developed a leadership team to communicate with the larger group	• Is there a formal letter of agreement to collaborate? If not, when would this be appropriate? • How is the management of such a large group working?	
Participation in the full range of development and implementation activities	Unsure about the type and level of participation		
Identification of or development of appropriate multiple measures of student growth (This can include the districts use of measures and evidence)	We could not find evidence of this in the materials provided	What is the strategy for addressing this model element?	This is a key area in which additional support and assistance to the pilot districts would be very valuable
Development of appropriate evaluation system forms	Forms include self-assessment, goal setting, planning conference, three templates for classroom observation report, and teacher lesson reflection and reflection on student progress	• How is the workload associated with the completion of these forms going to be managed? • Why are there three templates for classroom observation? Perhaps these are to be used as pilots? What is the rationale?	• To what extent does the completion of these forms facilitate the ongoing conversation about improving practice? • The teacher lesson reflection is very interesting! • What and how is the Student Learning Goal Assessment Results form being used? • How to develop ambitious and realistic goals for student learning and how are these results weighted and used?

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Professional development plan for principals and classroom teachers regarding the content of the new evaluation system	The PD plan begins on June 30, 2011 (pg. 15)	What is the plan for gathering feedback and suggestions from teachers and principals once the PD is underway and models begin implementation?	• The WEA has an online tool for gathering and analyzing feedback that might be useful. • How will other districts acquire the necessary resources to develop and train the teachers and building administrators?
Evaluator Training Plan	There is a plan for a two day training for building administrators (pg. 15)	Will there be any additional opportunities for training beyond the two days?	
4- Tiered Summative Evaluation System	Yes, there are four tiers	How will the qualitative/holistic model be translated into a final rating per domain and final overall rating?	
Clearly articulated Vision and Goals for District Project	Could not find evidence given the materials provided		
Clear and Consistent Communication Plan	On pages 12-14, there are specific descriptions of how district-association leadership, teams and schools have been involved in communication		Numerous efforts have already taken place to communicate the ongoing process
Ongoing Stakeholder Engagement	There is evidence on page 12. The district leadership and union president visited each of the 21 schools. What is the plan for ongoing engagement with stakeholders as the approach is implemented and refined?		

TPEP Model Review Template *Teacher*			
Othello			
Required Model Elements (Per RCW 28A.450.100)	**Evidence** (Use the plan to find evidence that demonstrates development in this area)	**Questions** (Please ask any questions you have about the model)	**Commendations/Suggestions** (Please provide any positive points and suggestions for improvement)
Rubrics map back to the revised evaluation criteria	Rubrics map back to criteria	• Do the elements in the rubric match back to the criteria? • Do the rubrics reflect high expectations and rigor? • Definition for each of the criteria might not align with the rubric for that	• Be sure rubrics have been crosswalked with criteria • Some vagueness in rubrics – how do you define some of the adjectives used in the rubrics in order to measure/determine teachers' effectiveness? • Appendices demonstrate what specifically? • Need to be more clear about what student achievement will be used
Agreement of the local associations representing classroom teachers and principals to collaborate with the district in the development of the evaluation systems.	Yes – staff and teacher feedback to demonstrate; meeting agendas	Know they have had meetings, but unclear who has actually been at the table when decisions have been made	# 7: • Provide documentation of people who have been there • Do the committees for teacher evaluation model and principal evaluation mode work together?
Participation in the full range of development and implementation activities including: development of rubrics for the evaluation criteria and ratings.	Yes – rubrics there; suggested development by entire team	• How well do the rubrics support one another? • Could there be – or is there – a common language between teacher and principal rubric?	Good starting point
Identification of or development of appropriate multiple measures of student growth **(This can include the districts use of measures and evidence)**	Have identified use of data	Have they identified comparative formative and summative assessments for use related, specifically #6?	Should the criteria be weighted?

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Development of appropriate evaluation system forms	Do not exist		• What will self-assessment form look like? • Will each teacher have their own professional development form?
Professional development plan for principals and classroom teachers regarding the content of the new evaluation system	Documentation exists – reflection pre and post		• What kind of support will be provided for both teachers and principals throughout the year? • Question about how they will thoughtfully engage people?
Evaluator Training Plan	Documentation exists		See above
4- Tiered Summative Evaluation System	Yes		• Inconsistent message around "rounding up" • How do the factors identified affect their evaluation? • Need to be clear that mathematical model includes rounding up/down • Inter-rater reliability – what does training look like, if it exists?
Clearly articulated Vision and Goals for District Project	Has an articulated vision and goals		Who decides about the length of the stay at a 2 looks like?
Clear and Consistent Communication Plan			• How will they build ownership? Not clear • How will they brand their message that this is about professional growth and practice?
Ongoing Stakeholder Engagement			Not much community engagement demonstrated
Other Areas			What is about performance and what is about discipline? Only principals and ed. associations being trained, not teachers.

TPEP Model Review Template
Teacher

Snohomish

Required Model Elements (Per RCW 28A.450.100)	Evidence (Use the plan to find evidence that demonstrates development in this area)	Questions (Please ask any questions you have about the model)	Commendations/Suggestions (Please provide any positive points and suggestions for improvement)
Rubrics map back to the revised evaluation criteria	Yes. Explicitly structured around criteria as "headers."	• To what extent has the Danielson model been modified such that it is her model in name only? • Is this Danielson crosswalk and her language in the rubrics? • Do the heading titles correspond, e.g. "emerging" = "basic"?	Criteria 2 does not seem to address key elements of lesson design: "Teach to an objective"; "gradual release"/practice guided, independent
Agreement of the local associations representing classroom teachers and principals to collaborate with the district in the development of the evaluation systems	2 union reps listed on steering committee? + 2 teachers		Size of pilot team seems very small compared to size of district
Participation in full range of development and implementation activities including: development of rubrics for the evaluation criteria and ratings	We don't know from material in notebook; references to teacher cadre and to presentation to various groups		
Identification of or development of appropriate multiple measures of student growth (This can include the districts use of measures and evidence)	+ Observation rubric + Self-assessment - "Aware" of need to develop rubrics		• Explicit statement that student growth data – "evidence that" – it "perverts" both sides of that process – what is that evidence? • Links standardized tests as only means of student growth data; why not include other "robust" measurements of student growth? • Seems to equate student growth data solely with standardized test data • Broad, dismissive statements that seem to reject out of hand any use of common assessment data

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Development of appropriate evaluation system forms	? We don't see the forms		
Professional development plan for principals and classroom teachers regarding the content of the new evaluation system	• District admin retreat • Full teacher workshop		
Evaluator Training Plan	• Documents don't reflect identification of evaluators or the training they will receive • No evidence that any data tracking process has been designed	Seem to be relying on the idea that they have not "really" changed their established system much	
4- Tiered Summative Evaluation System	Evidence of 4-tiered rubric, but no summative form that reflects the actual rating(s) can be found		• How does the language in the summative rubric relate to the language in summative evaluation? Why not use consistent language? • "Effectiveness of Practice" is in level 1, 3 not 4,2 • How can you use the terms effective/ineffective without referencing student achievement data?
Clearly articulated Vision and Goals for District Project			
Clear and Consistent Communication Plan	Document references a presentation/plan, but we don't see it		On page 10, professional practice at level 4, "student directed" notion of objectives – how does a K teacher, for example, demonstrate that students are leading learning, and is the even appropriate?

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Ongoing Stakeholder Engagement	No evidence provided		
Other Areas			

TPEP Model Review Template
Teacher

Wenatchee

Required Model Elements (Per RCW 28A.450.100)	Evidence (Use the plan to find evidence that demonstrates development in this area)	Questions (Please ask any questions you have about the model)	Commendations/Suggestions (Please provide any positive points and suggestions for improvement)
Rubrics map back to the revised evaluation criteria	✓	Why are teacher and principal documents in different formats?	• Really liked clarity of principals with evidence column • Some criteria had great sub-criteria – could use more of that around use of student data
Agreement of the local associations representing classroom teachers and principals to collaborate with the district in the development of the evaluation systems.	Good		
Participation in the full range of development and implementation activities including: development of rubrics for the evaluation criteria and ratings.	Good		

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Identification of or development of appropriate multiple measures of student growth (This can include the districts use of measures and evidence)	Not clear. Seems buried, especially compared to principal document.		Something along the lines of principal document would be great, especially Criteria 7 & 8 – could benefit from examples
Development of appropriate evaluation system forms	Not present		
Professional development plan for principals and classroom teachers regarding the content of the new evaluation system	Under-developed; not yet completed		Could use examples and clarity of terms. What does "exemplary" mean? How will it be qualified?
Evaluator Training Plan	Under-developed; not yet completed		
4- Tiered Summative Evaluation System	Present		Don't understand "looser" and "stricter" – not immediately clear what the process will be
Clearly articulated Vision and Goals for District Project	Great		
Clear and Consistent Communication Plan	Great		
Ongoing Stakeholder Engagement	Great		

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Other Areas			• Well-documented, good mix of participants. Impressive. • Where is teacher self-assessment? • Criteria 1 could use sub-criteria that addresses quality of questions such as ability to promote critical thinking and to help students ask great questions of themselves and each other
Supplemental Notes	1 • Use sub-criteria? • ID good questioning techniques • Teach kids how to ask good questions 2.1 • Sub-criteria makes it much easier to qualify rubric • Pretty solid 4 • Qualify "skillful" vs. "exemplary" • How to gather evidence of multiple preps? 7.2 • Why no sub-criteria? • How would evaluator distinguish levels of communication? Overall – Expand sub-criteria, provide examples		

TPEP Model Review Template *Principal*			
Central Valley			
Required Model Elements (Per RCW 28A.450.100)	**Evidence** (Use the plan to find evidence that demonstrates development in this area)	**Questions** (Please ask any questions you have about the model)	**Commendations/Suggestions** (Please provide any positive points and suggestions for improvement)
Rubrics map back to the revised evaluation criteria	• Self-assessment rubric • Evaluator rubric	Is all evidence required?	Order of evaluation rating form should match same order of rubric form
Agreement of the local associations representing classroom teachers and principals to collaborate with the district in the development of the evaluation systems.	Newsletters imply working together on project	Is there agreement in the contracts? MOU?	
Participation in the full range of development and implementation activities including: development of rubrics for the evaluation criteria and ratings.	• Raw score/conditions • Hybrid model	How were the rubrics decided?	
Identification of or development of appropriate multiple measures of student growth **(This can include the districts use of measures and evidence)**	• Implied in newsletter • Growth data referenced	How is site growth measured?	
Development of appropriate evaluation system forms	• Self-assessment • Evaluation • Summary		• Rubric and evaluation forms should look more similar • No summary form for all criteria
Professional development plan for principals and classroom teachers regarding the content of the new evaluation system	Not evidenced, not available		

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Evaluator Training Plan	Not available – training is planned	How will consistency between different evaluators occur?	
4- Tiered Summative Evaluation System	2 examples given	• Student growth – how is this tracked? • How is evaluation scored? Summative? Different weights for different criteria? • How does evidence tie to rating? • How does collaboration on rating form work if disagreed?	• Weighted criteria – prioritize what evidence is most important in rating • Examples/expectations of evidence (i.e. all evidence must be brought to pre-conference?) • Who conducts survey?
Clearly articulated Vision and Goals for District Project	Cover page		
Clear and Consistent Communication Plan	Newsletters		
Ongoing Stakeholder Engagement	• Parents • Website • PAC • Focus groups		

TPEP Model Review Template
Principal

North Mason

Required Model Elements (Per RCW 28A.450.100)	Evidence (Use the plan to find evidence that demonstrates development in this area)	Questions (Please ask any questions you have about the model)	Commendations/Suggestions (Please provide any positive points and suggestions for improvement)
Rubrics map back to the revised evaluation criteria	Each criteria is throughout	What document would you use for assistant principals and other district administrators?	Like the references to the 21 Marzano characteristics throughout the document
Agreement of the local associations representing classroom teachers and principals to collaborate with the district in the development of the evaluation systems.			
Participation in the full range of development and implementation activities including: development of rubrics for the evaluation criteria and ratings.			
Identification of or development of appropriate multiple measures of student growth (This can include the districts use of measures and evidence)	Menu of multiple measures		
Development of appropriate evaluation system forms	• Rubric but no self-assessment/final evaluation forms • Evidence forms		
Professional development plan for principals and classroom teachers regarding the content of the new evaluation system		Where does the principal's evaluation fit into the PD plan?	

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Evaluator Training Plan			
4- Tiered Summative Evaluation System			
Clearly articulated Vision and Goals for District Project			
Clear and Consistent Communication Plan			
Ongoing Stakeholder Engagement			

TPEP Model Review Template
Principal

Kennewick

Required Model Elements (Per RCW 28A.450.100)	Evidence (Use the plan to find evidence that demonstrates development in this area)	Questions (Please ask any questions you have about the model)	Commendations/Suggestions (Please provide any positive points and suggestions for improvement)
Rubrics map back to the revised evaluation criteria	Yes		Really liked clarity of document
Agreement of the local associations representing classroom teachers and principals to collaborate with the district in the development of the evaluation systems.	Yes		
Participation in the full range of development and implementation activities including: development of rubrics for the evaluation criteria and ratings.	Good		

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Identification of or development of appropriate multiple measures of student growth (This can include the districts use of measures and evidence)	Excellent		
Development of appropriate evaluation system forms	Excellent		
Professional development plan for principals and classroom teachers regarding the content of the new evaluation system	Under development		
Evaluator Training Plan	Under development		
4- Tiered Summative Evaluation System	Present		
Clearly articulated Vision and Goals for District Project	Excellent		Overall – impressive, clear, well-articulated
Clear and Consistent Communication Plan	Excellent		

New Initiative Best Practices	Evidence	Questions	Commendations/Suggestions
Ongoing Stakeholder Engagement	✓		
Other Areas			• How much input does staff have in principal evaluation? • Parents? • Students?

Participants: TPEP Practitioner Review Panel

May 31, 2011

Name		District/Org
Anderson	Dana	ESD 113
Anderson	Eric	Seattle
Anderson	Sue	Tumwater
Armstrong	Mack	WASA
Ashabraner	Karin	Peninsula
Banks	Patti	University Place
Brown	Ron	Wenatchee
Burke	Alan	OSPI
Campbell	Tammy	Spokane
Conditt	Shari	Woodland
Davies	Lois	Pateros
De Paoli	Barry	Chelan
Hall	Taryn	Nine Mile Falls
Hall	Pete	Spokane
Harmon	Jeanne	CSTP
Hartmann	Judy	Governor's Office
Hatch	Shellie	Prosser
Hattendorf	Ramona	PTA
Headrick	Kim	Cle Elum
Imler	Becky	Wapato
Karl-Robinson	Kelci	OSPI
Kernutt	Doug	Clover Park
Kipp	Gary	AWSP
Klewiada	Nora	Bellingham
Laughery	Patty	Moses Lake
Legary	Lisa	Clover Park
Lewis	Shawn	OSPI
Mahaney	Tina	Bremerton
Manchester	Starla	Vancouver
Miedema	Allen	Northshore
Parr	Randy	WEA
Plano	Gary	Mercer Island
Plecki	Marge	UW
Poirer	Scott	WEA
Priddy	Jennifer	Olympia
Quinn	Paula	AWSP
Randall	Ann	WEA
Rose	Leslie	OSPI
Rosier	Paul	WASA
Skerritt	Nancy	Tahoma
Werner	Kathleen	Aberdeen
Williams	Bill	PTA
Wisenburg	Suzanne	Bremerton
Young	Lucinda	WEA

Data/Tech

Driving Forces	Restraining Forces

• Data helps everyone get on the same page
• Data can reveal hidden issues
• What comes off the plate? (data collection, other PD)
• Data can significantly improve the system
• How do you collect and manage all of the data?
• How to effectively use data to make informed decisions
• Move staff to use technology more effectively
• How to analyze data
  o Time
  o Process
  o Use
• Highly variable methods and differences in the type of data available
• Have an opportunity to develop multiple measures tailored to specific circumstances
• Different weights and values of data
  o What data has the priority?
	• Safeguards around protecting data
	  o Mechanics, how much to expose
	  o How to make it useful
	• Hardware resources?
	  o Cost, implementation, sustainability?
	• Don't know what data to look at
	• Limits on all resources
	  o Time, people, money
	• These resources are all stretched – this is one more unfunded demand on those
	• What takes priority among the multitude of state data requests?
	• What is our data capacity?
	• How to use the data effectively?
	• Don't want to look at data
	• Confidentiality
	• Students
	• Teachers and principals
	• How do I link the data elements together?
	• Distributing and making sense of data
	• Is the data reliable and valid for the purpose?

Other Forces: Align evaluations with student achievement	

• Student data and alignment with evaluation tool
• Inter-rater reliability
• Can clear path, keep focused
	• Student data
	  o Need clear strategies
	  ▪ Ability to use
	  ▪ What will be used?
	  ▪ Not just summative to formative, but make-up of students (mobility, ELL, etc.)
	• IRR – within district and within state

Other Forces: Principals & Teachers in Need	

• Public account – as different from legal requirements
• Principal training groups
	• Huge operational change
	• Levels of principals
	  o Do they understand process?
	  o Are they "instructional experts"?

Finance

Driving Forces	Restraining Forces

• Inadequacy of current system
• Increasing emphasis on performance
• What is happening around rest of country
	• Current state fiscal situation
	• Time
	  o Time to develop and train
	  o Time to implement
	  o Fiscal issue
	• Thin D.O. staff
	• Increasing workloads for all
	• Span of control
	  o Principal workload (current)
	• Professionals not always wanting to embrace change – see the same being forced upon them
	• Timeline – State's
	  o Not realistic if meaningful and sustainable models are to be implemented
	• Money to implement PD for all staff and building leaders
	• Continuity of current decision makers who may not be around when full state implementation is to occur (2013-14)

Other Forces: External & Internal	
• Current perceptions of the profession • Desire of professionals to maintain credibility • Federal government	

HR

Driving Forces	Restraining Forces

• Change needed – everyone wants and needs good feedback	• Logistics of implementation
	  o Time to train, review
	  o Conduct baseline evaluation
	  o Collision in change of requirements
	  o Deciding on evaluation tool and bargaining its implementation

Other Forces: Student Data	

• Alignment with evaluation tool	• Need clear strategies
	  o Ability to use
	  o What will be used?
	  o Not just summative to formative but make-up of students (mobility, ELL, etc.)

Other Forces: Ownership & Inter-rater Reliability	
• Ownership – better for all • Inter-rater reliability	• Ownership - takes a lot of work to get there • Within district and within state

PD

How differentiated by organizational capacity & culture?	How supported at various levels?

• Differentiated by role
  o Within role by capacity
• Create common elements (power standards)
  o Calibration opportunities
  o Anchor points
	• Classroom teachers
	• Principals
	• System
	  o Learn
	  o Observe
	  o Practice
	  o Feedback
	  o Reflect
	  o Evaluate
	• Drawing up leadership capacity
	• Statewide resources
	  o Common elements
	  o Change process
	  ▪ Awareness
	  ▪ Informational
	  ▪ Personal
	  ▪ Management
	  ▪ Consequence
	  ▪ Collaboration
	  ▪ Refocusing

Readiness

Why change? **What is change?**	

Expert

on

We need:

Focus - Consistency - Coordination - Resources - Support

Principal

Driving Forces	Restraining Forces
• Current limitations of present evaluation system • Excitement around teaching and learning • New system about growth • Movement toward becoming a learning organization • Clarity of high quality teacher behavior • Same page with teachers and principals	• Time to complete process for all teachers • High needs teachers probation • Weight of all pieces • Funding for professional development • Are there elements of effective teaching not represented? • Evaluating specialists
Other Forces: Changing Traditions	
• New teachers have this philosophy – supports it, keeps them from falling • Framework – clarity of effective instruction • Teachers want to change	• Risk of stepping into the unknown • Time and effort • All stakeholders involved • Institutional policies
Other Forces: Student Data	
• Link outcomes with student data • Use student data – more intentionality to drive • Become the culture	o Agreement on pieces of data to use o How do we take into account: ▪ Demographics ▪ Multiple measures ▪ Growth models ▪ "This rough class" ▪ Teacher assignments (PE/art) ▪ Tools to measure each area management

Superintendent

Driving Forces	Restraining Forces
• Non-negotiables = Criteria/Four-tiered system • Teacher quality/Promoting student achievement/accountability • Clear and consistent research on what has worked ➔ defining effective practice • Acknowledgement that student achievement is core work • Universal acceptance of need to change • Momentum to push change in culture • Required to have public awareness of evaluational achievement • Alignment of teacher/principal/staff evaluations • Alignment of instructional framework • Low resources have moved work forward	• Fatigue factor – high demand/low support at state and federal level • Time/Money/Energy/Resources • The work is harder/more time consuming/more transparent for administrators
Other Forces: Values, Traditions ***Tough times (more difficult/more clear)***	
• Relationships – collaboration becoming norm • Belief the work is about kids'success • Work focused on right issues	• Strained relationships over budgets • State mandates between admin and unions • Other constraints and demands pull away from joint conversations
Other Forces: Legal, Polices, Contracts ***Transitions between (prior/next)***	

• Majority of staff are reasonable and hard working
  o Focused on kids
  o Want authentic growth conversations in the evaluation process
	• Existing contractual agreements
	• Need for clarity
	• Union reps don't always reflect whole group, yet take time and bring multiple issues
	• Principals/Admin may not have skillset or a depth of knowledge of instruction
	• Fear of change

Teacher

Driving Forces	Restraining Forces
• Quality instructive feedback (timely) • Clear common goal, language and expectations • Identify meaningful PD • Trusting relationships • Holistic examination of one's teaching • Legislation-mandated • Students deserve proficient teachers • Ease and value from experience	• Time and money (repeat) • Skill/knowledge base of evaluator/evaluate • Objectivity on both sides • Language ambiguity • Lack of trust • Inter-rater reliability • Buy-in by all parties • Right to appeal • Lack of authentic support • Data collection reporting and confidentiality • How much evidence is enough?
Other:	
Can we please get a demonstration of the teacher evaluation data service that you mentioned?	

APPENDIX G

OSPI School Employee Evaluation Survey PDF Summary

OSPI School Employee Evaluation Survey Data File (link to Excel spreadsheet)

http://www.k12.wa.us/Communications/StimulusPackage/SEES_Summary_Data_SY2009-10.xlsx

Kennewick School District

Washington State School District

Teacher and Principal Evaluation System Survey Results

2009-2010

Data as of 4/6/2011

	Certificated Classroom Teachers	Principals
1. Results of performance evaluations for certificated classroom teachers and principals: Count of satisfactory and unsatisfactory teachers and principals:	Count of Teachers and Principals	
Satisfactory	59,022	2,578
Unsatisfactory	459	41
2. Description of Rating System: Did you use a rating system that consisted of only two ratings (Satisfactory or Unsatisfactory)?	Count of Districts	
Yes	209	196
No	80	93
3. System Capabilities: Do you have a system that is capable of collecting counts at the criteria level?		
Yes	81	68
No	208	221
If "Yes", is the capability currently being used?	60	50
Is the system is electronic?	21	14
4. Evaluation Criteria: Did you evaluate based on the criteria set forth in the RCW?		
Yes	282	277
No	7	12
5. Question 5 requested specific attachments. There is nothing to report for this question		

Washington State School District
Teacher and Principal Evaluation System Survey Results
2009-2010
Data as of 4/6/2011

	Certificated Classroom Teachers	Principals
6 Frequency of observations: What is the district's minimum required observations for each of the following?		
Provisional Teachers		
Once a Year	2	
Twice a Year	179	
Three Times a Year	85	
Other	23	
Continuing Teachers		
Once a Year	33	
Twice a Year	237	
Three Times a Year	8	
Other	11	
PGO Teachers		
Once a Year	89	
Twice a Year	94	
Three Times a Year	15	
Other	91	
Principals		
Once a Year		107
Twice a Year		68
Three Times a Year		13
Other		101
7 Methodology: Select methodologies used in evaluation process (multiple items may be selected):		
Principal observation	274	
Superintendent observation	53	219
District office personnel observation	37	76
Teacher evaluation		27
Self evaluation	77	78
Student evaluation	9	4
Peer evaluation	6	3
Other	11	56

Washington State School District
Teacher and Principal Evaluation System Survey Results
2009-2010
Data as of 4/6/2011

	Certificated Classroom Teachers	Principals
8 Participants: Select the expected participants for evaluation process (multiple items may be selected):		
School Board	13	62
Superintendent	75	236
Principal	268	95
Teachers	119	27
Aides/Paraprofessionals	7	12
Student	4	3
Parents/Guardians	2	3
Other	20	51
9 Feedback Protocol: Select the feedback protocols used in evaluation process (multiple items may be selected):		
In-person review	282	266
Written documentation delivered to participant	277	257
Team collaboration and presentation	11	17
Other	8	17
10 Results: Indicate whether the results of the evaluation system are used in following processes (multiple items may be selected):		
Professional Development	257	241
Compensation	-	27
Promotion	24	52
Establishment of probationary period	244	171
Probable cause for non-renewal of contract	268	234
Instructional improvement	278	135
Leadership improvement	111	255
Other	7	17
11 Achievement: Are student achievement outcomes used in evaluations?		
Yes	33	63
If "Yes": Student achievement outcome included in collective bargaining agreement	5	5
No	256	226

Washington State School District
Teacher and Principal Evaluation System Survey Results
2009-2010
Data as of 4/6/2011

	Certificated Classroom Teachers	Principals
12 AYP: Is Adequate Yearly Progress used in principal retention?		
Yes		41
No		248

APPENDIX H

TPEP Pilot Site Interview Compilation Video

http://tpep.files.wordpress.com/2011/05/tpep-site-interviews.m4v

Office of Superintendent of Public Instruction
Old Capitol Building
P.O. Box 47200
Olympia, WA 98504-7200

For more information about the contents
of this document, please contact:
Michaela Miller, OSPI
E-mail: Michaela.miller@k12.wa.us
Phone: (360) 725-6116

To order more copies of this document,
please call 1-888-59-LEARN (I-888-595-3276)
or visit our Web site at http://www.k12.wa.us/publications

Please refer to the document number below for quicker service:
11-0027

This document is available online at:
http://www.k12.wa.us/EdLeg/TPEP/default.aspx
http://tpep-wa.org/

This material is available in alternative format upon request.
Contact the Resource Center at (888) 595-3276, TTY (360) 664-3631.

Office of Superintendent of Public Instruction
Old Capitol Building
P.O. Box 47200
Olympia, WA 98504-7200
2011