The Wyoming State Model Educator Support and Evaluation System

The Wyoming Advisory Committee to Select Education Accountability Committee

With Support from the Center for Assessment

DRAFT: November 27, 2013

Table of Contents

Introduction 3
Key Principles 3
General Evaluation Framework of the Wyoming Educator Evaluation State Model 4
Professional practice measures 5
Measures of student performance 5
Specific Measurement Framework for the Wyoming State Model 7
Standards for Professional Practice 7
Domain 1: Learner and Learning 7
Domains 2 (Content Knowledge) 8
Domain 3 (Instructional Practice) 9
Domain 4: Professional Responsibility 10
Data Collection Strategies 10
Documentation of Practice: Self Assessment, Goal Setting, and Artifact Collection 10
Observations of Professional Practice 11
Student Performance 12
Student Learning Objectives (SLO) 12
Calculating Student Performance Results in "Tested" Subjects and Grades 13
Shared Attribution 14
Student surveys 15
Performance Standards 17
Highly Effective 18
Effective 18
Needs Improvement 19
Ineffective 19
Combining Multiple Measures 20
Supports and Consequences 22
Assumptions 22
Supports 22
Consequences 23
Recommendations for Implementation 24
Review Process 26

Introduction

The Advisory Committee to the Wyoming Select Committee on Educational Accountability was charged with carrying out the recommendations put forth in the Wyoming Accountability in Education Act of 2012 (WEA 65) and House Bill 0072 (2013 Chapter 167). The specific charge for the Advisory Committee was to design a State Model for educator evaluation in Wyoming. The Select Committee was quite clear that they wanted a balance between state and local control and, in keeping with Wyoming's educational philosophy, the Select Committee placed considerable authority for making specific design and implementation decisions with local educational leaders and teachers. However, in order to best support the work of districts, the Advisory Committee produced this document: The Wyoming State Model for Educator Support and Evaluation System. This model system outlines methods and design decisions necessary for implementing an educator evaluation system and indicates where the Advisory Committee recommends where the requirements should be "tight" or more standardized across districts and where flexibility is expected and even encouraged. The Advisory Committee intends for the Model System described below to be able to be used by districts as the basis for their local systems if they choose. The Model System will not be "plug and play" in that local districts will still have many decisions in order to operationalize their local system, but this Model System is designed to make districts' jobs considerably easier.

A critical aspect of the model system, as reflected below in the key principles, is the intention to build both an internally coherent system and an educator evaluation system that is coherent with other educational accountability systems in Wyoming. A coherent system would use information from the school accountability system (and perhaps the district accreditation system) to supplement the information generated from the educator accountability systems. For example, if a school has demonstrated high achievement and the students are growing at admirable rates, there is good evidence of high quality education in the school. This suggests that the State (likely WDE) can trust that the school is comprised of high-performing educators. Relying on the larger sample sizes associated with the school than with any individual teacher means that the determinations are that much more reliable. This intent to build off of the information from the school accountability does not relieve school districts from implementing educator evaluation systems, but it could mean that the state would have to provide far less oversight of educator evaluation systems in high performing schools.

Key Principles

The following principles guided the development of the Wyoming State Model Support and Evaluation System. The Advisory Committee kept these principles at the center of its deliberations in the development of the various components of the system and these principles are at the heart of the recommendations discussed throughout this document. As noted below,

the primary purpose of the system is to maximize and improve student learning. The following principles support this primary purpose.

1. The primary purpose of Wyoming's educator evaluation system to support and promote increases in student learning in Wyoming schools such that all Wyoming students graduate ready for college or careers. The Advisory Committee believes that while related, career and college readiness require somewhat different knowledge, skills, and dispositions, but in both cases, beyond simply strong performance in language arts and mathematics. This is not to downplay mathematics and language arts knowledge and skills. These represent necessary and not sufficient dimensions of college and career readiness.
2. The system must be designed coherently to support a system of continuous school improvement. A coherent system will work with the school and leader accountability systems and foster collaboration among educators, administrators, and other stakeholders, including civic leaders, business representatives, and parents.
3. The State Model and locally-aligned versions of the system shall be designed to promote opportunities for meaningful professional growth of educators by providing specific and timely feedback on multiple aspects of professional practice and student learning.
4. The system must be designed and implemented with integrity. A system designed with integrity will be transparent such that all relevant participants clearly understand the expectations.
5. The State Model must allow for flexibility to best fit local contexts and needs. The local evaluation systems should be design collaboratively by administrators and educators, with input gathered from parents and community members.
6. The system will provide credible information to support hiring, placement, and career ladder decisions in a defensible manner.
7. The system must be supported by local and state policy makers to ensure that leaders and teachers have the proper opportunities and resources to successfully implement the system.

General Evaluation Framework of the Wyoming Educator Evaluation State Model

The general evaluation framework of the Wyoming State Model describes the overall approach for how local districts following the State Model would approach the data collection involved in evaluating educators. The State Model follows from the key principles outlined at the beginning of this document. There are four dimensions of educator practice along with evaluations based on student achievement. As part of the general measurement model, the State Model includes the use of multiple measures of each domain when possible as long as the multiple measures

improve the validity of the evaluation decision. All local educator evaluation systems shall include the elements discussed below.

Professional practice measures

A key aspect of the State Model is that it will contain five major dimensions, four domains of professional practice and one domain of student performance data. The four domains of professional practice noted below represent the overarching categories of the Interstate Teacher Assessment and Support Consortium Model Core Teaching Standards (InTASC Standards)[1].

- Learner and Learning
- Content Knowledge
- Instructional Practice
- Professional Responsibility

Districts will use a variety of tools to measure professional practice (e.g., Danielson's *State Model for Effective Teaching*; Marzano's *Art and Science of Teaching*). The Advisory Committee does not want to limit districts' options, but recommends that all locally-selected tools measure the four domains of effective teaching described in the InTASC Standards and capture the essence of all 10 standards. District leaders will be expected to document the degree to which its selected tool is able to meet this requirement.

Multiple approaches and measures will be used to collect data on educator practices such that the specific data collection approaches are tailored to the complex nature of teaching practice. The Advisory Committee recommends having each educator complete a self-assessment each year that will be used as the foundation of a goal setting meeting with the principal and/or peer coach (mentor). The self assessment and collaboratively established goals will be used to focus the professional practice data collection for the year in which the educator is being formally evaluated. For the years in which the educator is not undergoing a formal evaluation, the self assessment and goals shall be used to guide professional development and formative evaluation activities.

Measures of student performance

The Wyoming State Model uses three strategies for incorporating student achievement and growth into evaluations in order to attempt to maximize the benefits, while striving to minimize potential unintended negative consequences.

- ✓ Student Learning Objectives

[1] Council of Chief State School Officers. (2011, April). Interstate Teacher Assessment and Support Consortium (InTASC) Model Core Teaching Standards: A Resource for State Dialogue. Washington, DC: Author. http://www.ccsso.org/Resources/Programs/Interstate_Teacher_Assessment_Consortium_(InTASC).html

- ✓ Student Growth Percentiles (if applicable)
- ✓ The SLO and/or SGP results may be "shared" among multiple educators depending upon local theories of action around school improvement.

Student Learning Objectives (SLO) form the foundation of Wyoming's approach for documenting changes in student performance associated with a teacher or group of educators, therefore, all educators will have the results of SLOs incorporated into their evaluations. For educators in "tested" subjects and grades, those grades and subjects for which there is a state, standardized test as well as a state test in the same subject in the previous year, student performance will be evaluated using Student Growth Percentiles (SGP), and the results of SGP analyses, along with SLO results, will be used in the evaluations of educators in tested subjects and grades. Both SGP and SLO approaches are described in more detail in the Specific Measurement Model section of this document (below).

Both SGP and SLO approaches can be used to attribute the academic achievement and growth of students to individual educators or to appropriate aggregations of educators such as grade- or content-level teams or even the whole school. Distributing student performance results to multiple educators is referred to as "shared attribution." The Advisory Committee further recommends that at least part of the SLO and/or SGP results be shared among multiple educators depending upon local theories of action around school improvement.

The State Model is designed to promote coherence and integration among the five components. Therefore, the Advisory Committee recommends weighting each component, especially student learning, as equally as possible in the overall evaluation of each educator. Because there are often differences between nominal (intended) and effective (actual) weights, the Advisory Committee recommends that as each district pilots its system, it analyzes the data to determine the actual weight of the various components. This actual weighting will depend on the variability in the responses to the specific instruments used in each district. In the following sections, the major dimensions of the State Model are discussed in more detail.

The Advisory Committee recommends requiring that districts provide evidence that any tool used for evaluating teacher practices validly measures all four domains of teacher practices. The Advisory Committee recommends allowing districts to alter the weighting of the various student practice domains as long as student learning counts at least 20% and that all components of the system are fully evaluated for each teacher in each three year period. The Advisory Committee further recommends that teachers in each district have input into the weighting decisions of each district's system.

Specific Measurement Framework for the Wyoming State Model

The specific measurement framework adds the details to the general measurement framework discussed above to guide the data collection methods in order to successfully conduct educator evaluations. Such a detailed measurement model describes the type and frequency of data collection approaches for each of the major components of the model. The specific measurement approach presented here is still not detailed enough for districts to adopt in a "plug and play" fashion, but it is intended to provide recommendations for how data should be collected to support educator evaluation in Wyoming. The following section includes a brief review of the relevant InTASC standards, organized by major domain, and then provides recommendations for how the performance of educators related to each domain may be evaluated. Additional guidance by WDE, this Advisory Committee, or others will help fully describe the specific measurement procedures and policies to be enacted for the various educators in the system.

Standards for Professional Practice

The State Model uses InTASC Standards as the framework for evaluating teachers relative to the four domains of effective teaching. This recommendation is based, in part, on ensuring that the State Model is not tied to any commercial products, but to open source materials widely used by multiple states and districts. Local districts may adopt tools or approaches to add more specificity to the InTASC Standards, but the Advisory Committee recommends having districts document that any tools used in their local model are supported by research or at least best practice.

The following section presents the verbatim language of each of the ten standards, grouped into the four domains of professional practice, and recommendations for data collection approaches to measure educator performance relative to these standards and domains.

Domain 1: Learner and Learning

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

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

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

Well structured and multiple classroom observations may be used to collect data for evaluating educators in relationship to standards 2 and 3. However, such observations would be unlikely to reveal enough information about teachers' understanding of learner development (standard 1) to enable evaluators to make valid judgments. For example, planning documents that describe how the educator includes an understanding of learning theory and individual differences would be a source of information for judging educators. Similarly, evidence of reading and understanding relevant literature could provide documentation for educators' consideration of learner development as part of the teaching process. Of course, possessing this knowledge is only a first and insufficient step. Educators must be able to apply such theoretical and/or empirical reading to actual classroom practice. Some of this understanding could be revealed through reflection and planning documents, but also through pre- and post-observation conferences. Given the variety of information necessary to support decisions related to this domain, the **Advisory Committee recommends** that local evaluation systems include sources of evidence, similar to the examples described here, in the evaluation of educators' according to Domain 1.

Domains 2 (Content Knowledge)

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

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

This domain requires a teacher to demonstrate deep knowledge of disciplinary content and how to connect that content knowledge with appropriate instructional strategies. This is referred to as *pedagogical content knowledge*. Similar to Domain 1, it is unlikely that evaluators could collect information about content and pedagogical content knowledge simply through observations of practice. Content knowledge (standard 4) must be evaluated through collection of artifacts such as successful completion of programs of study and/or in-depth discussions with experts in the relevant content area. Once the evaluator has documented that the educator possesses solid content knowledge, the educator should include, as part of her/his self-reflection and goal setting, plans to stay current and improve her/his understanding of the discipline. The educator should be expected to document and reflect on her/his new understandings of the discipline as part the artifact collections.

Pedagogical content knowledge or the application of content to instructional practice (standard 5) should also be evaluated by examining planning and refection documents. However, evaluators

may gather critical information related to standard 5 through structured observations of practice that include pre- and post-observation conference to allow for reflections of this standard.

Domain 3 (Instructional Practice)

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

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

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

Information about the way in which an educator plans for instruction (standard 7) and uses assessment (standard 6) may be revealed through pre- and post-observation conferences, particularly planning for instruction, but examining artifacts such as unit plans, syllabi, and assessment tools would reveal important information about these standards. Further, the Advisory Committee is convinced that evaluators cannot validly judge how well educators understand and use assessment to improve learning (standard 6) without hearing or reading how educators use student work to reflect on what was revealed in the assessment process and what instructional decisions should be made based on these results.

On the other hand, capturing information about educators' use of appropriate instructional strategies (standard 8) would be very difficult without direct classroom observations. The Advisory Committee recognizes that any manageable schedule of observations will be necessarily "thin." In the years that the teacher is evaluated, the Advisory Committee **recommends** that teachers are observed formally on at least three different occasions and supplemented with more frequent "walk-through" or informal observations. The general time frame/unit of instruction for the observations shall occur in consultation with the educator, but at least some of the specific lessons observed shall be unannounced. At least one of the observations, but preferably most of them, should be tied to aspects of the curriculum that are the focus of the Student Learning Objectives (SLOs) in order to use information about what students have learned to triangulate the teacher practice information. Further, the observations should include an analysis and discussion of relevant documents associated with the unit of study being observed. These documents may include lesson plans, assessments, assignments, student work, and other relevant documents associated with the teaching, learning, and assessment of the unit.

Domain 4: Professional Responsibility

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

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

Professional responsibility may be observed informally through noticing how the educator interacts with colleagues, parents, or others, but it is unlikely that information about professional responsibility can be collected through formal classroom observations. The Advisory Committee recommends that the yearly self reflection and goal setting activities specifically address aspects of professional responsibility and establish the focus of professional responsibility of the given year. The Advisory Committee deliberated whether teachers new to the profession should be exempted from being evaluated on Domain 4, but overwhelmingly recommended that all educators should be expected to demonstrate their responsibility as a professional educator. One potential difference between novice and experienced educators is that novice educators may focus on more "inward-facing" aspects of this domain, as discussed in standard 9. On the other hand, experienced educators may continue to focus on these internal aspects of responsibility, but they would also be expected to become more "outward-facing" leaders whether in the school, the district, or the profession at large. The specific focus of the professional responsibility will guide the required data collection and reflection.

Data Collection Strategies

There are several data collection tools, such as observations, surveys, and artifact analysis described for many of the standards, but that does not mean that each one requires a separate data collection. The Advisory Committee recommends using the following data sources as part of each educator's evaluation:

Documentation of Practice: Self Assessment, Goal Setting, and Artifact Collection

The collection of artifacts is a critical component of WY's State Model and contributes data to multiple domains of teacher practice. The collection of artifacts is guided by educators' self-assessment, but also by the districts' elaboration of the Specific Measurement Model. The

Advisory Committee recommends having all educators establish yearly professional goals in consultation with their supervisor or designee and document the process and products associated with these goals through a selective collection of artifacts documenting teacher's professional practices and evidence of student learning. The Wyoming Department of Education (WDE) will produce resources designed to support the use of artifacts for evaluation purposes. The Advisory Committee recommends that each educator's evaluation incorporate the following components:

- Self assessment of strengths and weaknesses,
- Collaboratively (among educator, administrator, and perhaps peer team) established specific goals,
- A plan describing specific professional learning opportunities and needs for achieving the goals,
- Analyses of key artifacts such as student work from specific assignments, planning documents, and assessments related to the established goals, and
- Self-reflection at the end of the year to self-evaluate the extent to which the specific goals have been achieved.

The specific collection and use of artifacts should depend on each district's evaluation plan and the suggestions for data collection described throughout this document.

Observations of Professional Practice

As indicated earlier, evaluators will need to "see it" in order to document that educators are able to enact key aspects of professional practice. This evidence will be derived from classroom observation protocols designed to measure performance related to the ten InTASC standards and will likely involve the use of commercially-available or other existing tools. Most of these existing tools are able to provide evidence about the degree to which the educator is enacting expected InTASC practices. However, these tools may also be designed to collect data for practices not part of the InTASC framework. The district must ensure that it collects data on at least the InTASC standards. It may also collect data on non-InTASC-related practices, but not at the expense of InTASC standards. Therefore, districts must establish a clear and transparent approach for how it will enact the tools it has selected to use, both in terms of aspects of the tools for which educators will be accountable and for which they will not.

Unfortunately, many school leaders have not been adequately trained and/or do not have enough experience at accurately evaluating instructional practice, although they generally have extensive experience observing and providing feedback on classroom management behaviors. Therefore, each district must, as part of implementing its classroom observation tool, ensure that those responsible have training at a level that enables the evaluator to distinguish fine gradations in instructional practice. As importantly, these evaluators must also have training and support to enable them to provide actionable feedback to educators based on the results of the observations.

Student Performance

As stated in the first guiding principle of this State Model, the primary purpose of Wyoming's educator evaluation and the reason for engaging in this work is to support and promote increases in student learning in Wyoming schools. Therefore, the results of student achievement must be incorporated in the evaluations of all educators. While this sounds intuitively straightforward, it is one of the most complex aspects of new forms of educator evaluation. The Wyoming State Model uses a three part approach for incorporating student achievement and growth into evaluations in order to attempt to maximize the validity of educator evaluations, while striving to minimize potential unintended negative consequences.

Student Learning Objectives (SLO) form the foundation of Wyoming's approach for documenting changes in student performance associated with a teacher or group of educators and, as such, all educators will have the results of SLOs incorporated into their evaluations. For educators in "tested" subjects and grades, those grades and subjects for which there is a state, standardized test as well as a state test in the same subject in the previous year, student performance will be evaluated using Student Growth Percentiles (SGP). The results of SGP analyses, along with SLO results, will be used in the evaluations of educators in tested subjects and grades. SGPs and SLOs are described in more detail below.

Both SGP and SLO approaches can be used to attribute the academic achievement and growth of students to individual educators or to appropriate aggregations of educators such as grade- or content-level teams or even the whole school. Distributing student performance results to multiple educators is referred to as "shared attribution." The tradeoffs associated with shared attribution are also discussed below.

Student Learning Objectives (SLO)

> *SLO are content- and grade/course-specific measurable learning objectives that can be used to document student learning over a defined period of time. To boil SLO down, they provides a means for educators to establish learning goals for individual or groups of students, monitor students' progress toward these goals, and then evaluate the degree to students achieve these goals. The active involvement of the teacher throughout the process is a key advantage of the SLO approach over traditional test-centered approaches to accountability. It is designed to reflect and incentivize good teaching practices such as setting clear learning targets, differentiating instruction for students, monitoring students'*

progress toward these targets, and evaluating the extent to which students have met the targets[2].

All teachers, whether in "tested" or "non-tested" subjects and grades shall be required to document student academic performance each year using SLOs in accordance with Wyoming's SLO guidance (to be developed). Both SGP and SLO analyses shall produce results in at least three classifications of performance, to the extent possible, such as: high, typical/average, and low. The results of the SLO determinations shall be incorporated into the evaluation of all educators according to the rules described below in the section on combining multiple measures.

Calculating Student Performance Results in "Tested" Subjects and Grades

The growing interest in reforming long-standing approaches for evaluating and compensating teachers has been characterized by among other things incorporating student performance results in teacher evaluations. Advances in growth and value-added models in education have contributed to the interest in using changes in student test scores over time as part of educator accountability systems. Many districts, states, and non-governmental organizations have embraced these test-based accountability initiatives, but the initial focus has been on the content areas and grades for which there are state standardized tests, generally administered at the end of each school year, i.e., "tested subjects and grades." Student performance for these tested subjects and grades is generally evaluated using complex statistical models such as value-added or student growth percentile models. It is important to realize that while these statistical approaches have led to the popularity of incorporating student achievement results into teacher evaluations, they generally apply to approximately 25% of the teaching population. More specifically, in Wyoming, the results of such analyses could be applied to only those educators teaching math and language arts in grades four through eight.

There are several possible approaches that Wyoming could use for evaluating student performance in tested grades, but in order to adhere to the coherence principle, the Advisory Committee recommends using the same Student Growth Percentile model currently being used for the school accountability system. However, this is not necessarily as simple as it sounds to move from school to teacher accountability (subsequent guidance will describe the multiple considerations for using SGPs in educator evaluation).

WDE shall produce Student Growth Percentiles (SGP) results documenting the individual student and aggregate growth for students. These results will be reported for the whole school level and for identifiable student groups in the school. A student–level file will be provided to

[2] Marion, S., DePascale, C., Domaleski, C., Gong, B., & Diaz-Bilello, E. (2012, May). Considerations for analyzing educators' contributions to student learning in non-tested subjects and grades with a focus on student learning objectives. Center for Assessment. Retrieved from www.nciea.org September 3, 2013.

each district to use for aggregating SGP results according to the attribution rules in each districts' evaluation plan, whether for individual teachers, specific groups of teachers, or both. These SGP results, based on PAWS or other state-mandated assessments, shall be incorporated into teachers' evaluations either using a shared or individual attribution State Model.

Shared Attribution

The Advisory Committee recognizes the challenges of properly attributing the results of student performance to individual teachers. It is easy to think of many examples where it does not make sense to attribute the performance of students to any individual teachers, such as the case when grade-level teams of teachers place students into differentiated instructional groups and instruction is provided to students by educators other than the child's "regular" teachers. Therefore, the Wyoming State Model relies on a mix of shared attribution and individual attribution of student performance results. The SGP results, based on state tests in grades 3-8 should, depending on the specific theory of improvement for the particular school, be shared among educators at the same grade and/or teaching the same subject areas. SLO results, assuming groups of educators are working on the same SLO, may also be shared among educators at the same grade and/or content area. However, SLOs allow for more control and tailoring to specific courses than state test results so the Advisory Committee recommends that at least a portion of the SLOs used to document student performance be attributed to the individual educator of record. Like anything else in accountability system design, there are both advantages and disadvantages to using shared attribution.

One of the major concerns with attributing the results of student performance to individual teachers is that many fear that this could erode collaborative cultures at many schools, especially if the results are used in some sort of "zero sum game" accountability design. Shared attribution approaches, if implemented sensibly, can help promote both collaboration and internal accountability orientations, both of which are associated with high performing schools and organizations. Another concern for policy makers and accountability system designers are potential unintended negative consequences of having the mathematics and reading teachers in grades 4-8 evaluated in potentially very different ways than the other 70-80% of educators in the district. This could lead to higher rates of attrition from these subjects and grades or perhaps feelings of professional isolation. The requirement for all educators to participate in the SLO process is one hedge against this potential problem. However, sharing the results of all of the student performance indicators among multiple educators, as appropriate, is another way to recognize the contributions of other educators to student performance, especially in reading and math. Finally, one of the major concerns with tying student performance results to individual teachers involves the reliability concerns when dealing with such small groups of students. Aggregating the student performance results for multiple educators is one way to ameliorate, but far from eliminate, these reliability challenges.

This discussion could lead one to believe that if shared attribution has so many advantages, why would a system include any other approach? Of course there are potential disadvantages to shared attribution. One important disadvantage—that may be reduced with careful design—is that educators may be held accountable for results for which they may have little to no control. This was a considerable criticism of Tennessee's approach for including student performance results in the evaluations of teachers from non-tested subjects and grades. This threat is likely greatest when student performance on the state math and/or reading tests is attributed to all educators in the school as opposed to a finer-grained aggregation. Another potential disadvantage to shared attribution is that it may mask true variability in educator quality. If one believes that educator quality is truly variable in terms of being able to influence student performance, then pooling results among multiple educators could mask such differences. Of course, being able to separate the "signal" (true variability) from the "noise" (unreliability in the system) is not easy with such small samples. This more problematic at the elementary level with self-contained classroom of 20 students or so compared to a middle school where a teacher might be responsible for the math or reading instruction of over 100 students. The Advisory Committee is well aware that this assumption of greater numbers at the secondary level still may not hold true in many of Wyoming's small schools and districts.

Therefore, the Advisory Committee recommends sharing student performance results among multiple educators according to local theories of improvement and not based simply on reliability concerns. For example, if the focus of improvement activities is the grade level team, then attribution should be shared among educators at that grade and not at the whole school level. Therefore, the first step in implementing any sort of shared attribution approach involves a careful articulation of the school's locus of improvement actions. This theory of improvement (action) should also make clear which subjects are shared and with whom. For example, does the 5th grade team share both math and ELA results or just one subject? Finally, while the Advisory Committee favors shared attribution approaches in many cases and for at least some of the weight in the accountability determinations, it also recommends that at least some of the changes in student performance be attributed to individual teachers. This might best be accomplished with SLOs rather than SGPs because of the closer ties to the specific course, but the Advisory Committee suggests leaving this decision to local school districts.

Student surveys

The Advisory Committee discussed the merits and challenges associated with incorporating results from student surveys into teacher evaluation decisions. On one hand, using information from students solves a major "sampling problem" associated with both teacher observations and student test scores. Even an ambitious observation schedule of four or five one hour

observations in a year (and most would consider 2-3 ambitious) is still four or five hours out of a possible 720 instructional hours each year (180 days x 4 instructional hours each day). Student Growth Percentiles (or value-added models) based on PAWS or other state-mandated assessments, while technically strong, are only a sample of students' knowledge and skills and suffer from limited reliability based on small numbers of students in a given class or school in Wyoming's case. Student surveys, in contrast, collect information from those with the teacher essentially 100% of the teacher's instructional time. Further, by including enough questions (e.g. 25-40), it is possible to generate fairly reliable results. In fact, the student surveys were the most positive influence on the reliability of the composite rating of teachers' performance in Measures of Effective Teaching (MET) project when surveys, value-added model results, and observations were combined for a teacher rating.

On the other hand, increasing reliability does not mean increases in validity will automatically follow. Several researchers have raised concerns that having students participate in the consequential evaluation may change the "social contract" in the classroom. This concern should not be taken lightly and if surveys are used, care must be taken in the design to deal with potential challenges to the validity of the teacher evaluations.

The Advisory Committee has several recommendations if surveys are incorporated into district evaluation systems:

1. Survey questions must be predominantly "low inference" type questions that ask about specific practices (e.g., "how many times each week does your teacher ask you to explain your reasoning") compared with questions about feelings (e.g., "does your teacher care about you?").
2. Surveys should be piloted extensively so students can get used to completing surveys and school personnel can gain an understanding of how the surveys relate to other information about teachers.
3. Instead of incorporating the results of surveys into evaluations directly, districts should consider using surveys as an additional factor to raise or lower a teacher's overall rating.
4. In order to most conservatively provide the type of additional information called for in #3, districts and schools should consider using the surveys normatively. In other words, the survey results would only be a factor to adjust the evaluation results if the teacher's survey results were noticeably higher or lower than the average for other teachers at that same grade span.
5. Student surveys should be designed to provide information regarding the standards for which students would likely have meaningful insights. This would include most of Domain 1 as well as standards 5, 6, and 8.
6. Finally, given the capacity and cost required to produce valid and reliable surveys, the Advisory Committee recommends that WDE be charged with producing or selecting

model surveys that districts can use if they choose. There should be model surveys designed at least for each grade span and different content areas, if applicable.

Performance Standards

All Wyoming schools, as determined by their districts, will classify all licensed personnel, as illustrated by the State Model, as **highly effective, effective, needs improvement, and ineffective** based on data from measures of the standards for professional practice and measures of student performance. The evaluation system will produce an overall rating for each teacher. To arrive at an overall rating, a description of performance that characterizes the types of knowledge, skills, dispositions, and behaviors of an "effective" teacher (as well as other levels) must be described. Further, if there is any hope in comparable ratings across the state, common performance level descriptors must be used. Performance standards describe "how good is good enough" and the "performance level descriptor" (PLD) is the narrative component of the performance standard that describes the key qualities that differentiate educators at each of the various levels. These PLDs are critical to help guide the data collection and validity evaluation of the system.

The InTASC Standards provide performance descriptors for each of the ten standards, but they do not provide an overall description for various levels of teacher effectiveness. One might ask, why not require educators to meet the requirements on each of the ten standards in order to be classified as effective? This type of conjunctive system where candidates must meet every threshold in order to be classified as "effective" is both unrealistic and unreliable. We discuss various approaches and recommendations for combining multiple measures in the following section of the document.

The State Model provides PLDs for each of the four overall levels of the system. These descriptors connect the standards for professional practice with the various data produced by the measurement instruments used in the system. This overall description is necessary, because an effective teacher is not necessarily a simple sum of the scores on the various components/indicators in the system. The PLDs in this document present the Advisory Committee's recommendation for how the ten INTASC standards should be combined into an overall classification of educator effectiveness.

Ultimately, each district system must be able to validly classify its educators into four levels of performance as described by the following policy-level PLDs. Each PLD essentially describes the final evaluation of how well a teacher has performed in any given year based on all factors considered. **The Advisory Committee strongly endorses employing a set of common**

performance descriptors for Wyoming in order to promote comparable expectations for educators across districts.

Highly Effective

*Teachers performing at the highly effective level **consistently** advance growth and achievement of students at levels to ensure that students **meet or exceed** important growth and achievement targets. They set and maintain high expectations for learning and achievement for all students and create learning experiences and inclusive learning environments **consistently reflective** of individual differences.*

*Highly effective teachers demonstrate **extensive knowledge** of their **content** area, consistently making connections among concepts to engage learners. Highly effective teachers **consistently** use their expertise and skills to employ research-based strategies to **frequently engage** their students in **authentic, accessible, and meaningful learning** opportunities aligned to the content, standards and related skills. They are **knowledgeable in multiple** forms of assessment and incorporate these multiple assessment strategies to evaluate student learning and adjust instruction accordingly as part of their regular practice. **Highly** effective educators integrate technology into their instructional and assessment approaches in ways that advance student learning opportunities.*

*Finally, highly effective educators **consistently demonstrate leadership** in their contributions to their school's academic progress and culture of growth. They engage productively in learning communities and **continuously strive to maximize** their own self-directed professional growth and that of their colleagues. These educators consistently uphold **high** standards of professional practice.*

Effective

*Educators performing at the effective level **generally advance** student growth and achievement at levels for students **to meet** important growth and achievement targets. They set and maintain high expectations for learning and achievement for all students and create learning experiences that are **mostly reflective** of individual differences and inclusive learning environments.*

*Effective teachers demonstrate **strong knowledge** of their **content** area and **often** use their knowledge and skills to employ research-based strategies to **regularly engage** their students in **authentic, accessible, and meaningful learning** opportunities aligned to the content standards and related skills. They use assessment evidence to evaluate student learning and adjust instruction accordingly. Effective educators appropriately integrate technology into their instructional and assessment approaches to maximize student learning.*

Finally effective educators engage in learning communities, fostering their own self-directed professional growth, and frequently provide leadership to support improvements in their colleagues' performance, making regular contributions to their school's academic progress and culture of growth. These educators consistently uphold professional standards of practice.

Needs Improvement

*Educators performing at the needs improvement level **inconsistently advance** student growth and achievement such that only some students **meet** important growth and achievement targets. It is **not evident** that they set high expectations for learning and achievement for all students. They **are inconsistent at creating** learning experiences that reflect an understanding of individual differences and inclusive learning environments.*

*Teachers in the needs improvement category demonstrate **a basic knowledge** of their **content** area and **occasionally** employ research-based strategies to **engage** their students in **authentic, accessible, and meaningful learning** opportunities aligned to the content standards and related skills. Teachers in the needs improvement category use assessment evidence to evaluate student learning, but it is not evident if or how they adjust instruction based on assessment results. These educators use technology in their instructional and assessment approaches.*

Finally educators performing at the needs improvement level participate in learning communities, but inconsistently attend to their own self-directed professional growth. These educators uphold professional standards of practice.

Ineffective

*Educators performing at the ineffective level may advance some student growth and achievement, but **frequently fail** to have students **meet** important growth and achievement targets. There is little evidence that they have established ambitious and reasonable expectations for student learning for most students and generally do not engage students in appropriate learning opportunities.*

*Educators performing at the ineffective level may have a **limited** knowledge of their **content** and **rarely** employ research-based strategies to **engage** their students in **authentic** learning opportunities. Teachers in the ineffective category have a very limited assessment repertoire and there is little evidence that they use assessment results to adjust instruction. These educators have not fully or consistently integrated technology as part of their instructional and assessment approaches.*

Finally educators performing at the ineffective level may participate in learning communities, but there is little evidence that they make substantive contributions to their own professional

growth and/or support the growth of their colleagues. These educators generally uphold professional standards of practice.

Combining Multiple Measures

As discussed above, there are many approaches for combining multiple indicators to yield a single outcome: *compensatory*, *conjunctive*, *disjunctive*, and *profile* methods. Compensatory means that higher performance in one measure may offset or compensate for lower performance on another measure. Conjunctive means that acceptable performance must be achieved for every measure (e.g., AYP). Disjunctive means that performance must be acceptable on at least one measure. A profile refers to a defined pattern of performance that is judged against specific performance level descriptions. A profile approach is often operationalized using a matrix to combine indicators for making judgments. Given the challenges involved in characterizing the complexities of teaching, the State Model must employ a thoughtful approach for combining the multiple sources of data in order to produce the most valid inferences about overall teacher quality possible.

A compensatory approach recognizes that some degree of variability in performance across indicators may be expected. Such an approach has a higher degree of reliability because the overall decision is based on multiple indicators evaluated more holistically. Conjunctive decisions are less reliable because errors accumulate across multiple judgments meaning a teacher might fail to be classified as effective due to poor performance on the least reliable measure. A conjunctive approach does not appear to make much sense for an educator evaluation system. A disjunctive method is used when any one component is viewed as adequate assurance the teacher met expectations. Again, this does not appear to make much sense in a teacher evaluation system. Finally, profiles are useful especially when there are certain patterns that can be described that reflect valued performance that are not easily captured, usually because the combinations of criteria are judged to be not equivalent.

These approaches should not be regarded as mutually exclusive. It is possible, for example, to combine aspects of compensatory and profile 'rules' to arrive at a final result. For example, a compensatory approach may be used to aggregate the data from the multiple measures within any single domain, while a profile approach could be used to combine information across domains. A major advantage of a profile or decision matrix approach is that once established, the teacher can never receive an unexpected overall rating, whereas simple averages characteristic of compensatory approach can produce some surprising outcomes.

The Advisory Committee recommends using, as part of the State Model, an approach for combining the various sources of information that avoids mechanistic approaches such as simple averaging, but that takes into account the nature of the different sources of information. A

"panel" or "decision matrix" approach" for combining the multiple measures allows the goals of the system to be reflected explicitly and not buried in some numerical composite.

Each local educator evaluation system must be able to produce overall classifications for each educator in the district. The Advisory Committee recommends that the decision matrix described below be used by all districts to combine the results of teacher practice and student learning results. This will allow for at least some level of comparability across Wyoming districts. However, the Advisory Committee is willing to allow local districts to determine how to best combine data to arrive at these penultimate indicators.

NOTE to Advisory Committee: We need to deliberate on the specific matrix we want to put in the document as well as the values we want to put in the matrix.

EXAMPLE A 4x 3 Panel Approach for Combining Multiple Measures (based on an approximate 25/75 weighting between student performance and teacher practices)

"Professional Practice" Rating	**4**	Automatic Review	Highly Effective	Highly Effective
	3	Needs Improvement	Effective	Effective
	2	Needs Improvement	Needs Improvement	Needs Improvement
	1	Ineffective	Ineffective	Automatic Review
		1	**2**	**3**
		"Student Performance" Rating		

Again, this is just an **example** and this sort of 4 x 3 matrix might be useful with an immature system such as the type we would expect during early implementation phases. As the system matures and more data are available for each educator, particularly in terms of student performance, more expansive matrices may be appropriate, such as the **example** of 5 x 4 matrix below.

EXAMPLE #2: A 5x 4 Panel Approach for Combining Multiple Measures (based on an approximate 25/75 weighting between student performance and teacher practices)

"Professional Practice" Rating"					
	5	Needs Improvement	Effective	Highly Effective	Highly Effective
	4	Needs Improvement	Effective	Effective	Effective
	3	Needs Improvement	Needs Improvement	Effective	Effective
	2	Ineffective	Needs Improvement	Needs Improvement	Needs Improvement
	1	Ineffective	Ineffective	Needs Improvement	Needs Improvement
		1	**2**	**3**	**4**
		"Student Performance" Rating			

Supports and Consequences

Assumptions

As stated in the guiding principles, Wyoming's State Model is being designed such that it can support improvements in teaching and learning. As part of this design, the Advisory Committee emphasizes the importance of reporting detailed and actionable information so that educators and their leaders have the information they need to guide efforts to improve their practice. This means that educators need to receive information on each of the indicators in the system, while recognizing that the information at the indicator level is considerably less reliable than the total evaluation. This will require having thorough documentation produced for each local system, in terms of the components and indicators outlined in this document, so that all educators understand the nature of the information on which they will be evaluated.

The WY State Model and all local systems must produce an overall effectiveness rating that guides support, career development, and employment decisions. The overall rating can only be an overall flag to guide support since the detailed information is necessary to allow for focused support and development.

Supports

A critical support requires having each educator understand the rules by which they will be evaluated. Therefore, each district shall develop and implement a process for training all licensed personnel on the educator evaluation system including the consequences associated with the ratings.

One of the major guiding principles of Wyoming's educator evaluation system is that it should lead to improvements in educators' performance. Therefore, the Advisory Committee recommends that each Wyoming school district include well-specified and formalized processes of mentoring and support designed to improve the performance of all educators in the district. The support and mentoring systems *should* be designed collaboratively with teachers and administrators based on research and documented best practices.

Districts shall provide training for all personnel who will be conducting classroom observations as part of a defined training and qualification process. This training will help leaders better understand differences in instructional quality so that they can better support their teachers' improvement efforts. Additionally, all evaluators (administrators) must receive evidence-based training on how best to provide feedback to those evaluated in order to support understanding of the information derived from the evaluation system and to improve practice.

Consequences

Ultimately, the system will lead to certain consequences for educators falling well below or well above expectations. While the system is designed for improvement and a significant support system is required to help struggling educators, there will likely come a point where educators may need to be counseled out of the profession. The State Model includes the following expectations for such eventualities:

1. Educators rated *ineffective* or *needs improvement* in one year must be placed on *directed professional growth* (improvement) plan that includes receiving targeted support. These support systems must be research-based to the maximum extent possible. Further, the evaluations of the educators involved in a directed professional growth plan shall include additional data sources in the evaluation during the improvement plan year.
2. The State Model requires that an experienced, educator with **two consecutive** years of *ineffective* ratings lose her/his current (continuing contract) status and may be dismissed without additional cause. The Advisory Committee recognizes that such potential consequences will need to be incorporated into locally-negotiated personnel contracts.
3. After receiving a second consecutive "*needs improvement"* rating, the educator will be considered to have received his/her first year of an *ineffective* rating.
4. An educator rated *highly effective* for two consecutive ratings should receive recognition, as determined by the local district, and may assume a "teacher leader role" as part of the mentoring and support system.

Recommendations for Implementation

In this section, the Advisory Committee provides recommendations for enacting a process to efficiently implement this system.

1. The evaluation process starts with the educator's assessment of his/her strengths and shortcomings, which then leads into establishing goals for each year. This should be based on a review of existing data from past formative and summative evaluations, student performance results, and other relevant information. This cycle could/should start in the spring for continuing educators.
2. After the self-assessment is completed, the supervisor and/or mentor and educator meet to discuss the goals. They may revise the goals tied to specific standards, but will ultimately come to agreement on the goals.
3. The supervisor and leader agree on a data collection plan to best evaluate the yearly goals. However, each district will have certain data collection protocols, such as specific types of observations, surveys and assessment analyses, enacted for all educators.
4. Typically, in well-functioning evaluation systems there will be multiple progress monitoring meetings between the supervisor and educator, but the Advisory Committee strongly recommends that there is at least a yearly formative or summative evaluation for all educators in the district.
5. Additional data will be collected throughout the rest of the year based on the initial goals and the results of the mid-year conference.
6. The evaluation cycle concludes with a summative evaluation at the end of the third year. While this is technically the end of the evaluation cycle, it is also the beginning of the subsequent evaluation cycle. The specific beginning and end of each district's cycle will be locally determined based on negotiated contracts.

The following graphic represents the general process for implementing an educator evaluation system in Wyoming school districts.

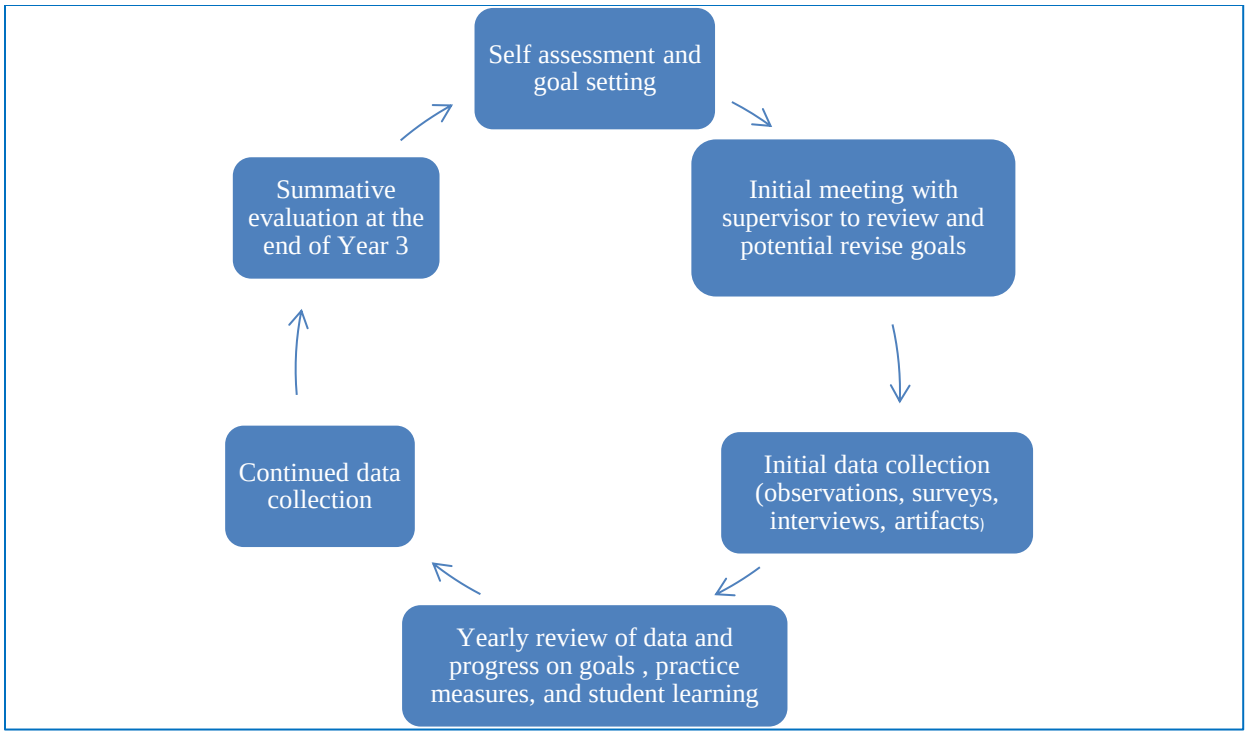

Figure 1. Educator evaluation cycle[3].

The Advisory Committee has been very thoughtful about designing a State Model for educator evaluation in Wyoming. We have attempted to outline a clear approach to addressing the complexities for designing and implementing educator evaluation systems in Wyoming. However, the Advisory Committee wants to stress that there are enormous challenges to implementing such systems in any locale. One positive aspect of having Wyoming follow other states and districts in this work is that we have the opportunity to learn from the experiences of others. One of the most striking things being learned is that significant time and thoughtfulness are needed to implement these systems well.

This would be true under conditions where the state standards and assessment systems were stable. As we know, Wyoming has recently adopted the Common Core State Standards which call for deeper levels of understanding on the part of students than ever before. Shifting instructional practices and curriculum will require considerable effort on the part of local school districts. Adding requirements for a new school accountability system will further stress systems. Therefore, the Advisory Committee appreciates that the educator evaluation system in Wyoming can be implemented with an extended pilot period to both gradually implement the system and to allow for formative feedback to make adjustment to the system before it is implemented operationally.

[3] Note: This is based on a three-year evaluation cycle for experienced and effective educators. Novice and educators not yet rated effective will be summatively evaluated each year.

The Advisory Committee has been sensitive to balancing the needs of creating a valid system with an understanding that the system or one like it must be implemented by all school districts without creating an unmanageable burden. While many states require a full evaluation of every teacher every year, the Advisory Committee quickly recognized that this would place an impossible and inefficient burden on WY schools. Therefore, the Advisory Committee recommends differentiating evaluations according to the experience and status of the schools' educators. Ultimately, each district shall enact a policy and set of procedures to differentiate evaluation systems for its different classes of educators (e.g., novice, veteran, and/or high performing, low performing) and to the specific evaluation questions to be investigated. Each educator shall be evaluated at least once, using the full system, within the first three years of implementation, with novice educators evaluated every year. To the extent possible, yearly evaluations shall include multiple years of student performance results.

Novice educators, defined as those within the first three years of the teaching profession, must be evaluated **every year** until they are rated "effective" for **two** consecutive years. Districts may decide to focus specific aspects of the evaluation for novice educators by reducing the demands of certain aspects of the systems and/or focusing the evaluation on specific standards.

Teachers with professional status (continuing contract) receiving an ineffective or needs improvement rating shall be evaluated every year until they receive "effective" ratings or better for **two** consecutive ratings or until other actions are taken. Once these teachers receive two consecutive effective ratings, they shall receive summative evaluations every three years.

In addition to multiple measures, the Advisory Committee recognizes the challenge of having enough expertise and time in any single individual to conduct all required evaluations. Therefore, the State Model includes the optional use of peer teams, in addition to building-level administrators, to participate and advise in the evaluation process.

Review Process

References

Data Quality Campaign. (2010a). Strengthening the teacher-student data link to inform teacher quality efforts. Retrieved from: www.DataQualityCampaign.org/resources/947.

Data Quality Campaign. (2010b). Developing a definition of teacher of record. Retrieved from: http://dataqualitycampaign.org/files/Teacher%20of%20Record.pdf.

National Research Council. 2010. Getting value out of value-added. H. Braun, N. Chudowsky, and J. Koenig (eds.). Washington, DC: National Academy Press.