Wyoming Educational Accountability: Phase I Report

SCOTT MARION, PH.D. & CHRIS DOMALESKI, PH.D.
CENTER FOR ASSESSMENT

PRESENTATION TO THE WYOMING JOINT EDUCATION COMMITTEE

JANUARY 17, 2012
Plan for Presentation

• Brief introductory remarks
• Review key provisions of the comprehensive accountability system framework
• Highlight processes for moving from framework to implementation
The Center for Assessment

- Non-profit consulting firm established in 1998 with the mission of improving student learning through improved assessment and accountability practices
- Current contracts with 30+ states/entities
  - Almost all are long-term contracts designed to provide technical and design support for a range of assessment and accountability issues
- Purposely small—12 full-time professionals
  - All with doctoral degrees, but who have worked in the “real world”
The Advisory Committee to the Select Committee is comprised of a smart and insightful group of Wyoming educators and other stakeholders who contributed tremendously to the proposed designs and this report:

- Bill Schilling, (Committee Chair), President, WY Business Alliance
- Tony Anson, HS Principal, Big Horn #4
- James Bailey, Superintendent, Uinta 1
- Sue Belish, SBE, former Superintendent, Sheridan 1
- Diana Clapp, Superintendent, Fremont 6
- Kris Cundall, Elementary Principal, Sweetwater 1
- Mary Kay Hill, Governor’s Office
- Brian Kaumo, Principal, Sweetwater 1
- Molly Kinsey, Instructional Facilitator, Sheridan #2
- Sheryl Lain, WDE
- Jack Patrick, HS English Teacher, Carbon
- Janine Bay-Teske, Board Trustee, Teton 1
Wyoming currently has several educational accountability systems in place:
- **Districts**: Accreditation and School Finance Model
- **Schools**: NCLB-AYP
- **Students**: Body of Evidence/Graduation requirements
The major charge of our work with WY

Too many states end up adding accountability and assessment components piecemeal—leads to incoherence

Wyoming is poised, given SF 70, to align major aspects of the assessment and accountability systems

The work of the Select Committee resulted in a comprehensive accountability framework to serve as a touchstone for WY educational policy going forward
The Comprehensive Accountability Framework

• Does...
  ○ Serve as a foundation for legislation in the 2012 session
  ○ Clearly communicates to stakeholders a vision for educational accountability in Wyoming
  ○ Provide a design for a school accountability
  ○ Outline a process and considerations for developing educator and student accountability systems

• Does Not...
  ○ Contain enough detail (e.g., business rules) to immediately work as the operational accountability system
  ○ Include empirical analyses (at least in this phase) to fully test the system
  ○ Specify the educator and student accountability systems at this point
The Comprehensive Accountability Framework...

- Articulates the values driving the overall system and its various components
- Explicitly describes the intended goals and outcomes of the system
- Outlines a “theory of action” that describes how the components are intended to work together to produce the desired outcomes
- Serves as a guide to the design of the overall system and specific components
- Provides specific steps necessary for designing the full operational system
The Comprehensive Accountability Framework

I. Background
II. Conceptual Foundations
III. The Multiple Accountability Initiatives
IV. Consequences, Support, and Capacity Building
V. Evaluation and other Technical Considerations
Goals

- National educational leader among states
- All students leave Wyoming schools “college or career ready”
- Increase the rates at which Wyoming students learn
- Reduce and eventually minimize gaps in achievement
- Improve quality of teaching and leading
- Maximize efficiency of Wyoming education
- Increase credibility of and support for Wyoming public education
Design Principles

• Support the Instructional Core
• Coherence
• Equity
• Transparency
• System Support and Improvement
• State-Local Partnership
• Shared Responsibilities—beyond education
Multiple Accountability Initiatives

- School
- Teacher
- Leader (all levels of the system)
- Student

- **Phase I**: Focus on school accountability and outline processes for developing teacher/leader and student accountability components
- **Phase II**: Development of educator and student accountability systems
Guided by Theory of Action

- Using a theory of action as a guide to the design allows us to recognize and capitalize on the differences between reporting and accountability.
- We recommend collecting data and reporting information on indicators and processes that support the theory of action, but we also generally recommend including in the accountability system the least corruptible outcome measures.
School Accountability Components

- **Achievement** – To what extent do students meet standards for proficiency on state tests?
- **Readiness** – Do high school students achieve outcomes meaningfully related to post-secondary success?
- **Growth** – To what extent do students show progress on state tests?
- **Equity** – To what extent do low performing students make progress toward proficiency?
- **Inclusion** – Are students served for the full academic year included in accountability measures?
Achievement refers to indicators that provide information about student academic performance with respect to Wyoming state standards.

Based on either proficiency rates on PAWS in reading, mathematics, science, and writing in applicable grades or an “index” system for the same tests:

- Index system provides partial credit for students scoring below proficiency and can allow for additional credit for students scoring at advanced levels.
Readiness

- Has generally been considered by the Advisory Committee as indicators of college or career readiness
  - Such as ACT scores, high school graduation rates
- The Advisory Committee is also interested in exploring readiness measures that can be employed validly at the elementary and middle school levels
- Readiness, however, is a very complex construct...
Dimensions of College Readiness (Conley)

- **Key Cognitive Strategies**: Also known as “habits of mind, e.g., inquisitiveness, persistence, intellectual openness
- **Key Content Knowledge**: Broken into Overarching (e.g., writing, research) and **Core Academic** (e.g., reading, math)
- **Academic Behaviors**: Self awareness and self-regulation
- **Contextual Skills**: Also known as “college knowledge”
The current focus of readiness discussions with the Select and Advisory Committees has focused almost exclusively on one sub-part of Conley’s framework (core academic skills) via the use of the ACT suite or other college entrance exams.

One could argue that high school graduation also taps “academic preparation,” although given the relatively low standard for graduation, this claim is hard to substantiate.

Therefore, in keeping with the committees’ recommended distinction between reporting and accountability, we can broaden the view of readiness...
Reporting system can include:

- Credit accumulation at key checkpoints (e.g., 9th grade)
- Course completion/ success
  - Enrollment and/or performance in AP/IB or other ‘advanced’ courses
  - Participation in joint-enrollment or other post secondary courses at the secondary level
- Student surveys of attitudes, dispositions, and behaviors
- Attainment of career/ industry certifications
- Follow-up surveys to report on achievement of post-secondary outcomes
Two categories of indicators are recommended, at this point, for inclusion in accountability determinations: academic performance and graduation rate

- Academic performance: EXPLORE, PLAN, and ACT
- Graduation Rate: outcome at the completion of high school
• ACT & PLAN for high school system
• The metric would be the percent of students meeting identified readiness benchmarks

• EXPLORE for middle schools
• An analogous metric would be used for middle schools
While graduation rate can be similarly incorporated into the accountability system, it may desirable to consider multiple levels of performance.

To accomplish this, an index can be created that awards points in proportion to the value of the outcome in year 4.

- For example, the State may wish to value (in terms of more points) 4-year graduation more than GED.

The score for this component is simply is the average of all student outcomes for the high school.
Growth Recommendations

- **Do NOT** legislate a specific growth model
- **DO** legislate:
  - A model that measures changes in student performance as fairly and accurately as possible all along the achievement continuum
  - A model that is “scale independent”
  - A model that is either open source or owned by the State
  - A model that includes both criterion and norm-referenced metrics
The Advisory Committee has recommended using Student Growth Percentiles (SGP) in reading and mathematics, but we argue that the key principles can apply to different models that meet the previous recommendations.
Student Growth Percentiles (SGP) works by evaluating current achievement based on prior achievement and describing performance relative to other students with the “same” prior achievement histories.

This provides a familiar basis to interpret performance – the percentile, which indicates the probability of that outcome given the student’s starting point.

This can be used to gauge whether or not the student’s growth was atypically high or low
The Binominal Distribution

How does it work?

Think of a group of students, where each student has two test scores – one for 2009 and one for 2010.

We could show the distribution of these scores at the same time as pictured.
Slicing the distribution at the Year 1 score

We could ‘slice’ through the picture to show the 2010 distribution for just one 2009 score. This is called a conditional distribution.

The red shaded curve shows the conditional distribution in 2010 for all students who scored 600 in 2009.
Comparing Year 2 scores for all who scored 600 in Year 1

Assume we are interested in just one score, 650, in 2010.

We could ask, what percentage of students who scored 600 in 2009 scored at or below a 650 in 2010?

In this case, that turns out to be 75%. In other words, a score of 650 is at the 75\textsuperscript{th} percentile.
It all starts from the individual student

<table>
<thead>
<tr>
<th>Advanced</th>
<th>Proficient</th>
<th>Part Proficient</th>
<th>Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 4 2006</th>
<th>Grade 5 2007</th>
<th>Grade 6 2008</th>
<th>Grade 7 2009</th>
<th>Next Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale Score</td>
<td>Achievement</td>
<td>Level</td>
<td>Percentiles</td>
<td></td>
</tr>
<tr>
<td>Profficient</td>
<td>High</td>
<td>65th - 99th</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part Proficient</td>
<td>Typical</td>
<td>35th - 65th</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1st - 34th</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Growth Percentile</th>
<th>Growth Level</th>
<th>Achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>46</td>
<td>Typical</td>
<td>Typical</td>
</tr>
<tr>
<td>85</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

Math

Achievement

Growth
14 States currently using or starting to use SGP model

- Colorado, Utah, Nevada, Washington, Hawaii
- Maine, New Hampshire, Massachusetts, Rhode Island, New York
- Indiana, Virginia, West Virginia, Georgia, Arkansas
How much growth?

• Whatever the approach to growth, it is important to identify an appropriate growth standard. What is ‘good enough’?
• Consider not only what is, but what should be.
Neither growth nor status provides a complete picture of school performance, but together provide a much richer view.

**Status/Growth Combinations**

- High Status
  - High/Low
  - High/High
- Low Status
  - Low/Low
  - Low/High

- Low Growth
- High Growth
Most states in which we have been working have determined that individual student targets must be created, evaluated, and reported

- Several states decided to establish individual student targets for students currently below proficient to reach proficient in 3 years or less or by 8th grade (whichever is first), while proficient/advanced students stay above proficient
- The target is based on a defined and meaningful criterion (proficient) and can be used in the aggregate to establish school and subgroup targets

These are just suggested targets, we can easily consider other targets...
Equity refers to specific features designed to track and promote progress of the lowest performing students.

The committees proposed to define this group collectively as all students in a school not currently meeting proficiency standards and establish the target outcome as achieving proficiency.

However, all indicators will be reported for all relevant demographic and educational groups!
Equity – Design

• Proposed approach:
  o Compute growth twice: once for the whole school and again for students below proficient

• This will place a strong focus on equity by:
  o Separately reporting the progress of low performing students (prevents masking in summary data)
  o Rewards schools making the most progress with low performing students and penalizes schools making the least progress
The next step in designing the accountability system is to consider how the indicators will come together to yield information about school effectiveness. A single outcome (score/rating) will be produced for each school.
Recommended Overall Performance Levels

- Exemplary/Exceeding Expectations
- Satisfactory/Meeting Expectations
- Approaching/Partially Meeting Expectations
- Priority Improvement/Not Meeting Expectations
Exemplary/Exceeding Expectations: Schools in this category, which is reserved for schools considered models of performance, have demonstrated high growth in all applicable content areas, have average to high levels of achievement (proficiency rates), and have high performance on graduation rates and other readiness indicators (if applicable).

Satisfactory/Meeting Expectations: Schools in this category have demonstrated either high levels of growth or high levels of achievement in all content areas and are meeting state targets for readiness indicators.
Approaching/Partially Meeting Expectations: Schools in this category have demonstrated either acceptable levels of growth or acceptable levels of achievement in some, but not all content areas. Schools in the “approaching” category may demonstrate average or lower performance on graduation or other readiness indicators.

Priority Improvement/Not Meeting Expectations: This category is reserved for schools with unacceptable performance on many or most indicators. Schools in the priority improvement category typically have low levels of achievement in all content areas and demonstrate low to average growth in the relevant content areas and fall short of expectations on graduation and other readiness indicators (if applicable).
While the committee recommended a single overall reporting category, they also recommended that the following seven major indicators are always reported, to the extent applicable along with the overall category:

- Mathematics achievement
- Mathematics growth
- Reading achievement
- Reading growth
- Science achievement
- Writing achievement
- Readiness
Recommendations for Consequences and Supports

• Consequences should be framed according to necessary supports
• The system should produce performance designation in multiple levels that are linked to increasing levels of required support
• The proposed legislation outlines a series of consequences and supports tied to school performance
• The Advisory Committee has proposed an approach for designing a system for improving Wyoming’s capacity for improving support and development
Each of the following levels must be addressed in a comprehensive support system:

- Support/intervention for low performing students
- Support/mentoring for teachers needing to improve
  - Induction for new teachers and leaders
- Support/mentoring for school leaders
- Capacity building for schools and districts with lower than acceptable levels of achievement or growth
- Capacity building for the state as a whole to support continuous improvement
- The role of institutions of higher education in building capacity and preparation especially in terms of P-16 coordination
Combining indicators has the advantage of being clear to stakeholders and can guard against potentially irresponsible attempts to produce a summary outcome.

However, more detailed information is needed to inform decisions about supports and program improvement.

For this reason it is important to develop a reporting system that equips educators, leaders, and stakeholders with ample information to support a variety of uses.
At a minimum we recommend reporting information by content area, by grade, and by subgroup for:
- Mathematics achievement
- Mathematics growth
- Reading achievement
- Reading growth
- Science achievement
- Readiness

The figure at right shows a small ‘slice’ of the possible reporting levels.
Reporting

- In general, reports should be accessible to stakeholders to ensure that those closest to the classroom have the information needed to inform instructional decisions.
- Reports should be accompanied by resources to help users interpret and use results.
- We recommend Wyoming explore innovative reporting practices that draw on dynamic reporting technology (e.g. interactive data tables) and data visualization (e.g. graphs and plots).
The Colorado system provides a good example of the possibilities.
• Content standards should be credibly aligned to college and career ready outcomes
• Content standards should be well articulated across grades to facilitate meaningful measures of growth
• If Wyoming wants to be compared (favorably) with other states, we need a basis for comparison
• Therefore, the Select and Advisory Committees recommended that Wyoming adopt and implement the Common Core State Standards as soon as possible
Implications for Assessment

- **Alignment**: Assessment must be credibly aligned to CCSS or other rigorous standards
- **Rigorous**: Assessment must tap rigorous content and process and sufficient depth of knowledge. Most likely requires open-response items
- **Reliable**: Must be reliable throughout the score scale
- **Scaling and Linking**: Must avoid ceiling and floor effects and, most importantly, scores must be validly linked across years
- **Signal**: The state assessment should serve as a signal for the types of activities we would like to see in classrooms and on local assessments
This section is an introduction to the very complex challenges associated with designing an educator evaluation system that includes measures of student academic performance.

We intend for the framework to identify the many issues and decisions that must be addressed.
A comprehensive and defensible system incorporates multiple measures that go beyond student performance on state tests.

These may include some or all of the following:

- Direct observations of educators by principals or peers
- Student surveys
- Parent surveys
- Analysis of artifacts (e.g. student work, instructional activities, lesson plans etc.)
It sounds so easy and sensible on first blush, but it’s not!

“Tested” subjects and grades: Courses for which there is a state test and a prior state test in the same subject area (e.g., grades 4-8 reading and math)

“Non-tested” subjects and grades: Every grade and course not included in the “tested” category

Both are incredibly challenging to do well, but the non-tested subjects and grades issue is even more challenging
Isolating educator contribution on academic gains is very challenging

Important to engage in research to:

- Strengthen the association between instruction and outcomes
- Show ‘proof of concept’
- Rule out rival explanations
Define the type and manner of reported results.

Establish performance levels that must be produced and the purposes for which they will be used.

- Tension between reporting high-level results that are more reliable and the desire to report more nuanced but less precise outcomes for multiple indicators.

Combine indicators and setting performance thresholds.
Educator Evaluation Recommendations

- Convene a broad-based group of educators and other key stakeholders (e.g., parents, students, business) to work through the multitude of very complex value and design decisions
- Learn from other states
- Focus on improving the majority of educators not simply eliminating the lowest 5%
- Go slow! Any system needs to be piloted and carefully evaluated prior to implementation
Senate File 70 recommends considering end of course exams (EOC) to potentially replace the Body of Evidence (BOE) System

However, the legislation is silent on how this should occur

There are many decisions to be made when implementing an EOC testing system

There are many additional considerations when such a system is to be used for student accountability decisions
Student Accountability Considerations

- Definition of a Wyoming graduate
- Knowledge, skills, and dispositions
- Accountability Decisions
- Assessment system
- Support and Interventions
A Design Committee Must Address At Least...

- What framework or approach will be used to organize the EOC exams?
  - How will scores on the various exams be translated into graduation decisions?
- Which courses will include a state EOC exam?
- Which standards will the tests be designed to measure?
- What are the participation rules for any or all of the exams?
- At what level should the passing scores be set?
- What consequences will be associated with the results?
- Will retesting be included? How many opportunities?
- Can other sources of evidence replace EOC scores?
- Will there be an appeal process for students not meeting graduation standards on the testing system?
- How will the system address issues of student mobility?
• Convene a broad-based group of educators and other key stakeholders to work through the issues and decisions raised on the previous slides

• Recognizing that there will likely NOT be EOC exams in all relevant content areas, we recommend situating the EOC system within a larger framework for graduation accountability

• For example, EOC exams could be required components of districts’ BOE systems, but districts would still be required to evaluate students on content areas not tests by EOC exams
Very broadly, for both the school and educator evaluation system, the process for implementation involves the steps shown at right.

In our work, we’ve focused on the first 4 components for the school accountability system.

However, there are:
- unresolved decisions and
- a level of detail needed for implementation that goes beyond what is addressed in the framework.
Decisions and Business Rules

- As the state works toward implementation there are a number of key decisions that will need to be made to finalize the model. For example:
  - Select and implement analytic approach to growth – who will generate and evaluate the results? Is an external contract needed?
  - Setting performance levels for indicators - how much is ‘good enough’?
  - How to ‘weight’ components - should reading count more than science?
- Details matter! A number of business rules will need to be developed, such as:
  - Minimum n-size to compute results
  - What to do about missing data
  - Handling schools of different types and grade configurations
- These decisions are key to the success of the model and should be carefully vetted and tested before roll-out
Analyze Impact

- Review distribution of outcomes for all proposed reporting units and aggregated to various summary levels.
- Special attention should be given to:
  - Examining results based on differences in student populations
  - Variability of outcomes across grades, content areas, and years
- Results should be investigated for reasonableness and compared to any credible existing information
  - For example, are national board certified teachers rated higher or lower than other teachers?
- Use this information to go back and refine decisions about indicators and design.
  - For example, if the results show that high poverty schools cannot earn favorable scores or that nearly all teachers attain the same performance level, the system isn’t producing credible results and the design must be revisited.
Piloting the System

- The school accountability component of this framework shall be pilot tested with data from the 2011-2012 school year
- This shall serve as an opportunity to determine how the system works and its likely impact
- Results should be delivered to the Select Committee by September 2012
Once the system is ‘launched’ the work isn’t done!

Implementation plan should include a process for ongoing monitoring, evaluation, and support.

Investigate the claims and assumptions in the theory of action. For example:

- Are educators and leaders using the information to improve practice?
- Are incentives effective?
- Are remediation and support strategies effective in improving outcomes?
- Are more students enrolling and succeeding in college?

Again, use outcomes to refine system design and/or supports in order to maximize effectiveness.
Questions?