Wyoming's Statewide Assessment System: Recommendations from the Wyoming Assessment Task Force

Written on behalf of the

Wyoming Assessment Task Force

for the

Wyoming State Board of Education

By:

Joseph Martineau, Ph.D. and Scott Marion, Ph.D.
National Center for the Improvement in Educational Assessment

October 12, 2015

WYOMING ASSESSMENT TASK FORCE MEMBERS

Stephanie Czarobski
Middle School Teacher, Hot Springs #1

Sharla Dowding
HS Teacher, Converse County #2

Christopher Dresang
Principal, Natrona County #1

Kim Ferguson
Business Representative, Sheridan

JoAnne Flanagan
Assistant Superintendent, Fremont #25

Molly Foster
Parent Representative, Gillette

Crystal Graf
Teacher, Albany County #1

Cindy Gulisano
Curriculum Director, Goshen County #1

Shannon Harris
Principal, Natrona County #1

Cassie Hetzel
Business Representative, Thermopolis

Ellen Kappus
English Language Learner Teacher, Teton County #1

Wanda Maloney
Curriculum Director, Sweetwater County #1

Kevin Mitchell
Superintendent, Park County #1

Anne Ochs
School Board Member, Campbell County #1

Audrey Kleinsasser, Ph.D.
Professor, University of Wyoming

Mary Charles Pryor
Parent Representative, Park County #6

Kevin Roberts, Ph.D.
President, Wyoming Catholic College

Sue Stevens
Principal, Laramie County #2

Bryon Stutzman
Superintendent, Crook County #1

Sonya Tysdal
K-12 Math Facilitator, Weston County #1

Kathy Vetter
President, Wyoming Education Association, Cheyenne

Rebecca Weston
Teacher, Uinta County #6

Nicole Novotny Wonka
School Board Member, Johnson County #1

Marty Wood
HS Principal, Niobrara County #1

Ex Officio Members:
Brent Young
Deputy Superintendent, WDE

Deb Lindsey
Assessment Director, WDE

Paige Fenton Hughes
Interim Executive Director, State Board of Education

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Wyoming Legislature directed the State Board of Education to convene the Wyoming Assessment Task Force to evaluate Wyoming's current state assessment system and make recommendations for its future. The twenty-four task force members[1] met seven times between June 1 and October 1, 2015 to deliberate over many technical, policy, and practical issues associated with implementing an improved assessment system. This report presents the results of those deliberations in the form of recommendations to the Wyoming State Board of Education and the Wyoming State Legislature.

A Comprehensive Assessment System

The Task Force identified key challenges associated with the current assessment system and articulated its goals and intended uses of a new system. Some of the main concerns articulated were the lack of stability of Wyoming's state summative assessment, the incoherence among the multiple required assessments, and the lack of quality of information provided to improve educational programs. To overcome these key challenges, the Task Force focused on important uses they would like to see supported by high quality assessments. These uses included having data to help make instructional decisions for students, providing meaningful information to parents, having results that contribute to school improvement efforts, and serving as meaningful indicators in a school accountability system.

One of the most important discussions and a constant theme throughout the Task Force deliberations was the value of a comprehensive assessment system that coherently connects the results of various assessments intentionally designed to serve multiple uses and multiple stakeholders. However, the Task Force recognized that creating an assessment system that links the results of local and state assessments would challenge issues of local control. Therefore, the Task Force focused its recommendations for a comprehensive system around the summative and interim assessment components. Two of the most important sets of recommendations related to comprehensive assessment systems consider high school assessments and interim assessment programs.

High School Assessments

The Task Force recommended that the state-required standards-based summative assessments extend continuously from grade 3 through grade 10. This is in contrast to the current system that requires assessments tied to one set of standards in grades 3-8 (Wyoming State Content Standards) but to a different set of content standards in high school (ACT). The Task Force felt strongly that being able to continuously and coherently measure student growth through grade 10 would provide rich data to high schools that was previously unavailable. While the Task Force acknowledged attempts by the state to document growth among the ACT high school assessments, they noted that because the ACT was not built to measure the Wyoming standards, these assessments provided little useful information about how well schools had helped students learn those standards.

[1] There were 26 original members, but two members resigned during the course of the project due to other commitments.

The Task Force also recommended including the 10th grade assessment results as part of Hathaway scholarship eligibility requirements, but acknowledged that the mechanics of how to accomplish this goal should be left to the Hathaway Advisory Committee. The Task Force emphasized that including the 10th grade standards-based assessment results as part of Hathaway would help motivate students to take the test seriously and connect at least part of the eligibility to student performance on the Wyoming content standards.

A related Task Force recommendation was to provide students opportunities to pursue specialized areas of study in grades 11 and 12 for college or career readiness. Having a single statewide summative assessment at the end of 11th grade, as is currently required, works counter to this recommendation. By moving the Wyoming standards-based high school assessments to grades 9 and 10, schools will have the opportunity (and the expectation) to help students pursue individualized pathways, whether college or career-technical preparation, that will afford students meaningful postsecondary and life opportunities. Schools should be expected to expand options such as technical certifications, dual enrollment opportunities in community/technical colleges, AP or IB courses, and career specializations ("concentrators" in Perkins terminology). The Task Force was committed to having an assessment system that facilitated a smooth transition from high school to postsecondary-education. Tying the grade 10 assessment requirements to at least the minimum expectations for initial credit-bearing courses in community colleges—while allowing students to pursue specialized options in grades 11 and 12— will allow students to hit the ground running once they graduate from high school no matter which specialized path they pursue.

The Task Force appreciated the legislature's support for ACT census testing, but members expressed concerns that given the high numbers of Wyoming students focused on career readiness, requiring essentially all students to take a college entrance examination may not be best for Wyoming's students. Therefore, the Task Force recommended requiring all Wyoming 11th grade students take *either* a college-entrance *or* career-readiness assessment (or an alternate assessment for students with the most significant cognitive disabilities). Further, the Task Force was concerned that the ACT was overly influential in high school accountability (WAEA) scores since it is the main or sole determinant of achievement, growth, and readiness in high school. Therefore, the Task Force strongly recommended using the Wyoming standards-based grade 9 and 10 assessments as the achievement, growth, and equity indicators in the high school WAEA and reserving use of the college-readiness assessment (ACT currently) as part of the readiness indicator. Making this adjustment should be left to the technical staff at WDE.

Interim Assessments

The Task Force recommended that interim assessments be procured as part of the summative assessment request for proposals (RFP) so that both are designed to measure the same learning targets, using the same item (test questions) specifications, and same item formats. The recommendation was based on the Task Force's desire to create coherence between the interim and summative systems. The current state-required interim assessment is created by a different company than the state summative system, is designed to measure different learning targets than the state summative assessment, and has a very different approach to test and question design. The Task Force was concerned that such differences introduced unnecessary incoherence into the Wyoming educational system, and resolved that having the two systems linked more coherently would provide richer data to support school improvement opportunities.

The Task Force further recommended not requiring districts to administer the interim assessment according to a specific schedule (e.g., fall/spring) or even not requiring districts to administer the state-provided interim assessments at all. Such flexibility would allow each district to determine what approach will work best in its local context. The Task Force considered multiple design options for a state-provided interim assessment. These are discussed in detail in the full report and in Appendix D. Each district would also have the option of using another (non-state provided) interim assessment, but would be responsible for paying the cost on its own.

District Assessments

The Task Force had lengthy discussions about the role of the required district assessment system in a comprehensive assessment system. While all Task Force members acknowledged the importance of the district assessments for documenting that Wyoming students have received an opportunity to learn the "basket of goods[2]," they were insistent about maintaining a firewall between the district assessments and state assessment, reporting and accountability systems. Task Force members acknowledged the work of the District Assessment Steering Committee and generally supported the recommendation that districts be expected to document that district assessments are aligned to the Wyoming state content standards. However, the Task Force also acknowledged the need for more support and professional development regarding local assessment development and use. Additionally, Task Force members were concerned about inconsistency in quality and lack of useful feedback from reviews of district assessment systems performed as part of the district accreditation process. To address these concerns, the Task Force recommended that WDE take various steps to improve the quality and utility of these reviews by improving the expertise of those conducting the assessment system reviews. Finally, the Task Force members were adamant that district assessment results remain separate from any school accountability determinations. Task Force members felt strongly that in order to tie district assessments to accountability scores, the quality would have to improve well beyond current levels and that doing so would be an immense burden on district staff without a clear benefit in terms of student achievement.

Formative Assessment

The Task Force strongly recommended that formative assessment remain under the complete control of local districts and classroom teachers. That said, the Task Force supported WDE's current professional development efforts for formative assessment and would like to see increased state support for capacity building in this critical area.

Key Summative Assessment Recommendations

Given the centrality of the summative assessment in the legislature's charge to the State Board, the Task Force made the following additional critical recommendations regarding the design and implementation of the state summative assessments:

1. The summative assessment system selected should be used in multiple states, such as in a consortium, collaborative (in which multiple states share items and other aspects of their

[2] As a result of school funding lawsuits and resulting legislation, Wyoming students must be provided an opportunity to learn "the basket of goods," defined as the knowledge and skills in the nine subject areas represented by the content standards in each of those subjects.

assessment systems) or a commercial product. The Task Force argued forcefully that such a stance would allow Wyoming to capitalize on the pooled expertise of other states, make the cost structure much more efficient than a custom state assessment, and reduce the likelihood of year-to-year changes in the assessment since any change would have to be negotiated with multiple states.

2. The assessment design must start with the design of the reporting system, capitalizing on the latest data visualization and assessment literacy techniques. The Task Force recognized the importance of high-quality, transparent, data-secure reports in making educational decisions. It recommended that a rigorous report development process be implemented to target reports to various audiences (e.g., students, parents, teachers, administrators, policymakers, and the general public) of state assessment to address each audience's critical needs while minimizing the possibility of misinterpretation.
3. Summative assessment testing time should require no more than one-percent (1%) of the school year. This translates to between 9-11 hours of testing, depending on the grade level. This limit provides enough time to allow for high-quality assessment of complex knowledge and skills *and* to restrict testing time to a reasonable level. Further, to help schools and districts minimize disruptions to daily instruction, the Task Force recommended that the Department of Education work with a group of stakeholders to provide flexibility in the amount of time devoted to each testing session in the summative assessment.
4. Summative assessment should be designed using the most up-to-date assessment design principles and should include item (question) types capable of measuring the full depth and breadth of the Wyoming state content standards. This will ensure that the assessments can provide meaningful results to key stakeholders and will serve an important signaling function for local educators in the development of their own instructional and assessment tasks. Additionally, the Task Force recommended the inclusion of writing in the English language arts assessment to ensure fully measuring the depth and breadth of the standards, to signal that the state standards on writing are important, and to improve both the learning and instruction of writing.
5. Assessment should be developed to capitalize on the advantages afforded by online administration. The Task Force recognized past problems with online assessment and provided a comprehensive set of recommendations to ensure a smooth transition. Key among these recommendations is that schools, districts, and the state be given until the spring of 2018 to implement the new state summative assessment. Such lead time is critical for a successful and smooth transition. Further, to facilitate quick return of results, assessments should be scored using automated scoring to the extent that is practical.
6. The Task Force indicated that a balance must be struck between scheduling the state summative assessment as late as possible in the instructional year and returning the results in time for use in school improvement activities. This will aid school districts in maximizing instruction time within designated school year months, while evaluating and adjusting interventions, curriculum, and programming during the late summer months. This, in essence, requires giving the test later and getting the results back sooner, which is a difficult and potentially expensive task, but essential to the assessment's instructional utility.

The Need for Policy Coherence

One of the key findings of the Task Force was the need for a policy environment that supports the development of a coherent set of Wyoming assessments. The full report offers several general and specific recommendations for how the legislature and State Board of Education can support a policy

environment conducive to sound assessment use in Wyoming. The Task Force strongly recommends that the legislature focus on creating statutes to set broad goals and articulate the appropriate intended uses of assessments (e.g., measuring student growth, for use in school accountability determinations). The legislature and the State Board should prioritize creating a coherent, comprehensive, and efficient assessment system designed to measure student learning of Wyoming content standards and to support school improvement efforts. On the other hand, the legislature should avoid drafting legislation that speaks to the specifics of assessment design (e.g., types of items to be included on the assessment) or even requiring assessments for specific purposes (e.g., requiring a 3rd grade reading assessment).

The Task Force recognizes that each time the legislature adds an assessment (e.g., ACT) or a specific requirement (e.g., multiple-choice items only), it is for well-intentioned reasons, often in response to constituent concerns. Unfortunately, while every action might be well-intentioned, it is not long before a once coherent assessment system is no longer so. In other words, good intentions can often lead to unintended negative outcomes. Finally, the legislature should never name a specific product in legislation or write statutory requirements so narrowly that only one product or vendor meets the qualifications. It is rare that the legislature possesses the specialized knowledge necessary to recommend a specific assessment product, but most importantly, naming a specific product or too narrowly defining requirements puts the state in a poor negotiating position. The report also points out several existing statutes that will need to be amended or eliminated in order for the recommendations in this report to be enacted.

Limitations of this Report

The Task Force acknowledged that there are many other assessments that comprise the Wyoming assessment system, including the Alternative Assessment on Alternate Achievement Standards, the English Language Learner Proficiency Assessment, the K-2 reading assessments, and the many assessments that make up the career and technical assessment certification and Perkins requirements. However, the Wyoming Assessment Task Force did not address these assessments in this report for two main reasons. First, the time frame for deciding on these recommendations and issuing this report was compressed. The Task Force devoted considerable effort in a short amount of time to complete this report by the deadline. Second, and more importantly, the Task Force did not feel it possessed the specialized expertise necessary to provide recommendations for many or most of these other assessments. Therefore, the Task Force recommends that WDE be charged, perhaps with legislative support, to convene small advisory groups for each of these specific assessments and issue recommendations that adhere to the general framework for comprehensive assessment systems outlined in this report.

SECTION 1: APPROPRIATE CHARACTERISTICS AND USES OF ASSESSMENT

Introduction

In 2015, the Wyoming Legislature passed Enrolled Act 87, authorizing the State Board of Education to evaluate Wyoming's current state assessment system and create the Wyoming Assessment Task Force. Specifically, Section 6 of the act authorizes:

> *The state board shall assemble a task force to assist with the assessment review and evaluation. The task force shall be comprised of representatives of small and large school districts and schools from all geographic regions of the state and shall at minimum include representatives from district and school administration, school district assessment and curriculum program administrators, elementary and secondary school teachers, school district board members, state higher education representatives, member of the Wyoming business community and parents of children enrolled in Wyoming public schools.*

The twenty-four task force members[3] met seven times between June 1 and October 1, 2015. Three of these meetings were held in person, one of which was for two full days, and the remaining four meetings were three-hour webinars. This report presents the results of the task force deliberations. Before moving to the discussion of the task force recommendations, we first present in this section of the report some critical definitions and background assessment information.

We begin by defining two broad categories of assessment use: (1) *high-stakes accountability uses* and (2) *lower-stakes instructional uses*. Stakes (or consequences) may be high for students, teachers or administrators, or schools and districts. For students, test scores may be used for making high-stakes decisions regarding grades, grade promotion, graduation, college admission, and scholarships. For educators, student test scores may formally or informally factor into periodic personnel evaluations. In addition, students, teachers and administrators are affected by high-stakes uses of test scores in school and district accountability: identification as a school or district in need of intervention often leads to required interventions intended to correct poor outcomes.

Lower-stakes instructional uses of test scores for teachers and administrators include informing moment-to-moment instruction; self-evaluation in teaching a unit and adjusting subsequent plans accordingly, evaluating one's own instructional effectiveness; and evaluating the success of a curriculum, program, or intervention.

As described above, within the *high stakes accountability* and *lower stakes formative* categories of use, there are many potential uses and there can be many uses that blur these distinctions. The multiple appropriate uses of the various types of assessment introduced below may fall into both broad categories.

[3] There were 26 original members, but two members resigned during the course of the project due to other commitments.

Types of Assessments and Appropriate Uses

While there are several possible categorizations of assessment by type, we focus on the distinction among *summative*, *interim*, and *formative* assessment[4] because of the direct relevance to the Task Force's work. We define and outline the appropriate uses of the three types of assessment below. These definitions are critical to understanding what each type of assessment can and cannot do. Appendix B provides an at-a-glance summary of the typical characteristics, appropriate uses, and examples of each type of assessment.

Formative Assessment

Formative assessment, when well-implemented, could also be called formative instruction. The purpose of formative assessment is to evaluate student understanding against key learning targets, provide targeted feedback to students, and adjust instruction on a moment-to-moment basis.

In 2006, the Council of Chief State School Officers (CCSSO) and experts on formative assessment developed a widely cited definition (Wiley, 2008):

> *Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievements of intended instructional outcomes* (p. 3).

The core of the formative assessment process is that it takes place during instruction (i.e., "in the moment") and under full control of the teacher to support student learning. Further, unless formative assessment leads to feedback to individual students to improve learning, it is not formative! This is done through diagnosing on a very frequent basis where students are in their progress toward learning goals, where gaps in knowledge and skill exist, and how to help students close those gaps. Instruction is not paused when teachers engage in formative assessment. In fact, instruction should be inseparable from formative assessment processes.

Formative assessment is not a product, but an instruction-embedded process tailored to monitoring the learning of and providing frequent targeted feedback[5] to individual students. Effective formative assessment occurs frequently, covering small units of instruction (such as part of a class period). If tasks are presented, they may be targeted to individual students or groups. There is a strong view among some scholars that because formative assessment is tailored to a classroom and to individual students that results cannot (and should not) be meaningfully aggregated or compared.

Data gathered through formative assessment have essentially no use for evaluation or accountability purposes such as student grades, educator accountability, school/district accountability, or even public reporting that could allow for inappropriate comparisons. There are at least four reasons for this:

1) if carried out appropriately, the data gathered from one unit, teacher, moment, or student will not be comparable to the next;

[4] In defining formative, interim, and summative assessment, this section borrows from three sources (Perie, Marion, & Gong, 2009; Michigan Department of Education, 2013; Wiley, 2008).
[5] See Sadler (1989).

2) students will be unlikely to participate as fully, openly, and honestly in the process if they know they are being evaluated by their teachers or peers on the basis of their responses;
3) for the same reasons, educators will be unlikely to participate as fully, openly, and honestly in the process; and
4) the nature of the formative assessment process is likely to shift (i.e., be corrupted) in such a way that it can no longer optimally inform instruction.

Because there is considerable confusion about what formative assessment is, further definition and four vignettes[6] describing formative assessment in action are provided in Appendix A to clarify the meaning using concrete ideas. The first two vignettes are also presented in condensed form in the one-page summary of formative, interim, and summative assessment in Appendix B.

Summative Assessment

Summative assessments are generally infrequent (e.g., administered only once to any given student) and cover major components of instruction such as units, semesters, courses, credits, or grade levels. They are typically given at the end of a defined period to evaluate students' performance against a set of learning targets for the instructional period. The prototypical assessment conjured by the term "summative assessments" is given in a standardized manner statewide (but can also be given nationally or districtwide) and is typically used for accountability or to otherwise inform policy. Such summative assessments are typically the least flexible of the various assessment types. Summative assessments may also be used for "testing out" of a course, diploma endorsement, graduation, high school equivalency, and college entrance. Appropriate uses of such standardized summative assessments include school and district accountability, curriculum/program evaluation, monitoring educational trends, and informing policy-makers and other stakeholders. Depending on their alignment to classroom instruction and the timing of the administration and results, summative assessments may be appropriate for grading (e.g., end-of-course exams).

Less standardized summative assessments are also found in the majority of middle- and high-school classrooms. Such assessments are typically completed near the end of a semester, credit, course, or grade level. Common examples are broad exams or projects intended to give a summary of student achievement of marking period objectives, and figure heavily in student grading. Such assessments tend to be labeled "mid-terms," "final projects," "final papers," or "final exams" in middle and high school grades. Elementary school classrooms also have similar summative assessments but these do not have a consistent label in elementary grades. Classroom summative assessments may be created by individual teachers or by staff from one or more schools or districts working together.

Summative assessments tend to require a pause in instruction for test administration. They may be controlled by a single teacher (for assessments unique to the classroom), groups of teachers working together, a school (e.g., for all sections of a given course or credit), a district (to standardize across schools), a group of districts working together, a state, a group of states, or a test vendor. The level at which test results are comparable depends on who controls the assessment. They may be comparable within a classroom, across a few classrooms, within a school, within a district, across a few districts, within a state, or across multiple states.

Assuming they are well-design, appropriate uses of such summative assessments include:

[6] Informed by Wiley (2008).

- student grading in the specific courses for which they were developed,
- evaluating and adjusting curriculum, programming, and instruction the next time the large unit of instruction is taught,
- serving as a post-test measure of student learning, and
- as indicators for educational accountability.

Interim Assessment

Many periodic standardized assessment products currently in use that are marketed as "formative," "benchmark," "diagnostic," and/or "predictive" actually belong in the interim assessment category. They are neither formative (e.g., they do not facilitate moment-to-moment targeted analysis of and feedback designed to student learning) nor summative (they do not provide a broad summary of course- or grade-level achievement tied to specific learning objectives).

Many interim assessments are commercial products and rely on fairly standardized administration procedures that provide information relative to a specific set of learning targets—although generally not tied to specific state content standards—and are designed to inform decisions at the classroom, school, and/or district level. Although infrequent, interim assessments may be controlled at the classroom level to provide information for the teacher, but unlike formative assessment, the results of interim assessments can be meaningfully aggregated and reported at a broader level. However, the adoption and timing of such interim assessments are likely to be controlled by the school district. The content and format of interim assessments is also very likely to be controlled by the test developer. Therefore, these assessments are considerably less instructionally-relevant than formative assessments in that decisions at the classroom level tend to be *ex post facto* regarding post-unit remediation needs and adjustment of instruction the next time the unit is taught.

Common assessments developed by a school or district for the purpose of measuring student achievement multiple times throughout a year may be considered interim assessments. These may include common mid-term exams and other periodic assessments such as quarterly assessments. Many educators refer to "common formative assessments," but these tend to function more like interim assessments. This is not a negative connotation because there is tremendous transformative power in having educators collaboratively examine student work.

Standardized interim assessments may be appropriate for a variety of uses, including predicting a student's likelihood of success on a large-scale summative assessment, evaluating a particular educational program or pedagogy, identifying potential gaps in a student's learning after a limited period of instruction has been completed, or measuring student learning over time.

There are three other types of interim assessments currently in use beyond the "backward looking" interim assessments described above. All are "forward-looking." One useful but less widely used type is a pre-test given before a unit of instruction to gain information about what students already know in order to adjust plans for instruction before beginning the unit (teachers may do these pre-instruction checks on a more frequent, formative basis). Such forward-looking assessments may be composed of pre-requisite content or the same content as the end-of-unit assessment. A second type of forward-looking assessment is a placement exam used to personalize course-taking according to existing knowledge and skills. Finally, a third type of forward-looking assessment is intended to

predict how a student will do on a summative assessment before completing the full unit of instruction. The usefulness of this last type of interim assessment is debatable in that it is unlikely to provide much instructionally relevant information and there is often other information available to determine who is likely to need help succeeding on the end of year summative assessment.

SECTION 2: DESIRED CHARACTERISTICS AND USES

The Task Force recognized that assessment design is always a case of optimization under constraints[7]. In other words, there may be many desirable purposes, uses, and goals for assessment, but they may be in conflict. Any given assessment can serve only a limited number of purposes well. Finally, assessments always have some type of restrictions (e.g., legislative requirements, time, and cost) that must be weighed in finalizing recommendations. Therefore, a critical early activity of the Task Force was to identify the purposes and uses for a new Wyoming assessment system.

Task Force members initially were asked to ignore constraints, and identify their highest priority purposes and goals for assessment and their desired uses of assessment data. Task Force members, working in small groups, identified their highest priority uses and then reviewed the work of other subgroups. After each subgroup's highest priority uses and purposes were reviewed, each individual panelist identified their three highest priorities. The full task force then discussed possible patterns emerging from the activity.

In general, Task Force members desired a Wyoming assessment system that is capable of serving the following broad purposes:

- Provide instructionally-useful information to teachers and students (with appropriate grain-size and timely reporting),
- Provide clear and accurate information to parents and students regarding students' achievement of and progress toward key outcomes, such as progress toward meeting grade-level standards and progress toward readiness for post-secondary education and/or career training,
- Provide meaningful information to support evaluation and enhancement of curriculum and programs, and
- Provide information to appropriately support federal and state accountability determinations.

Detailed top priority uses and characteristics that were similar were consolidated in the broad purposes listed above. In consolidating, important differences in each contributing uses/characteristics were incorporated into the consolidated description. Appendix B provides more detailed information regarding this prioritization activity.

An important outcome of this activity is that no single type of assessment (formative, interim, or summative) is applicable to all of the high-priority desired uses and characteristics and that all three types would be needed to address the various purposes and uses. In other words, to accomplish the full set of uses and characteristics, a system of assessments would be required that span the range of assessment type (formative, interim, and summative) and assessment level (classroom, district, and state). This can be accomplished by combining state and local assessments to create a coherent system and eliminating unnecessary assessment.

[7] See Braun (in press).

SECTION 3: INTENDED OUTCOMES AND ISSUES TO BE ADDRESSED

In developing recommendations for a new state summative assessment, the Task Force deliberated on issues it intended to address in three areas: state summative assessment, interim assessments, and district assessment systems. In other words, Task Force members were asked what "problems" they were trying to solve with their recommendations. What follows is a brief discussion of these issues. The bulleted statements characterize information reported by Task Force members and, in fact, each statement can start as follows: "Task Force members reported that…"

State Summative Assessment

Timing, Stability, and Comparability

- The state summative assessment is administered too early in the year to reflect a full year of instruction, and on the flip side results sometimes come too late for use in school improvement activities such as program and curriculum evaluation. The assessment needs to be administered later in the year *and* results need to be returned in time for use in school improvement activities, which is generally by the beginning of August.
- The use of state test scores for school improvement activities is tenuous because the test or the cut scores defining achievement levels on the test change too often. The state assessment needs to remain stable for many years to allow for analysis of policies, programming, and curriculum over time.
- Comparing results from Wyoming state assessment to other states is not possible because the assessment is unique to Wyoming. It is important that Wyoming be able to compare its results with other states with similar content standards to inform state and local policy.

Test Quality

- The quality and usefulness of student achievement and growth reports needs to be improved, given the high-stakes use of state test results.
- It is important that the state assessment include tasks and questions that require deep thinking from students intended to signal the kind of activities the Task Force expects Wyoming students to engage with as part of classroom instruction. Multiple-choice-only tests are inadequate in that they signal that Wyoming puts a priority on easy-to-measure knowledge and skills.

Concerns about Appropriate Use

- Educators need adequate professional development in appropriate uses of state assessment data and appropriate preparation for success on the assessment. Teachers need confidence that they can appropriately use state assessment data to improve their own practice.
- Current use of ACT goes beyond what is appropriate. The ACT is a college entrance examination that is designed to measure ACT's college readiness standards. It was not developed to measure the Wyoming state content standards. As such, it is inappropriate to use the ACT as the sole accountability assessment in high school. The use of college entrance assessment scores should be limited to the use for which it has been validated: predicting first year grade point average in postsecondary institutions.

- The use of ACT as the sole high school accountability assessment has resulted in confusion about the high school learning targets: the official Wyoming state standards or the ACT college readiness standards? Wyoming high school educators need the high school learning targets to be clear in order to appropriately focus their instruction.

Interim Assessment

The Task Force expressed concern about the incoherence between the existing state assessment and the current interim assessment product. It is important for the state and interim assessments to provide consistent information about individual students and groups of students to assure that difference seen in the results are not simply artifacts of differences between the tests in terms of format, quality, and content coverage. Put simply, Task Force members were concerned that the first questions asked when interpreting assessment results should not be: "did I even teach this or was I supposed to?"

District Assessment Systems

While Wyoming districts have been responsible for developing local assessment systems for a long time, the review of the technical quality of such assessment systems has been inconsistent over time. The following general issues with district assessments were identified:

- There are varying levels of coherence of district assessment systems with the state assessment and with interim assessments, leading to confusion in conclusions drawn from the various assessments.
- The quality of district assessment systems is inconsistent across the state.
- There is inadequate local capacity to develop and validate high-quality local assessment systems.
- The evaluation and support of the quality of local assessment systems has been inconsistent over time.

Intended Outcomes of a Comprehensive Assessment System

Based on desired characteristics and uses of assessment developed in Section 2 and on issues identified above, the Task Force developed intended outcomes of a new Wyoming Comprehensive Assessment System in several broad areas, as shown below.

Integrate Assessment and Instruction

- Prioritize the Wyoming state content standards in a transparent way so that educators clearly know what knowledge and skills will be included on the test and that the complete set of test-eligible content is feasible to teach in the allotted instructional time.
- Improve day-to-day integration of assessment with instruction by encouraging both teacher-level collaboration and ongoing professional development for teachers and leaders.
- Provide teachers and administrators with timely data on individual students' strengths and weaknesses, and their current and predicted future achievement of desirable outcomes.

Improve Student Engagement

- Assist students to become more engaged in their own education through a greater knowledge of their strengths and weaknesses and their current academic achievement by providing feedback from formative assessment as well as from interim and summative assessment. Further, students should be provided opportunities to learn to become self-assessors and to develop the skills to direct their own learning

Provide Useful Information to Parents

- Provide parents and guardians with rich information about their student's current academic achievement by providing feedback from classroom, interim, and summative assessments.

Achieve Alignment, Coherence, and Stability

- Achieve alignment of curriculum, instruction, and assessment with the officially adopted Wyoming state standards in every district to ensure that every Wyoming student is provided a high-quality opportunity to learn the "basket of goods."
- Achieve coherence of local, interim, and state assessments.
- Achieve stability of local and state assessments to allow for a single-minded focus on improving instruction rather than adapting to new assessments.

Improve Student Academic Achievement and Growth

- Better inform educational policy improvement by providing high-quality data, stable across many years, to high-level policymakers.
- Hold schools and districts appropriately accountable for better measured and more desirable student outcomes.
- Provide valid data to local educators in order to adjust programs and curriculum to target areas of weakness.

Improve the Quality of Assessment

- Improve the quality of district assessment systems.
- Expand beyond multiple-choice items to include other types of tasks on the state assessment better suited to measuring high-level knowledge and skills.
- Convey to all Wyoming education stakeholders that writing is a valuable skill that must be effectively taught and learned in Wyoming public schools.

Enhance the Grade 11 and 12 Experience

- Reserve testing time in grade 11 and 12 for individualized college entrance, work readiness, Advanced Placement (AP), and International Baccalaureate testing.
- Provide freedom for students in grades 11 and 12 to pursue individualized pathways in Career & Technical Education (CTE) including competency-based certificates (e.g., Microsoft, Cisco), college preparation programs such as AP, and dual enrollment options.

Section 4 provides an overview of the system recommended by the Task Force. Section 5 provides detailed recommendations. Sections 4 and 5 are presented separately because it is difficult to get a coherent picture of what the assessment system would look like from the various detailed recommendations.

SECTION 4: OVERVIEW OF THE RECOMMENDED WYOMING ASSESSMENT SYSTEM

Introduction and Context

Wyoming stakeholders have determined that they want an assessment system that will serve multiple purposes, including documenting Wyoming student academic achievement and growth rates as well as supporting local instructional and program evaluation needs. A thoughtfully-designed system of state, local, and classroom assessments will be necessary to achieve these goals. Such a system will yield high-quality data from all levels of the education system to support a variety of purposes. The Task Force strongly supported minimizing redundant assessments while maximizing coherence of the results. The Task Force prioritized the following broad purposes for the Wyoming Assessment System:

- Producing instructionally-useful information for teachers and students,
- Providing clear and accurate information to parents and students regarding students' achievement of and progress toward key outcomes,
- Producing meaningful and useful information for school administrators and policymakers to support evaluation and enhancement of curriculum and programs, and
- Providing appropriate information to support state and federal accountability determinations.

This section of the report describes the Task Force's recommendations for a Comprehensive Wyoming Assessment System, attempting to paint a picture of an assessment system that blends high-quality state and local assessment results to support the multiple purposes described above. Wyoming's educational system, in spite of the centralized funding model, is strongly based on local control. Therefore, the Assessment Task Force recommended an approach to assessment that supports the multitude of uses described above, but that strongly values and improves the quality of locally-generated information.

The assessment system recommended by the Task Force is comprised of statewide, standards-based summative assessments in English language arts, mathematics, and science; a set of interim assessments intentionally linked with the summative assessments; district assessments designed to ensure that students have had an opportunity to learn the "basket of goods;" and formative assessment practices controlled at the school and classroom levels. The Task Force supported employing summative and interim assessments that can accurately measure deeper levels of student thinking, but to do so as efficiently as possible so that the summative assessment does not occupy an oversized place in the overall system. The Task Force emphasized that implementation of formative assessment is exclusively a local endeavor, but welcomed expanded partnership and support from the Wyoming Department of Education to increase local assessment literacy to support high-quality practice at the local level. Finally, the Task Force recognized that the perceived and actual usefulness of any assessment system is limited by the quality of data and reporting capabilities. While the Wyoming Department of Education has made significant strides in capitalizing on modern data visualization techniques to facilitate accurate interpretation of the school accountability results (WAEA), more work is required to develop a reporting structure that enhances the utility of the results from state-provided assessments while minimizing potential misinterpretations.

Proposed Wyoming Assessment System

Because the Task Force was generally comprised of general stakeholders of Wyoming education, the Task Force provided the recommendations in this report for general academic assessments administered to the general population of students in Wyoming's public schools. Therefore, the recommendations in this report are not necessarily applicable to alternate assessments for students with significant cognitive disabilities, English language proficiency assessment, early (K-2) literacy assessment, or Wyoming's career/technical education assessments. For these specialty assessments, the Task Force recommended that the Wyoming Department of Education convene small committees of specialists to review the recommendations in this report. The purpose would be to identify recommendations in this report that should apply to the specialty assessments, those that should not apply, those that should be modified, and additional recommendations that should apply to specialty assessments. To assure the most coherent system possible, the small committees should attempt to depart as little as possible from the recommendations in this report.

The Wyoming Assessment Task Force recommended designing and implementing an assessment system that relies on local assessment results to provide rich information to support instructional and evaluative decisions (such as curriculum and program evaluation), while relying on state summative assessments to support accountability decisions. This is done by focusing on improving assessment practice and the quality of data produced by four main assessment system components:

1. Classroom formative assessment practices designed and implemented by teachers to inform moment-to-moment monitoring of student learning and allow for immediate adjustment of instruction, and to provide high-quality feedback to engage students in monitoring and furthering their own learning.
2. District assessment system used to document students' opportunities to learn the "basket of goods."
3. State-supported interim assessments in English language arts and mathematics are designed to provide checks on student performance a few times during the school year and/or provide feedback on how well students have learned key clusters of academic knowledge and skills.
4. State, standards-based summative assessments in grades 3-10 designed to support school (and perhaps district) accountability systems, serve program evaluation needs at local, regional, and state levels, and to audit local assessment results.
5. State-provided college entrance *or* career readiness assessments in grade 11 designed to give students choices matching individual goals for pursuing post-secondary education at institutes of higher education, career training, or technical education.

For the first four categories of assessments to work coherently in Wyoming, they must, at a minimum, be designed to measure student learning of the appropriate Wyoming content standards.

Classroom Formative Assessment

The Wyoming Assessment Task Force acknowledged the critical importance of classroom formative assessment practices for improving student learning, but emphatically argued that it should remain relatively silent on recommendations in that area. Task Force members noted that formative assessment is the purview of districts (actually, schools and classrooms) and, for the most part, should not be considered a state program. The Task Force, however, acknowledged that it would

make sense for the state and districts to partner in providing high-quality professional development to support improvements in local formative assessment practices.

District Assessment System

In response to State Supreme Court decisions and legislative mandates, Wyoming requires districts to document that students have had an opportunity to learn the "basket of goods," defined as the content standards in nine subject areas. A comprehensive assessment system must address how the state will monitor student learning of this basket of goods. The combination of district assessment systems and state summative assessments in English language arts, mathematics, and science are required to meet these mandates. The legislature and State Board of Education have had quality assurance requirements for district assessment systems in place for more than 15 years. In spite of this history, the Task Force members expressed concern about the effectiveness of these requirements and the utility of the feedback and supports provided to districts for improving their assessment systems.

The Task Force noted that district assessments play multiple roles, contingent upon their intended uses. Districts have designed a variety of approaches to meet local needs and work within the constraints of capacity. District summative assessments are expected to be aligned to the relevant Wyoming content standards in the given grade level or course, but the specific assessment approach may vary considerably across districts. For example, districts may choose to use single, large-scale tests at the end of a grade or grade span or they may rely on multiple unit-based assessments tied to the applicable Wyoming content standards. In another example, district assessments may serve both an auditing function for individual teachers' understanding of their students' learning, and a signaling function of the kinds of knowledge and skills that should be prioritized in daily instruction and classroom assessment.

Even so, Task Force members expressed frustration that in spite of the mandate that districts design and implement local assessment systems in at least nine content areas, there was little clarity regarding the state-required purposes and intended uses of these systems. As explained previously, assessments work best when designed for a specific use (in fact, we argue that is the only way that assessments are useful) and if the intended purposes of the district assessment systems are vague, the utility of the results will be limited. Many districts have designed assessment systems that meet local needs. This may be appropriate, but it makes it difficult to outline specific quality criteria if the assessments across districts are designed for considerably different purposes. The Task Force strongly recommended having common requirements of assessment quality, but supported local flexibility on specific assessment designs and uses.

There was interest among some legislators, as expressed in Enrolled Act 87, in using district or other local assessments for state and/or federal accountability purposes while reducing the amount of statewide summative testing. However, the Task Force declined to move in that direction at this time. Task Force members were concerned that meeting the quality requirements for district assessments to serve accountability uses could overwhelm district personnel. After examining the data and reviewing the existing literature, the Wyoming Assessment Task Force recommended that, at the current time, district assessment results should not be used as part of school accountability determinations. The Task Force acknowledged that such a stance may relegate district assessment results to a lower status than the state assessment. At the same time, Task Force members were

concerned that it was not practically feasible in the short term to dramatically improve the quality of district assessments so they could be used as accountability indicators fairly across the state.

However, the Task Force recognized the need for improving the quality of district assessments through the use of multiple strategies including increasing the assessment expertise of those who reviewed district assessments as part of district accreditation processes and to foster local assessment expertise through state support of district assessment consortia.

Interim Assessments

The Wyoming State Legislature has required and paid for the implementation of a common interim assessment program for all Wyoming school districts. The State supported two administrations of the interim assessment each year—fall and spring—but many districts paid to support winter administration as well. Many district leaders found value in the commercially-selected interim assessment products, using them for a variety of purposes including documenting within-year growth and identifying students in need of remediation. On the other hand, the Task Force members expressed some concern that it was difficult to coherently interpret the results of the interim assessments in light of the summative assessment expectations because the two were designed to measure different learning targets and to do so in different ways (e.g., different item formats).

The Wyoming Assessment Task Force's major recommendation on the interim assessment was that the State should require the development of an interim assessment system based on the same assessment framework and tied to the same learning targets as the state required summative assessment. Districts could optionally administer the state-provided interim assessments, and would have local control over how they would administer the tests and use the results. Additionally, districts would have the option of purchasing/developing an interim assessment system not tied to the state assessment system, but such districts would be responsible for the costs.

In a critically-important move to help inform WDE's procurement process the Task Force made additional recommendations regarding the specific interim assessment design. A key consideration for interim assessment design is whether the assessments are "forward-looking," "backward-looking," or a "mini summative assessment" design. Forward-looking assessments are provided prior to instruction to gain an understanding of student readiness for learning new concepts and skills. Conversely, backward-looking assessments are those that are designed to help educators and students know how well students learned material that had been taught, generally recently. They can be designed as modules to evaluate student learning of discrete aspects of grade level content (e.g., numbers and operations).

Mini-summative designs are those in which each instance of the interim assessment (2, 3, or 4 or more times each year) is designed to replicate the summative assessment blueprint[8]. Because they are intended to be on the same scale (often a vertical score scale), such mini-summative interim assessment designs are often used for evaluating student growth throughout the year. On the other

[8] A test blueprint is generally in the form of a matrix where the content categories (e.g., standards, objectives) to be tested are represented on one axis and the level of cognitive demand (in the form of process skills or depth of knowledge) required is represented on the other axis. The cells then document the number of test items or score points for each content category by each level of cognitive demand that can be expected to appear on the test.

hand, there is a substantial body of research indicating that vertical scales are not necessary for documenting student progress. Many Task Force members indicated that it is important for interim assessments to "predict" end-of-year summative assessment performance, and thought that the mini-summative designs were the best way to meet this need. However, the technical facilitators (Martineau and Marion) pointed out that it would be relatively easy to create prediction equations for almost any pair of reasonably well correlated assessments.

Task Force members were intrigued by having a set of modules, tied to key aspects of grade-level content, as the potential interim assessment design. In order to keep costs in check, the modules would be focused on a limited number of the major concepts of the discipline (e.g., 3-5 modules) and designed so that districts could administer the modules when and where they fit best within each district's curriculum. The modules offer promise for providing feedback to educators and students on more narrowly-specified sets of knowledge and skills than the broader set of content associated with a mini-summative design. See Appendix D for a pictorial representation and detailed explanation of the different designs. Such modules could also effectively serve an auditing function for district assessments, which should be designed to measure similar knowledge and skills. Finally, a modular approach to interim assessment offers the potential for simultaneously reducing the time associated with the summative assessment and generating more instructionally-useful information for educators, because it could eliminate the need for "subscores" on the summative assessment. Because this possibility may seem counterintuitive, additional explanation is provided in Appendix E.

In order to achieve this goal, it may be necessary to customize an existing assessment to some degree. Given the recommendations that follow about not using a custom-designed large-scale summative assessment in Wyoming, existing assessments would need to be capable of a degree of customization without the loss of the benefits that an existing assessment offers. This will likely be possible by 2018. Another potential benefit that such an approach offers is reducing the amount of student time devoted to state summative assessments.

The Task Force also discussed types of questions that should appear on the interim assessments. The members knew that using selected-response items (e.g., multiple-choice) to populate the interim assessments would allow for instant reporting and would keep costs down. However, the Task Force recommended that interim assessment questions reflect the types of questions found on the large-scale summative assessment designed to probe students' deep understanding of critical content and skills. At the same time, the Task Force also strongly recommended that the interim assessment scores must be returned to schools within one week of completing the test. This tradeoff would allow for questions that might take a little longer to score than instant multiple-choice items, but might not allow for the full array of extended-response tasks.

Finally, the Task Force issued recommendations around existing and future requirements associated with the interim assessments. The Task Force recommended that requiring districts to implement assessments in order to conduct evaluations of specific programs could easily become unwieldy and result in a hodgepodge of assessments instead of the coherent system that the Task Force is promoting. The Bridges program is a case in point. This intervention program is designed to provide supplemental educational opportunities to traditional educationally-disadvantaged student groups or other students struggling with grade-level knowledge and skills. These opportunities are often provided during the summer, but may be offered after school and on weekends during the regular school year. While well-meaning, the notion of requiring the administration of interim assessments

early in the school year to help evaluate the Bridges program has the effect of making the "state" assessment a three times per year event and, most importantly, may miss important aspects of the Bridges program. It is generally assumed that a fall interim assessment allows for calculation of change in students' scores from spring to fall after experiencing the Bridges summer school. However, as noted above, Bridges funds are commonly used to implement instructional interventions other than summer school, such as weekend programs throughout the school year, meaning that a fall interim test for Bridges evaluation may be limited. It is beyond the scope of this report to discuss alternative evaluation designs for the Bridges program. Rather, the Task Force emphasized that the legislature and other policy bodies should avoid requiring additional assessments without carefully thinking about how such assessments fit within a comprehensive assessment system.

State Summative Assessment

The Task Force indicated that the state summative assessment must comply with state and federal laws, industry best practices, and professional standards. Further, the assessment should be designed using a principled-assessment design approach. The Task Force strongly recommended that in content areas where it is possible, the state summative assessment selected for Wyoming should be used in at least one other state (preferably many states). There are two reasons for this: to allow for comparison of Wyoming educational outcomes to other states and to encourage a stable state summative assessment over time. In other words, changes to the state summative assessment should be minimized by requiring negotiation with other states and/or a vendor in order to make changes to the assessment system.

The Task Force recommended limiting testing time for state-required summative assessments to no more than *one percent* of the Wyoming required instructional hours for any grade. This translates to a limit of 9, 10.5, and 11 hours of testing time for elementary, middle, and high school grades, respectively. The Task Force did *not* recommend that the full limit of hours be used, only that this should be the maximum allowable, while allowing the time to include questions measuring high-level knowledge and skills on the assessment. State tests are not timed in Wyoming so the Task Force recommended that required testing time be estimated as the amount of time needed for at least 85 percent of students to complete testing. These estimates will improve in accuracy over time.

The Task Force recommended that state, standards-based summative assessments be required in English language arts (including writing) and mathematics in grades 3-10 as well as in science in at least one grades each in elementary, middle s, and high school. These assessments must be designed to fully measure the Wyoming content standards and to assess whether students are on track towards college and career ready outcomes. The Task Force recommended that the grade 10 state summative assessment count as part of the Hathaway scholarship[9] determinations to explicitly tie the scholarship to the official Wyoming content standards and to assure adequate student motivation[10].

[9] The Hathaway scholarship is a program where Wyoming high school students who complete a required set of courses, have a certain grade point average (GPA), and achieve the required composite score on the ACT. There are various levels of the scholarship award ranging from $1640 to $840 per semester (for 2015 graduates) depending on the specific GPA and ACT scores. It was beyond the scope of the Task Force's work to recommend exactly how the grade 10 scores may be included as part of the Hathaway determination, but the Task Force was confident that this was not an insurmountable problem.

The Task Force pointed out that it is not appropriate to include all of the Wyoming high school standards on a test given in grade 10, because students still have at least two more years of school remaining. Therefore, the Task Force recommended having the Wyoming Department of Education convene a standards review committee to determine which of the state high school content standards are eligible for testing by the end of 10th grade. Because grades 11 and 12 remain important, the Task Force recommends that district assessment systems be required to cover the Wyoming high school content standards that do not appear on the state summative assessment. The Task Force acknowledged that this should be relatively easy to accomplish for English language arts, but mathematics could be more challenging. For example, should the 9th grade assessment focus largely on algebra 1 standards while the 10th grade assessment target geometry knowledge and skills? This sounds intuitively sensible, but is not without its challenges. The Task Force noted that such prioritization could occur easily with a custom assessment program, but would have to be negotiated if the state procures a consortium, collaborative, or other existing assessment system.

The Task Force also recommended that the state continue to fund in-school administration of a college entrance examination in grade 11. However, the Task Force argued that career readiness was as important as or more important than college readiness in many parts of Wyoming. Therefore, the Task Force recommended requiring all students to participate in *either* a college entrance examination *or* an analogous career readiness assessment. The provision of an in-school opportunity for college entrance or career readiness testing (rather than a traditional Saturday administration) is intended to maximize the number of students thinking about post-secondary opportunities.

The recommendations to have the last required state standards-based summative assessment at the end of 10th grade was designed to encourage students to specialize during their last two years of high school. The lack of state mandated standards-based testing in grade 11 and 12 was designed to help junior and senior students focus on highly individualized pathways through either college preparation (e.g., through Advanced Placement (AP), dual enrollment, or other programs) or specific career/technical areas where students may become "concentrators." It also facilitates the transition from high school into college or career training by strengthening the connection between grades 11-12 and post-secondary education or training.

In order to improve reporting timelines for use in school improvement and other evaluation activities, the Task Force recommended administering state summative assessments online except in isolated situations with emergent needs for paper and pencil. Safeguards for assuring a successful transition to online testing are described near the end of this section of the report. The Task Force recommended administering the summative tests in a three-week window near, but not at, the end of the school year to maximize the amount of instructional time before the test, but also to assure return of results in time to support summer school improvement activities and district program evaluation needs.

The Task Force recommended that the state summative assessments serve both an auditing function for district assessment results and a signaling function of the kinds of knowledge and skill that should be prioritized in district assessments (e.g., deeper levels of thinking).

---

[10] The Task Force does not have a specific recommendation for how the grade 10 assessment results should be incorporated into the Hathaway determination, but suggests that the legislature direct the Hathaway Advisory Committee investigate how best to accomplish this goal.

However, the task force was concerned that including too many performance or other extended-response tasks on the state summative assessment may lead to unacceptable testing times. Therefore, the Task Force strongly recommended that the state summative assessment include only the number of such test questions necessary to both signal the types of assessment tasks the state would like to see on classroom and district assessments and ensure that the state assessments can provide information about achievement on the full depth of the Wyoming state content standards.

Supports and Conditions

To improve fidelity of implementation at the classroom, school, district, and state levels, the Task Force noted that certain supports are critical.

Data and Reporting Systems

The Task Force recommended the use of a comprehensive assessment system to maximize the coherence of information produced from various assessment tools. However, without a well-designed and implemented reporting system, the hopes for a comprehensive assessment system will fall well short. The world of data visualization has opened up exciting new possibilities for placing useable information in the hands of users in ways they can easily understand. Score reports are the only ways assessment designers communicate with stakeholders, yet it is often the last thing attended to in design deliberations[11]. Therefore, the Task Force strongly recommended that Wyoming devote the resources necessary to produce a high-quality digital reporting system that capitalizes on modern data visualization techniques and facilitates accurate assessment interpretations while minimizing opportunities for misconceptions. Such a reporting system could be included in vendors' bid in response to the state assessment RFP, but the Task Force was aware that such systems would likely come from more specialized vendors. The Task Force commended WDE's efforts in reporting the results of Wyoming Accountability in Education Accountability system (WAEA), but wanted to go much further to help users understand the assessment results and potential educational implications of the scores.

Assessment Literacy

Having high-quality and intuitively useable reporting systems is a big step toward improving assessment literacy. Unfortunately, it is probably not enough. The Task Force recognized WDE's current efforts to promote formative assessment practices, but recommended expanding the state's efforts to promote assessment literacy and effective assessment. It is beyond the scope of this report to fully outline approaches to meet these goals. The Task Force recommended implementing a thoughtful approach or set of approaches to improve local assessment practices and products (e.g., classroom and district assessments).

Evaluation

Finally, the Task Force recommended that the state should contract for an ongoing evaluation of (1) the quality of the state assessment; (2) the degree to which intended outcomes are being achieved; (3) the degree to which anticipated and unintended consequences have been observed and

[11] Attributed to Ron Hambleton.

minimized (for the unintended, negative consequences); and (4) after three to five years, a summary report including potential improvements to the system to address any issues identified.

Ensuring a Successful Transition

The Task Force recommended a multi-year transition strategy to ensure a successful transition to online state summative assessment and high-quality interim assessment systems. Allowing enough time to act on these recommendations is critical to assuring that the transition is successful. The first all-online administration of the state summative assessment should take place in the spring of 2018 and the transition must be smooth. The Task Force recommends a comprehensive, detailed set of safeguards to assure a smooth transition, presented in subsequent sections.

SECTION 5: DETAILED DESIGN AND TECHNICAL RECOMMENDATIONS

Before presenting the specific, detailed design and technical recommendations, it is important to note that Task Force members and State Board of Education members wanted to ensure that the recommendations in this report did not unreasonably limit the number of *potential* products that could qualify if these recommendations are enacted. To address those concerns, *potential* qualifying vendors and products are listed in Appendix F.

I. Classroom Formative Assessment

The Wyoming Assessment Task Force acknowledged the critical importance of classroom formative assessment practices for improving student learning, but emphatically argued that other than briefly discussing formative assessment in this report, the Task Force should remain relatively silent on the issue. Task Force members noted formative assessment is the purview of districts (actually, schools and classrooms) and for the most part should not be part of the "state" comprehensive assessment system. The Task Force, however, acknowledged that it would make sense for the state and districts (perhaps organized regionally) to partner in providing high-quality professional development to support high-quality local formative assessment practices.

II. District Assessment System

The major issues identified with district assessment systems by the Task Force include uneven quality, uneven coherence with state assessment, and practical challenges for districts to design and implement high quality assessment systems. The following recommendations attempted to address these concerns:

A. To improve quality and assure consistency of reviews, the state should contract with one or more qualified professionals to perform audits of district assessment systems as a part of the accreditation process.
B. The state should incentivize and/or support collaborative efforts among districts to improve the quality of locally-developed assessment tasks and the quality of data use for informing educational decisions. This could include hosting for educators to obtain access to intact assessments, banks of high-quality tasks and test questions, and appropriate professional development on using the resources.

The Task Force recommended NOT using the district assessment results as an indicator in WAEA at this time because considerable improvements in district assessment systems would be required to support high-stakes use and there is concern that districts do not have the time and capacity to meet such requirements at this time.

III. State-Provided Interim Assessment

A. Governing Principles

The Task Force recommended that the state support an interim assessment system tied directly to the summative assessment to encourage consistency across the state and coherence with the

accountability performance targets. The use of interim assessments should be governed by the following principles:

1. To reduce required testing time and to tailor to specific uses, districts should not be required to administer any interim assessments, but may choose to use interim assessments as the district sees fit to support identified local uses.
2. Districts choosing to use the state-provided interim assessment would not be responsible for the cost of the assessment. Districts choosing to administer a different interim assessment would do so at their own expense.
3. The interim assessment supported by the state should be coherently tied to the state summative assessment. It should be constructed to measures the same content standards, and should use the same types and formats of test questions to assure a consistent experience for students and educators across state summative and interim assessment.
4. To assure coherence with the summative assessments and to achieve competitive pricing, the interim assessment should be procured as part of the summative assessment.
5. To provide an outside audit of the district and other local assessment results, interim assessments should provide a check on the big ideas associated with the grade level learning targets.

B. Two "Flavors" of Interim Assessment

The Task Force discussed two basic forms of interim assessment (see Appendix D):

1. A "mini-summative" version in which the interim assessment is a shorter version of the end-of-year state summative assessment (e.g., the interim assessment blueprint is representative of the summative assessment blueprint, but results in a shorter test[12]). This allows for monitoring students' growth within a school year on an overall content area and for predicting student performance on the end-of-year summative test.
2. A module-based version in which the blueprint of the summative assessment is broken into 3-5 subsets of content categories, and each interim assessment module measures only one subset. Each module should allow for at least two subscores to be reported within the subset. This allows for measuring achievement of mid-sized units of instruction.

The Task Force recommended that an RFP for state assessments should include both mini-summative and module-based interim assessment designs with the timing of interim assessments being left entirely to local discretion to best meet local needs. However, because of concerns about potential cost increases from providing both types of interim assessment, the Task Force indicated a preference for the module design starting with at least 3-5 modules per grade and subject if a choice of either modular or mini-summative must be made.

[12] A test blueprint is generally in the form of a matrix where the content categories (e.g., standards, objectives) to be tested are represented on one axis and the level of cognitive demand (in the form of process skills or depth of knowledge) required is represented on the other axis. The cells then document the number of test items or score points for each combination of content category and level of cognitive demand that can be expected to appear on the test.

D. Item and Task Types

The Task Force recognized the importance of the interim assessment mirroring the summative assessment as much as possible to assure that complex knowledge and skills are measured on both. The Task Force also recognized that near-immediate reporting is needed to maximize the usefulness of interim assessments. The inclusion of complex item types (see the section on *Alignment to the Wyoming State Standards* on page 31) means that human scoring may be required, which increases the time between completing an assessment and reporting. To address this conflict, the Task Force recommended the following compromise:

1. Interim assessment results should be returned no more than one week after completion of an assessment.
2. All items types used on the summative assessment should also be included in the interim assessment, so long as they do not preclude returning interim assessment results in no more than one week.

IV. State Summative Assessment

A. Governing Principles

To assure that Wyoming is able to procure a high-quality assessment, the Task Force recommended that the technical quality of the assessment should be well-documented according to research and/or best practices as referenced by some or all of the following:

- Principled assessment design (e.g., *Evidence Centered Design*[13], *Knowing What Students Know*[14])
- Universal Design for Learning[15]
- The AERA/APA/NCME *Standards*[16]
- CCSSO/ATP *Best Practices* for Statewide Assessment[17]
- Applicable state and federal law and regulation
- Federal peer review requirements

B. Avoiding an Exclusive Wyoming Assessment

In order to provide stability, cost savings, enhanced quality, and comparability of Wyoming test results to other states, the Task Force recommended the following:

1. Each content area test must be used in some form in at least one other state (preferably several other states) for the following reasons:
    - Improve technical quality through the increased capacity and expertise in collaboration among multiple states.

[13] Mislevy & Riconscente (2006).
[14] Pellegrino, Chudowsky, & Glaser (2001).
[15] Thompson, Johnstone, & Thurlow (2002).
[16] APA, AERA, & NCME (2014).
[17] CCSSO & ATP (2013).

- Facilitate comparison of results from the Wyoming assessment to results from other states.
- Reduce cost through collaboration among multiple states.
- Provide stability by requiring changes to the assessment to be negotiated with at least one other state and/or vendor.

2. To maximize market competition, the ability to meet Wyoming's needs, and negotiating power, recommendations in this section should be required only where there are at least two options available.

C. Standards-Based Assessment vs. College/Career Entrance Assessment

The Task Force recommended that a distinction be made between assessments up to grade 10 and assessment after grade 10 in order to maintain the benefits of a college entrance examination, and to provide greater freedom for juniors and seniors to pursue individualized pathways.

1. Assessments after Grade 10.
    a. Reserve grade 11 and 12 for college entrance, work skills, CTE and other certifications, Advanced Placement, and International Baccalaureate assessments. Do not add standards-based state summative assessment in grade 11 or 12.
    b. To provide schools incentives to help upper level high school students develop individualized pathways through a career and technical education program or a college preparation program, do not use grade 11 and 12 assessments for school accountability purposes other than as part of the "readiness indicator" of WAEA.
    c. Require grade 11 students to take *either* a college entrance examination or a work skills examination. This should be administered in school on a regular school day.
    d. The Department of Education should be provided with funding for a contract to provide districts with one or more resources to gather and report on students' career/college interest to facilitate local development of individualized high-school pathways.
2. Assessment in Grades 3-10
    a. Require standards-based, state summative accountability assessment in grades 3-10.
    b. The Department of Education should be provided with funding for a contract to conduct studies to develop predictive relationships between the grade 9 and 10 assessments and the college readiness and work skills assessments.
    c. To ensure both student motivation on the grade 10 assessments and alignment of the Hathaway scholarship criteria with the official Wyoming content standards, the Task Force requests that the legislature and the Hathaway Advisory Committee investigate how the grade 10 assessment might be incorporated into the criteria for Hathaway scholarship eligibility.

D. Alignment to the Wyoming State Standards

The Task Force recommended signaling the importance of complex knowledge and skills described in the Wyoming state standards through the following:

1. The grade 3-10 assessments should be aligned to the depth and breadth of Wyoming's state content standards, including complex knowledge and skills that are not easily measured.

2. The assessment should include both multiple choice items and more complex item types better suited to measuring more complex knowledge and skills (e.g., enhanced multiple choice, technology enhanced items, short constructed response, extended constructed response, performance tasks). However, the number of more complex item types included in the assessment must allow for meeting the testing time limits.
3. To avoid market restriction, vendors proposing "naked" writing tasks may still be considered "qualified bidders" assuming they meet other requirements. However, after qualified bidders have been identified, vendors proposing writing tasks that require a text (evidence)-based response should receive more points for writing than vendors proposing naked writing tasks[18].

E. Content Coverage

To ensure compliance with federal laws and to signal the importance of the core content areas of English language arts (including writing), mathematics, and science, the Task Force recommended the following:

1. Require assessment of English language arts and mathematics in every grade.
2. Require coverage of writing (as a part of English language arts) in *at least* one grade each in the elementary, middle, and high school grade spans.
    a. If it is possible to do so within the limits for testing time, include writing in each of grades 3-10.
    b. The English language arts assessment should include at least two writing samples per student to adequately measure the Wyoming writing standards.
    c. Contextualized writing tasks should be preferred over "naked" writing tasks (e.g., writing tasks should require referring to provided text, charts, and/or tables).
3. Require coverage of science in *at least* one grade each in the elementary, middle, and high-school grade spans.
    a. Wait to bring the state science assessment into compliance with the requirements of this report until new Wyoming state science standards are adopted.
    b. Include in the RFP for state assessment services a range of dates in which the contractor could reasonable expect new science standards to be adopted.
4. To clearly identify what content is eligible to appear on the grade 10 test in each content area, the following should be enacted:
    a. The Department of Education should facilitate a standards review committee with the charge of specifying which of the Wyoming content standards are expected to be taught and learned by end of grade 10.
    b. The committee should be comprised of K-12 content specialists, district curriculum directors, and higher education content specialists.
    c. After the standards review committee completes its work, the Department of Education should convene a small advisory group of educators to assist it with determining appropriate content to appear on specific grade 9 and 10 assessments in mathematics[19].

[18] This assumes a bid evaluation process in which vendor bids are first scored to determine whether they meet a threshold for qualifying to provide the state with assessment services, followed by a review of the qualifying bids for a few areas in which select vendors may receive extra credit for proposing "value added" beyond the requirements of the request for proposals (RFP).

[19] The content of English language arts is reasonably well specified by high school grade level, but mathematics content standards are not. This requires a careful parsing of the high school mathematics content standards for not only what

d. Any remaining high-school content in the Wyoming state standards should be covered in district assessment systems.

F. Testing Time

In combination with eliminating the requirement to use a state-provided interim assessment, the Task Force recommended limiting the amount of time that may be required for state summative assessment.

1. Limit actual testing time for state-required summative assessment to no more than 1% of the required instructional hours for a given grade level (based on Chapter 22 of Wyoming Department of Education rules, this translates to approximately 9, 10.5, and 11 hours of testing time for elementary, middle, and high school, respectively)[20].
2. "Actual testing time" means the time that students are actually responding to assessment tasks (not additional time used for test preparation, breaks, gathering students, logging students, or reading test instructions)[21]. Because Wyoming state assessments are not timed, "actual testing time" should be based on estimated testing time needed for 85% of students to complete the test. These estimates should be updated annually based on actual test administration.

G. Test Timing and Test Windows

In order to balance maximizing the amount of instructional time before state summative assessments and typical end-of-year school activities, and the need to receive results in time for school improvement activities, the Task Force recommended the following:

1. State testing should occur during a three- to four-week testing window which is the same for every grade, with the last allowable testing day being in the first half of May.
2. All aggregate reports (other than statewide aggregate reports) should be available by August 1 to facilitate school improvement activities (with consideration that in the first year of any new program, reports are likely to be delayed).
3. Acting within the constraints of the first bullet in this list, the Department of Education should work with a committee of stakeholders to finalize testing windows (e.g., the first and last allowable testing days each year) and to address local needs for flexibility in scheduling assessment activities[22]. If possible, start and end dates should be later to maximize instruction before assessment, but should also consider typical year-end school activities and the time needed to return data to schools in time for use in school improvement activities. This committee of stakeholders should include school and district staff with two sets of responsibilities: (1) calendaring, and (2) managing state assessment activities.

---

must be taught by the end of 10th grade, but also for what should appear on each of the grade 9 and grade 10 assessments in mathematics.

[20] Required testing time may be less than these limits.

[21] This definition of "actual testing time" is provided to avoid district-to-district variation in the time devoted to activities wrapped around actual testing.

[22] For example, allow for flexibility in length of test sessions to coincide with the length of class periods (to avoid unnecessary disruption of daily instructional activities).

H. Moving Assessment Online

The Task Force recommended that test administration be moved fully online to expedite return of assessment results and the use of data in school improvement activities. While other states generally less ready than Wyoming have successfully made the transition to line assessment, Wyoming's previous experience with statewide online assessment prompted the Task Force to recommend several safeguards to assure a smooth transition. The most important of these is that the new assessment system should be developed and implemented in a deliberative manner. If these recommendations are acted upon quickly, a new assessment system could be in place by spring of 2018. The recommended safeguards include the following:

1. Schools and districts should be notified immediately, upon acceptance by the legislature of the recommendations in this report, that they must be ready for online assessment by spring of 2018.
2. The state should contract as soon as possible for a high-quality comprehensive technology infrastructure audit for the state as a whole and for every school and district. The state audit should at a minimum cover adequacy of the state internet backbone. District audits should, at a minimum, cover adequacy of available bandwidth, stability of connections to the state backbone and/or other networks. School audits should cover adequacy of available bandwidth, stability of connections to district/state systems, adequacy of wireless school network capacity, adequacy of the number of devices capable of administering the assessment, and the adequacy of the operating systems used on those devices.
3. The state contractor should work with each school district to assist in performing the audit (including fully conducting the audit if necessary) to assure a consistent application across all districts.
4. The state contractor should produce a public report including sections for the state as a whole (including a summary of district and school reports), each district (including a summary of each school report), and each school. The report should identify specific gaps in technology infrastructure in each section of the report and identify minimum actions required to close those gaps.
5. After the full set of audit reports has been produced, it may be necessary for the legislature to consider whether there are any critical, targeted funding needs to fill the identified gaps.
6. To improve schools' confidence in the process, all appropriate state agencies that support school technology infrastructure should clearly describe how they will support preparing all schools and districts for online assessment by spring 2018.
7. At least ten months in advance of the first statewide online administration, all schools, districts, and the state contractor should conduct a simultaneous load test simulating all of Wyoming's students logging on and taking the test simultaneously to attempt to "break" the system. Any breaks or near breaks in the system as a result of the load test should be used to increase capacity in any areas necessary before the first administration.
8. A paper and pencil option should be available to address isolated emergent needs that cannot be resolved in a reasonable amount of time to allow for online testing.
9. Schools should have reasonable flexibility on scheduling testing within the test window to accommodate the use of online assessment with a limited number of devices (e.g., the length and number of test sessions for each student).
10. It should be communicated often to both parents and educators that prior to taking assessments online, students should be provided with adequate experience in the classroom using devices they will take the test on. This should include at a minimum specific focus on

navigating a screen and keyboarding. The Department of Education should gather a workgroup of educators to develop guidelines for providing adequate experience.

I. Claims to Be Supported for Individual Students

In order to support important educational decisions made by parents, students, and teachers, the Task Force recommended that the assessment must support the following claims for each individual student:

1. How each student achieves relative to Wyoming standards, including more difficult to measure, higher-level knowledge and skills.
2. How each student's year-to-year growth compares relative to peers.
3. That student achievement and growth scores are accurate across the range of student achievement, meaning that:
    a. Scores are generally free of floor or ceiling effects.
    b. Scores support claims about whether novice, typical, and advanced students are being well educated.

J. Claims to Be Supported for Classrooms, Schools, Districts, and the State

In order to support important educational decisions made by teachers, administrators, policymakers, and the public, the Task Force recommended that the assessment must support the following claims for each classroom[23], school, district, and the state:

1. The magnitude of achievement and growth gaps for key demographic groups (e.g., sex, race/ethnicity, economic disadvantage, special education, and English learners).
2. The change in achievement and growth gaps over time.
3. The percentage of Wyoming students meeting proficiency targets.
4. The percentage of Wyoming students meeting growth targets adequate to remain proficient (for already proficient students) or to achieve proficiency (for not yet proficient students) within a reasonable number of years.
5. Produces valid and reliable group reports (at the class, school, district, and state level) on strengths and weakness in both proficiency and growth in a small number of sub-areas of each content area. To the degree that these sub-scores provide different information (see Appendix E for detail on issues with sub-scores), this supports school improvement activities, post hoc evaluation of instructional practices, curriculum, and programming, and high level policies. This could be accomplished using green/yellow/red light reports that show for each group the sub-areas in which a group's achievement is better than, similar to, or worse than its overall content area achievement[24].

[23] Access to classroom-level aggregate reports should be limited to educators responsible for that classroom to protect student privacy.

[24] For example, group average subscores can be compared to overall scores within a content area to identify whether in each sub-area, the group perform better than, similar to, or worse than they did in the overall content area. Each of those group average scores could also be compared to the thresholds for the different performance levels.

K. Reporting

Without thoughtfully designed and useful reports, the quality of the assessment system is moot. To assure that investment in the quality of the assessment is returned, the Task Force recommended the following:

1. Reports must be designed to meet the needs of the following four key groups of stakeholders:
    a. Students and parents
    b. Teachers
    c. School and district leadership teams
    d. Business community, media, State School Board, State Superintendent, Joint Legislative Education Committee, Legislature at large, Governor, and general public
2. Individual student reports must be designed with stakeholder groups "a" and "b" in mind.
3. Aggregate reports (e.g., classroom and school reports) showing individual student data must be designed with stakeholder groups "b" and "c" in mind.
4. Aggregate report showing group summary data must be designed with all four groups of stakeholders in mind.
5. Unless it is possible to adequately serve the needs of multiple stakeholder groups with a single report format, each report should be developed with a format specific to each audience.
6. The format and elements of each report should be determined by conducting focus groups and/or multiple rounds of workshopping, with a focus on the following for each report element:
    - Identifying the critical "so-what" message(s) for the intended audience(s).
    - Assuring that the "so-what" message(s) are clearly and transparently conveyed.
    - Designing reports to minimize probable misinterpretations.
    - Assuring consistency with AERA/APA/NCME standards for score reporting[25].
7. The reporting system should allow for teachers to receive dynamic individual reports for just their current students, and aggregate reports for their current and past students.
8. The reporting system should allow for each audience to obtain the desired information using intuitive navigation and assistance in finding reports to answer specific questions. Report users should be able to retrieve data to answer their questions with a minimum number of clicks through guided selection of options. Where access to data is appropriate, report users should be able to easily retrieve data about achievement and growth for individual students and demographic groups at the student, classroom, school, district, and state level; with simple navigation between levels.

L. Wyoming Educator Participation in Ongoing Development

After qualified bidders have been identified, vendors whose proposals are consistent with recommendations in this section should receive extra credit[26]. Although avoiding an exclusive

[25] APA, AERA, & NCME (2014).

[26] This assumes a bidding process in which vendor bids are first scored to determine whether they meet a threshold for qualifying to provide the state with assessment services, followed by a review of the bids for a few areas in which select vendors may receive extra credit for proposing "value added" beyond the requirements of the request for proposals (RFP).

Wyoming assessment means that development will already be completed, it is desirable that Wyoming educators have the opportunity to be involved in ongoing development and maintenance of the assessment. Therefore, in order to improve the fit of the assessment to the Wyoming context, and to assure understanding of the assessment by Wyoming educators, the Task Force recommended the following:

1. Wyoming educators have substantive say in ongoing development activities including item development, item review, rangefinding, and other development activities.
2. Wyoming educators have the opportunity to review test questions for specific Wyoming sensitivities.
3. If there are alternative test questions available to replace those flagged as problematic by Wyoming educators, WDE is able to replace the flagged questions.
4. Wyoming educators are involved in scoring student responses requiring human scoring for tests completed by Wyoming students
5. The Wyoming Department of Education defines and oversees Wyoming educator involvement.

M. Test Security

In order to avoid the considerable stress and disruption to students, educators, and families caused by test security breaches, the Task Force recommended the following:

1. The Department of Education should review its existing policy documents and associated training using industry standards on test security.
2. The policy document and training must include clear policies, protocols, and guidelines to comprehensively address test security in all aspects of testing including at least the following areas:
    - Professional development
    - Prevention of test security breaches
    - Detection of test security breaches (including balancing protection for whistleblowers and minimizing the impact of malicious allegations)
    - Investigating potential security breaches
    - Protocols for evaluating evidence to make conclusions
    - Protocols for appeals of conclusions
    - Follow-up activities to a substantiated or suspected security breach
3. The Department of Education's test administration vendor must assist with test security to supplement agency capacity in each of the areas listed in the previous recommendation.
4. The Department of Education's test administration vendor must document its own security procedures throughout its processes.

N. Data Security and Privacy

In order to protect the privacy of individual student data and to comply with state and federal student privacy laws, the Task Force recommended that the vendor must document that its corporate policies on data security and privacy comply with all applicable state and federal statutes and regulations, that those policies are adequately strong to prevent data security breaches, and that those policies are rigorously enforced.

O. Program Evaluation and Its Relationship with System Stability

In order to determine whether the State's investment in a new comprehensive assessment system is achieving the intended results, the Task Force recommended the following:

1. The state should contract for an independent summary report evaluating the degree to which the intended outcomes of the state summative assessment have been realized after five years of implementation.
2. The evaluation should include the following at a minimum:
    - The quality of the state assessment
    - The degree to which intended short-, mid-, and long-term outcomes are being realized
    - The degree to which anticipated unanticipated unintended consequences have been observed
3. In addition to its Technical Advisory Committee (TAC) responsible for consulting with the Department to monitor the technical quality of its assessments, the Department of Education should empanel from this point forward a statewide assessment policy advisory committee (PAC) that meets at least twice a year to monitor for concerns from the perspective of Wyoming education stakeholders. This panel should include teachers, administrators, technology coordinators, and assessment coordinators. Because stability of the state assessment is paramount, the first activity of this committee should be defining thresholds for recommending changes to the system. These definitions should strongly privilege stability of the system over time, meaning that thresholds concerns about the assessment must meet before changes are made must be high.

P. Specialty Assessments

The Task Force focused its efforts on designing a coherent assessment system for the general student population in the content areas comprising the basket of goods. The Task Force also recognized the importance of coherence of its recommendations in four additional specialty areas:

1. Alternate assessments based on alternative achievement standards for students with significant cognitive disabilities (the "1%")
2. English proficiency assessment for English language learners
3. Early literacy assessment in grades K-3
4. Wyoming Career Technical Assessment (WyCTA) for career and technical education concentrator students

However, the Task Force was largely comprised of general educators, and recognized the need for specialists in each of these areas to make appropriate recommendations for these assessments. Therefore, the Task Force recommended that in each of these three areas, the Department of Education convene small committees of experts to review the recommendations for state summative assessment presented in this report and then make recommendations for those assessments to be coherent with the general content area assessments by determining which of the recommendations in this report are appropriate for those assessments, which are inappropriate, which need to be modified, and to identify any additional recommendations that may be needed.

SECTION 6: RECOMMENDATIONS FOR POLICY COHERENCE

Introduction

The Task Force took great care in ensuring that the recommendations put forth in this report are technical and practically sound. However, the Task Force was aware and concerned that several of the recommendations contradict existing statute, and that current state procurement policies may lead to problems with maintaining a stable state assessment system. In this section, we therefore list specific statutes that will need to be amended or repealed in order to implement the recommendations issues here. Prior to offering specific recommendations to the legislature, we offer general guidelines for legislating assessment requirements.

General Guidelines for Legislating Assessment Requirements

The Task Force spent considerable time discussing and trying to outline a coherent and efficient assessment system for Wyoming. One of the key features of a coherent assessment system is that each assessment in the system is designed to measure the same learning targets in complimentary ways. Further, in order to create an efficient system that minimizes redundancy, each assessment must be carefully designed to produce the intended inferences and to thoughtfully occupy a place in the overall system. It is easy to start adding assessments to meet specific needs (e.g. to support the evaluation of the Bridges program), but this can quickly lead to an incoherent and inefficient set of assessments that no longer function as a system.

Therefore, the Task Force strongly recommended that the legislature create statutes to set broad goals and articulate the intended uses of assessments (e.g., measuring student growth, for use in school accountability determinations). The legislature should prioritize creating a coherent, comprehensive, and efficient assessment system designed to measure student learning of Wyoming content standards and to support school improvement efforts. On the other hand, the legislature should avoid legislation regarding the specifics of assessment design (e.g., types of items to be included on the assessment) or even requiring assessments for specific purposes (e.g., requiring a 3rd grade reading assessment). The Task Force was aware that each time the legislature adds an assessment (e.g., ACT) or adds a specific requirement (e.g., multiple-choice items only), it is for well-intentioned reasons often in response to constituent concerns. Unfortunately, while every action might be well-intentioned, when we look back after a few years, a once coherent assessment is no longer so. Finally, the legislature should never name a specific product in legislation or write statutory requirements so narrowly that only one product or vendor meets the qualifications. It is rare that the legislature possesses the specialized knowledge necessary to recommend a specific assessment product, but most importantly, naming a specific product puts the state in a terrible position for negotiating a contract.

Designing and implementing a stable, efficient, and coherent assessment system requires high levels of technical and practical knowledge. Therefore, we compliment the legislature for appointing the Assessment Task Force, a representative group of citizens, to try to bring more coherence and stability to the Wyoming assessment system. Further, statute tends to last longer than rules and they are often much more difficult to change, especially considering that the Wyoming legislature is in session only 20 or 30 days each year, while the State Board of Education meets monthly to allow for more rapid modification of rules and requirements.

Thus the Task Force recommended that whenever a new potential purpose for assessment arises in state-level policymaking, the following activities should take place:

1. The legislature (and other responsible policymakers) should evaluate with education stakeholders whether the purpose is sufficiently important to justify expanding and disrupting the current assessment system.
2. If deemed sufficiently important, the legislature (and other responsible policymakers) should request a plan from the Wyoming Department of Education (WDE) for how existing assessments could reasonably fulfill that purpose, how existing assessments might be minimally expanded, or how a new assessment might fulfill the purpose (in order of preference).
3. WDE should develop a plan to avoid introducing new assessments if possible. If additional testing time or a new assessment is required, the plan should address thoughtful integration into the existing assessment system in a manner that will minimize disruption of student and educator activities. The Wyoming Department of Education should include stakeholders in developing the plan from both a broad cross section of education stakeholders and experts in the area of the intended purpose.

Wyoming State Statutes Needing Amendment or Repeal

With that framework, we outline the following recommended changes to existing statute to allow the recommendations presented here to be enacted.

1. *W.S. 21-2-202 (a)*[27]*: administering a standardized, curriculum based, achievement college entrance examination, computer-adaptive college placement assessment and a job skills assessment test selected by the state superintendent to all students in the eleventh and twelfth grades throughout the state in accordance with this paragraph.* This clause basically requires the ACT and a placement exam such as Accuplacer. The Task Force recommendations would still require the provision of a college entrance or work readiness exam, but the Task Force made no such recommendation for a placement exam. Such an exam may be useful once students enroll in a postsecondary institution, but not as part of the state assessment system. Further, the language of "curriculum based, achievement college entrance exam" is an example of trying to limit the potential successful bidders and the Task Force recommends a more neutral requirement for a college entrance and career readiness exam.

2. *W.S. 21-2-304 (iv)*[28]*. Effective school year 2013-2014, and each school year thereafter, require district administration of common benchmark adaptive assessments statewide in reading and mathematics for grades one (1) through eight (8) in accordance with W.S. 21-3-110(a)(xxiv).* The Task Force recommended the optional (at the district level) use of interim assessments, but most importantly to have the interim assessment procured as part of the state assessment RFP. The Task Force did not recommend the use of an adaptive assessment, per se, but for an interim system that best fit the instructional needs of districts. This is an example of what might be considered over-specification of the interim assessment requirement.

[27] Also found in W.S. 21-3-110
[28] Also found in W.S. 21-3-110

3. *W.S. 21-2-304 (v) (B). Effective school year 2012-2013, and each school year thereafter, be administered in specified grades aligned to the student content and performance standards, specifically assessing student performance in reading and mathematics at grades three (3) through eight (8). In addition, the statewide assessment system shall assess student performance in science in grades four (4) and eight (8).* As seen earlier in this report, the Task Force is recommending administering the state assessment system in English language arts and mathematics continuously in grades 3-10. The Task Force suggests leaving the science assessment in place until new content standards are adopted.

4. *W.S. 21-2-304 (v) (C). In addition to subparagraph (a)(v)(B) of this section, measure student performance in Wyoming on a comparative basis with student performance nationally*. While this requirement has not been implemented previously, except through the National Assessment of Educational Progress (NAEP), the Task Force supports the intention of this clause.

5. *W.S. 21-2-304 (v) (E). Use only multiple choice items to ensure alignment to the statewide content and performance standards*. The legislature already knows this is a problematic clause, but has been waiting for recommendations from the Task Force to deal with this clause. The Task Force has made clear that it wants to be able to include the types of test questions necessary to fully and deeply measure the Wyoming content standards and not be limited in the types of questions available to use. This is also an example of the type of specification that should not be in statute.

6. *W.S. 21-3-401: Reading assessment and intervention*. The Task Force did not have the time or the specific expertise necessary to address the reading assessment requirements, but recommends that WDE convene an expert advisory panel to make recommendations regarding K-3 reading assessment. While there is often a desire to produce comparable (standardized) data, early childhood reading assessments must yield information so that teachers can understand students' unique strengths and weaknesses. This might require the use of individually-administered assessments tied to each district's specific reading program.

7. *W.S. 21-13-334 (h)(iv) Implement a structured common assessment evaluation of program effectiveness*. While not specified in this clause, the common, adaptive interim assessment required under *W.S. 21-2-304 (iv)* has been the defacto common assessment used as the evaluation instrument for this program. As noted in this report, the Task Force argued that the timing of the common interim assessment was not necessarily appropriate for providing data to evaluate the efficacy of the program. Therefore, the Task Force recommends removing this requirement and replacing it with a requirement for districts to provide an appropriate evaluation of their specific program. WDE should be charged with providing guidance to districts on how best to collect evaluation data tied to the specific requirements of each program.

There are likely other statutes related to statewide and district assessment requirements, but the statutes outlined above are the highest priority targets for modification in order to implement the Task Force recommendations.

A Recommended Variance in Standard Wyoming Procurement Practice

The Task Force understood that typical Wyoming state procurement practices limit contracts to three years. However, this can cause instability in a state assessment program. Changes in contractors introduce changes in the assessment program, even if the same product is used. The changes to the product may be minor, but the state, districts, and schools have to divert attention from other important activities to adapting to new processes and/or products used by a new vendor.

To maximize stability of the State assessment system over time, the Task Force recommended that the legislature direct Wyoming procurement officials to grant a variance from standard procurement practice as detailed below.

1. A new contract to provide Wyoming's state assessments should be awarded for 5-8 years, with the option for 1-5 extension years, with the length of the original contract and number of extension years being negotiated between state procurement officials and the Wyoming Department of Education.
2. The number of contract years available through both the original contract and extension years should be targeted at 9 years to coincide with the required lifespan of Wyoming content standards.
3. Vendors should be required to include in their pricing specific costs for each of the 5-8 original contract years.
4. Vendors should be required to include in their pricing objective methods for determining costs for each of the 1-5 extension years, based primarily on pricing for the original contract years and national economic conditions.
5. Because in long-term contracts, contract changes are inevitable, vendors should propose fair methods for determining contract change prices, based primarily on pricing for the original contract years and pricing for similar activities carried out for other clients.
6. The larger number of available contract years should not limit the state's ability to respond to issues of contract non-performance.
7. When the program is rebid, the cost of a change to the assessment system should be weighed against any cost savings proposed by vendors proposing a different product to determine best value for the state. However, to avoid market restriction, the weighting of the costs to schools and districts should be relatively weak compared to the weights assigned to ratings of the proposals themselves. This is intended to assure that competitors have a reasonable probability of success if they propose a high-quality assessment at a competitive price.

SECTION 7: ABBREVIATED THEORY OF ACTION

This section does not provide a complete theory of action for the recommendations in this report[29]. Instead, in the following table, this section gives an abbreviated theory of action showing connections between a few key recommendations and specific intended outcomes, potential unintended negative consequences of implementing key recommendations, and potential mechanisms to deter such unintended negative consequences. The intended outcomes are summarized from other sections in this report. Mechanisms for deterring unintended consequences tend to be drawn from other recommendations made by the Task Force in anticipation of the unintended consequences.

Table 8.1 *Abbreviated Theory of Action for Key Task Force Recommendations.*

Key Recommendation(s)	Intended Outcomes	Potential Unintended Negative Consequences	Deterring Unintended Negative Consequences

• Standards-Based Assessment in Grades 3-10.
• *Either* College Entrance *or* Career Readiness Assessment in Grade 11.
	• Educators and policymakers use continuous achievement and growth data from grade 3-10 to inform:
	  ○ Yearly instructional planning
	  ○ Yearly curriculum and program evaluation
	  ○ Policymaking
	• Clarify that the Wyoming High School learning targets are the official Wyoming state standards.
	• Retain the benefits of a college entrance examination.
	• Better meet the needs of high school students with career and technical education goals.
	• Allow and encourage specialized pathways for grade 11 and 12 students, improving student engagement and opportunity.
	• Strengthen ties between Wyoming high schools and Wyoming institutions of higher education, career training, and technical education.
	• Limit testing time by ending standards-based accountability assessment in grade 10.
		• Official Wyoming state standards are ignored in grade 11 and 12.	• District assessment systems address high school standards not eligible to appear on the grade 10 assessment.
			• Improve quality control of district assessment systems through accreditation, training, and support.

[29] A full theory of action would explicitly tie together issues identified with the current system; intended uses of a new system; recommendations for a new assessment system and how they address the issues with the current system and the intended uses of the new system; connections between various components; near-term, mid-term, and long-term intended outcomes associated with each component as well as the whole; and measures to monitor those outcomes. Each of these components is addressed in this report. However, the critical connection is between recommendations and intended outcomes so that an evaluation plan can be developed.

Key Recommendation(s)	Intended Outcomes	Potential Unintended Negative Consequences	Deterring Unintended Negative Consequences
• Implement a standards-based summative assessment used in multiple states. • Allow an extended contracting period beyond that normally allowed.	• Improve stability of the state summative assessment by requiring changes to be negotiated with other states and/or a vendor. • Increase the number of years between considerable changes to the state assessment because of changes in contractors and/or products. • Allow comparison of Wyoming students' achievement and growth to that of students in other states. • Increase Wyoming's leverage to address issues as they arise through applying collaborative pressure with other states. • Reduce costs through economies of scale available through collaboration among multiple states. • Reduce costs through the incentive of an extended contracting period. • Increase test quality by gaining access to a broader set of expertise available through collaborative efforts.	• Slow collaborative response to critical needs. • Contractor becoming too comfortable and/or inadequately responsive. • Costs increase unacceptably over time.	• Split contracting years between guaranteed contract years and optional extension years. • Tie cost increases in extension years to original costs and economic conditions. • Maintain strong remedies for contract non-performance. • Require a defined process for developing costs for contract changes keyed to original contract costs and costs to other clients for similar services.
• Move all state-provided assessment online	• Allow state summative test administration closer to the end of the school year (to maximize instruction time before assessment). • Allow return of results before August 1 of each year (to make results available for school improvement activities in the summer). • Eliminate logistical challenges inherent in paper and pencil testing. • Increase flexibility for test administration, scoring, and reporting through use of information technology tools.	• Statewide breakdown of the test administration system. • Localized breakdowns of the needed information technology infrastructure. • Test security breach. • Data security breach. • Increased logistical challenges for districts and schools.	• Safeguards recommended to assure a smooth and successful transition to online assessment. • Review and enhance test security sections of existing state policy documents to address new issues in test security posed by online assessment. • Require vendors to demonstrate strong security policies and adherence to those policies. • WDE should empanel a policy advisory committee to monitor and advise on logistical and other issues.

Key Recommendation(s)	Intended Outcomes	Potential Unintended Negative Consequences	Deterring Unintended Negative Consequences
• Procure state-provided interim assessments with the state summative assessments • Implement modular interim assessment design at a minimum.	• Eliminate inconsistencies between state summative assessments and interim assessments. • Eliminate inconsistencies between the Wyoming state content standards and the interim assessment. • Signal the importance of high-level student knowledge and skills on both state summative assessment and interim assessments. • Improve the usefulness of interim assessment results by targeting smaller units of content and reporting on finer-grained categories. • Control costs by bundling multiple products.	• Concerns from districts accustomed to the current interim assessment.	• Design reports for interim assessments to assure usefulness to educators. • Increase flexibility for districts on timing and number of interim assessments. • Eliminate requirement to use interim assessment (provide as a state service for districts to implement to best serve local needs)

This table should be used as a starting point for the recommended five-year evaluation of the new system.

REFERENCES

AERA, APA, & NCME. (2014). *Standards for Educational and Psychological Testing.* Washington, DC: AERA.

Braun, H. (Ed.). (in press). *Meeting the Challenges to Measurement in an Era of Accountability.* Washington, DC: National Council on Measurement in Education.

CCSSO & ATP. (2013). *Operational Best Practices for Statewide Large-Scale Assessment Programs.* Washington, DC: Authors.

Coladarci, T. (2002). Is it a house…or a pile of bricks? Important features of a local assessment system. *Phi Delta Kappan, 83*(10), 772-774.

Michigan Department of Education. (2013). *Report on Options for Assessments Aligned with the Common Core State Standards.* Retrieved June 20, 2015, from http://www.michigan.gov/documents/mde/Common_Core_Assessment_Option_Report_441322_7.pdf.

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-Centered Assessment Design. In T. M. Haladyna, & S. M. Downing (Eds.), *Handbook of Test Development* (pp. 61-90). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). Knowing What Students Know: The Science and Design of Educational Assessment. Washington, DC. Retrieved September 4, 2015, from http://www.nap.edu/openbook.php?record_id=10019&page=R1.

Perie, M., Marion, S., & Gong, B. (2009). Moving towards a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice, 28*(3), 5-13.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). Universal Design Applied to Large Scale Assessments (Synthesis Report 44). Minneapolis, MI: University of Minnesota, National Center on Educational Outcomes. Retrieved September 5, 2015, from http://www.cehd.umn.edu/NCEO/onlinepubs/synthesis44.html.

Wiley, E. C. (2008). *Formative Assessment: Examples of Practice*. Retrieved August 11, 2015, from http://ccsso.org/documents/2008/formative_assessment_examples_2008.pdf.

APPENDIX A: UNDERSTANDING FORMATIVE ASSESSMENT

Definition of Formative Assessment

Formative assessment has also been called formative instruction. The purpose of formative assessment is to evaluate student understanding against key learning targets, provide targeted feedback to students, and adjust instruction on a moment-to-moment basis.

In 2006, the Council of Chief State School Officers (CCSSO) and experts on formative assessment developed a widely cited definition (Wiley, 2008):

> *Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievements of intended instructional outcomes* (p. 3).

In addition, Wiley (paraphrased from p. 3) lists five critical attributes of formative assessment:

1. They are based on clear articulations of learning goals as steps toward an ultimate desirable outcome.
2. Learning goals and the criteria for success are clearly identified and communicated to students in language they can understand.
3. Students are frequently provided with feedback directly linked to the learning goals and criteria for success.
4. Students engage in self- and peer-assessment against the criteria for success.
5. Students and teachers jointly own (collaborate on) monitoring student progress over time.

While the practice of formative assessment in general embodies these five attributes, not every example of formative assessment incorporates every attribute. The definition and five critical attributes are based on research linking such practices to student learning gains. The core of the formative assessment process is that it takes place during instruction (i.e., "in the moment") and under full control of the teacher to support student learning while it is developing. Thus, formative assessment is an integral part of instruction; instruction need not be paused to engage in formative assessment. This embedded assessment is done through diagnosing on a very frequent basis where students are in their progress toward fine-grained learning targets such as those covered by a single class period. This ongoing diagnosis shows both teachers and students where gaps in knowledge and skill exist, and helps both teacher and student understand how to close those gaps.

The definition and critical attributes make clear that formative assessment is not a product, but a process tailored to the details of ongoing instruction to individual students. Effective formative assessment practices occur very frequently, covering very small units of instruction (such as part of a class period). If tasks are presented, they may vary for students depending on where they are in their learning. However, formative assessment processes often occur during regular and targeted questioning of students in small or large groups, observing students as they work in groups and/or engage in tasks. Formative assessment practices may be facilitated using certain technology and related tools. There is a strong view among some scholars that because formative assessment is tailored to the specific context of the classroom and to individual students that results cannot be meaningfully aggregated or compared. Many of these scholars question whether the observations from formative assessment should even be scored.

Another implication is the critical importance of providing frequent feedback to individual students. Providing each student such frequent and targeted feedback develops his or her ability to continuously monitor the quality of their own work against a clear learning target. It is this targeted and frequent feedback to students that is the most crucial part of the formative assessment process[30].

The nature of formative assessment implies that the frequently used term *common formative assessment* is a result of confusion about the nature of formative assessment. Other types of assessment may be used formatively for periodic progress monitoring (e.g., to inform mid-course corrections or modifications to curriculum and programming), but only formative assessment as described above is capable of informing instruction on a moment-to-moment basis. Effective formative assessment is tailored to a specific instructional plan and a specific group of students at defined points in their attainment of learning targets. The critical characteristics of formative assessment practices should be common across all teachers, and tools teachers use to implement formative assessment may be common across many teachers, but formative assessment is too tailored to a unique classroom to be common.

Data gathered through formative assessment have limited to no use for evaluation or accountability purposes such as student grades, educator accountability, school/district accountability, or even public reporting that could allow for inappropriate comparisons. There are at least four reasons for this: (1) if carried out appropriately, the data gathered from one unit to the next, one teacher to the next, one moment to the next, and one student to the next will not be comparable; (2) students will be unlikely to participate as fully, openly, and honestly in the process if they know they are being evaluated by their teachers or peers on the basis of their responses; (3) for the same reasons, educators will be unlikely to participate as fully, openly, and honestly in the process; and (4) the nature of the formative assessment process is likely to shift in such a way that it can no longer optimally inform instruction.

These implications create a distinct difference from summative and interim assessment (described below), which are intended to assess student achievement after an extended period of learning. Simply giving students an assessment in the classroom does not mean that the assessment is formative. Use of assessment evidence in a formative manner requires teachers to achieve insight into individual student learning in relation to learning targets, to provide effective feedback to students about those insights, and to make instructional decisions based on those insights. During the formative assessment process, feedback to students and student involvement is essential. Teachers seek ways to involve the student in "thinking about their thinking" (metacognition) to use learning evidence to close the gap and get closer to the intended learning target.

Because there is a great deal of confusion over what constitutes formative assessment, the next part of this appendix provides vignettes of formative assessment in practice. The four vignettes describe the work of four different educators to help readers to better understand what is meant by "formative assessment."

[30] See Sadler (1989).

Vignettes of Formative Assessment in Practice[31]

High School – Chemistry Mid-Period Check In

As part of instructional planning, a high school chemistry teacher develops both true and false statements related to a micro-unit covering a half hour in high school chemistry. Statements were strategically developed to assess whether students hold anticipated misconceptions. Following the micro-unit, students show thumbs up, thumbs down, or thumbs to the side to indicate whether each statement is true, false, or they don't know. Based on the prevalence of thumbs down and to the side, the teacher may select one of at least four options:

1. Reteach that micro-unit using a different instructional plan the next day.
2. Use pre-planned strategies to address a small number of misconceptions.
3. Strategically group students who put thumbs down or to the side with confident students to discuss their conclusions and monitor group discussions.
4. Work briefly with a one or two students needing additional assistance while the rest of the class engages in the next activity.

Middle School – English End of Period Check In

At the beginning of a seventh grade English class period, a middle school English teacher shares with her students what the three learning targets are for the day. At the end of the period, she asks each student to fill out and hand in a slip confidentially rating their attainment of each learning target in one of the following four categories:

1. I can teach this.
2. I can do this on my own.
3. I need some help with this.
4. I don't get this at all.

The teacher adjusts the next day's lesson plan by creating a simple task asking small groups of students to practice a learning target on which about half the students felt confident. The small groups are strategically selected to include students that are both confident and not confident with the learning target. She also reviews with the entire class another learning target on which few students felt confident. To do so, she asks two students to explain their approach on a specific problem. After gauging current understanding, she decides whether to instruct on that learning target again using a different strategy and different examples than the previous day.

Elementary School – Monitoring Development of Mathematical Understanding

After a successful unit on simple two-digit addition (without regrouping), an elementary school teacher wants students to learn both a regrouping algorithm and why the algorithm works. He demonstrates to his students that their current knowledge and skills are inadequate to accurately deal with two-digit addition requiring regrouping. He does this by assigning small groups of students to solve a problem either using the addition algorithm they already know or by using counting objects. In a subsequent whole-class discussion, the teacher highlights the conflicting answers and asks his

[31] Informed by Wiley (2008).

students to think about how place value place might explain why the groups got different answers. He then asks each small group to work on developing its own solution to the problem. After visiting and probing each group to survey current understanding and developing strategies, he asks strategically chosen groups to share their developing solutions, and builds post-activity instruction on the regrouping algorithm around them.

High School – English Capstone Project

As a capstone project for a unit on persuasive writing, a high-school English teacher assigns her students to individually write a persuasive essay incorporating each of the unit learning targets. Each student is to:

- Choose a position on a controversial topic important to him,
- Identify reliable resources for information on his position and a contrary position commonly taken on the topic,
- Summarize the arguments for both positions,
- Use the logical devices taught in the unit to argue for his position,
- Use logical tools to argue the logical superior of his position, and
- Incorporate work in all five previous steps into a coherent persuasive essay.

The teacher divides the capstone project into four subunits (with associated assignments):

1. Choosing a topic, a personal position, an opposing position, and identifying reliable resources;
2. Summarizing arguments for at least two positions on the topic;
3. Arguing for the personal position and against an opposing position on a logical basis;
4. Incorporating into a complete and coherent persuasive essay.

Along with other formative practices, the teacher spends class time making each sub-unit's learning targets explicit and instructing on them. She also uses class time on the day each assignment is due to have students peer-review each other's work, focusing on the learning targets and working on revisions. As assignments are turned in, the teacher provides formative feedback based on the learning target rather than grading each assignment. Only after providing at least one round of formative feedback on each assignment does the teacher grade the final product. She does this to ensure that the formative feedback fulfills its purpose and her evaluation of each student's performance represents what was learned by the end of the unit.

APPENDIX B: ONE-PAGE SUMMARY OF FORMATIVE, INTERIM, AND SUMMATIVE ASSESSMENT

	Formative Assessment	Interim Assessment	Summative Assessment
Characteristics	• Facilitate effective instruction (does not pause instruction) • Learning goals and criteria are clear to students • Students self-/peer-monitor progress toward learning goals • Students and teachers receive frequent feedback • Jointly controlled by each teacher and her students • Covers a micro unit of instruction • Very frequent (e.g., multiple times per period) • Tailored to a set of students and an instructional plan • Might be comparable for a classroom, but not beyond • *Not a product (e.g., quiz, test, bank of questions/tests)*	• Pauses instruction for evaluation • Controlled solely by a teacher, school, district, or state (or by a consortium of teachers, schools…) • Covers a mid-sized unit of instruction • Somewhat frequent (e.g., weekly to quarterly) • Administered before and/or after a mid-sized unit • Based on who controls assessment, results may be comparable across students, teachers, schools, districts, and/or states • A product	• Pauses instruction for evaluation • Controlled solely by a teacher, school, district, or state (or by a consortium of teachers, schools…) • Covers a macro unit of instruction (e.g., semester, course, credit, grade) • Infrequent (e.g., yearly, finals week) • Administered after completing a macro unit • Based on who controls assessment ,results may be comparable across students,…, and/or states • A product
Uses	• Engage students in learning/metacognition through frequent feedback and self-/peer-evaluation • Monitor moment-to-moment student learning • Diagnose individual students' immediate instructional needs • Diagnose immediate group instructional needs • Immediately adjust instruction • Differentiate instruction • Self-evaluate micro-unit instructional effectiveness • *Student results from formative assessment are not appropriate for use in grading or accountability; however, ratings of the quality of formative assessment practice may be appropriate for use in accountability*	• Evaluate achievement after a mid-sized unit • Monitor progress within a macro-unit (e.g., semester, course, credit, grade) • Corroborate formative assessment • Pre-test to tailor unit instructional plans for the group and individual students • Identify post-unit remedial needs • Mid-course self-evaluation and adjustment of teacher classroom practices • Mid-course evaluation and adjustment of school and district policies and programs • Predict performance on summative assessment • Grading (and possibly accountability)	• Evaluate achievement after a macro unit • Monitor progress across multiple macro-units • Corroborate interim assessment • Evaluate readiness for the next macro unit • After-the-fact evaluation/adjustment of broad instructional practices by individual teachers and of curriculum/programming policies by administrators • Predict later student outcomes • Grading and accountability
Examples	• Following a micro-unit, students show thumbs up/thumbs down to indicate whether statements developed around anticipated misconceptions are true. Based on prevalence of misconceptions, the teacher reteaches parts of his lesson using a different instructional strategy, strategically groups students to discuss their conclusions, or works briefly with one or two students. • At the end of class, students hand in a slip confidentially rating their attainment of each learning target as: (1) *I can teach this*, (2) *I can do this on my own*, (3) *I need some help with this*, or (4) *I don't get this at all*. The teacher adjusts her next-day group assignments and planned activities accordingly.	• Classroom unit quizzes and homework • Individual and group unit projects • Pre-unit exams of unit pre-requisites • Pre-unit exams of unit content • End of unit exams • Mid-term exams • Marking period exams not covering a full macro-unit • Quarterly assessments • District placement tests	• Classroom final exams, projects, and papers • School or district final exams, projects, or papers • District/state assessments for testing out of a credit • District graduation/diploma-endorsement tests • Typical state accountability tests • High school equivalency tests • District graduation tests • College admission tests

APPENDIX C: DETAILED HIGHEST PRIORITY USES AND CHARACTERISTICS

The Task Force's highest priority uses and characteristics are presented in detail in Table B1 below. These uses and characteristics were evaluated by the facilitators using the definitions and appropriate uses of formative, interim, and summative assessments discussed in Section 2 of this report. The evaluation also incorporates differences between classroom-, district-, and state-owned assessments to show the complexity of an assessment system that would be needed to fulfill all of the Task Force's highest priority uses and characteristics. This evaluation is reflected in additional elements added to Table B1. Those elements identify whether each type and level of assessment has full, some, minimal, or no applicability to the use or characteristic in each row. In addition, in each row the applicability of the various types and levels of assessment to each use or characteristic is briefly explained.

Table B1. *Task Force Highest Priority Uses and Characteristics.*

Total[1] Score	Number of Votes by Priority			Desired Uses and Characteristics of Wyoming Assessment	Applicability[2]					
					Type			Level		
	1st	2nd	3rd		Formative	Interim	Summative	Classroom	District	State
38	10	3	2	Provide information to parents, students, and educators regarding individual student achievement and growth within and across years, including readiness for the next level in a student's K-12 progression *- Classroom formative: continuous achievement/growth/readiness data on micro-units* *- Classroom/district/state interim: periodic achievement/growth/readiness data on mid-sized units* *- Classroom/district/state summative: yearly achievement/growth/readiness data on macro-units*	◑	◑	◑	◑	◑	◑
27	6	4	1	Provide feedback on progress toward standards to inform instruction on more than a yearly basis *- Classroom formative: continuous achievement and progress data inform daily instruction* *- Classroom/district/state interim: periodic unit achievement & progress data informs remediation* *- District/state summative: interim results might be rolled up for summative determinations*	●	●	◔	●	●	●
16	0	5	6	Allow for comparisons within the state and across states *- State interim: provides within-state comparability if adopted statewide* *- State summative: provides within-state comparability* *- State interim/summative: provides cross-state comparability if the assessment is used in multiple states*	○	◑	●	○	○	●
13	2	2	3	Provide reliable and valid data to evaluate program/curriculum effectiveness and alignment to standards *- District/state interim: can provide information to inform within- and between-year evaluations* *- District/state summative: can provide information to inform between-year evaluations*	○	●	●	○	●	●

Total[1] Score	Number of Votes by Priority 1st	2nd	3rd	Desired Uses and Characteristics of Wyoming Assessment	Applicability[2] Type: Formative	Interim	Summative	Level: Classroom	District	State
11	3	1	0	Be student-centered (e.g., student is not a number) - *Classroom formative: micro-unit diagnostic data to tailor instruction* - *Classroom/district/state interim: unit diagnostic data to tailor remediation* - *Classroom/district/state summative: macro-unit data to inform critical yearly decisions*	●	◑	◔	●	◑	◑
8	0	3	2	Encourage collaboration and sharing best practices - *Classroom formative/interim/summative: foster teacher collaboration on teacher practices* - *District/state interim/summative: foster teacher collaboration on using non-classroom data* - *District/state interim/summative: foster educator collaboration on curriculum/programming* - *Limit use of classroom assessment for evaluation to* quality of practices and support for collaboration	●	●	●	●	◑	◑
7	1	2	0	Continually inform instruction with timely feedback - *Classroom formative: continual micro-unit diagnostic data to inform daily instruction* - *Classroom/district/state interim: periodic unit data to inform post-unit remediation*	●	◔	○	●	◔	◔
6	1	1	1	Validly inform decisions about post-secondary education/training - *State summative: likely to provide based on ties to post-secondary outcomes (onerous for a district)*	○	○	●	○	◔	●
2	0	0	2	Consistency over time to facilitate the intended outcomes of assessment in Wyoming - *District interim/summative: stable longitudinal data can improve decision making* - *State interim: stable longitudinal data can improve decision making* - *State summative: likely to improve decision-making because of school/district accountability uses*	○	◑	●	○	◑	●
				Number of desired uses/characteristics with unique and full applicability	2	0	3	3	0	3
				Number of desired uses/characteristics with full applicability	4	3	5	4	2	5
				Number of desired uses/characteristics with some applicability	1	4	1	1	4	3
				Number of desired uses/characteristics with unlikely applicability	0	1	2	0	2	1
				Number of desired uses/characteristics with no applicability	4	1	1	4	1	0

1. Each panelist identified one characteristic as her highest priority, second highest priority, or third highest priority. These were given scores of 3, 2, and 1 respectively. The scores were summed across panelists to give a total score for each desired use/characteristic.
2. ●,◑,◔, and ○ indicate desired uses or characteristics for which the type or level of assessment has full applicability, some applicability, minimal or unlikely applicability, and no applicability, respectively.

APPENDIX D: MINI-SUMMATIVE VS. MODULAR INTERIM ASSESSMENT DESIGNS

To help illustrate the differences between a mini-summative and modular design, we present an abbreviated pictorial representation of the two designs below. In a mini-summative design, the interim assesmsents are in essence, just shorter versions of the summative assessment. In a modular design, the interim assessments focus on specific portions of what was covered by the complete summative assessment to give more fine-grained information about student achievement within the content area of the summative assessment. A more detailed explanation of how this might be accomplished is given on the following pages.

Figure D.1. Mini-summative Interim Assessment Design Schematic.

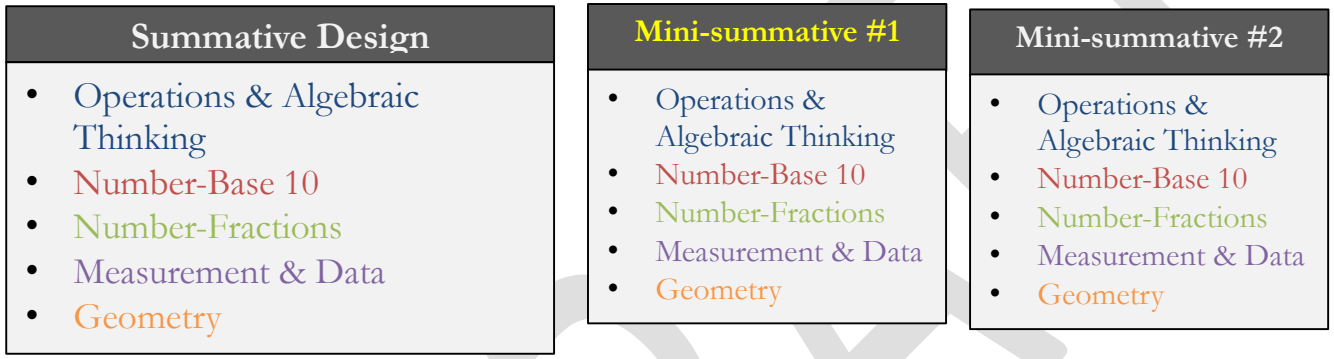
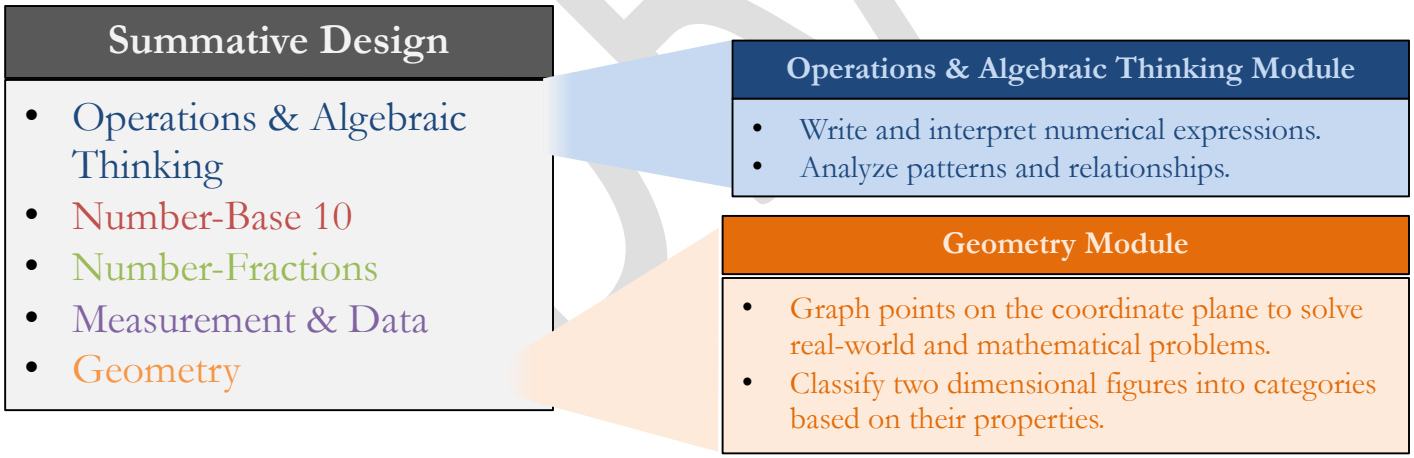


Figure D.2. Modular Interim Assessment Design Schematic.



As an aid in further understanding assessment design, we first describe the general hierarchical format that content standards take by providing an example from grade-5 mathematics:

Content Category
Operations & Algebraic Thinking Write and interpret numerical expressions *Use parentheses, brackets, or braces…* *Write simple expressions that record calculations…* Analyze patterns and relationships *Generate…numerical patterns…given rules…*
Number & Operations in Base Ten Understand the place value system *Recognize [digit values increase tenfold when one place… left]* *Explain patterns in…multiplying by powers of 10…* *Read, write, and compare decimals to thousandths* *Use place value understanding to round decimals to any place* Perform operations…to hundredths *Fluently multiple multi-digit whole numbers…* *Find whole-number quotients of whole numbers…* *Add, subtract, multiply, and divide decimals to hundredths…*
Number & Operations—Fractions Use equivalent fractions…to add and subtract fractions *Add and subtract fractions with unlike denominators…* *Solve [fraction word problems by comparison…]* Apply and extend…multiplication and division *Interpret a fraction [as a division problem]…* *[Extend whole number] multiplication to…fractions…* *Interpret multiplication as scaling (resizing)…* *Solve…problems [with] multiplication of fractions…* *[Extend division to involve unit fractions]*
Measurement & Data Convert like measurement units [in the same] system *Convert among different sized measurement units…* Represent and interpret data *Make a line plot to display [data with fractional units]…* Geometric measurement: understand…volume *Understand volume as an attribute of solid figures…* *Measure volumes by counting unit cubes…* *Relate volume to [multiplication and division]…*
Geometry Graph points on the coordinate plane to solve… *Use [two] perpendicular lines…to define a coordinate…* *Represent… points in the first quadrant…* Classify two-dimensional figures…on…properties *[Know category] attributes [apply] to all sub-categories…* *Classify…figures in a hierarchy based on properties*

To aid in explanation, the broadest content categories (at the top of the hierarchy) are displayed in bold. Sub-categories are indented presented in the same color as the broad category they belong to. Sub-sub-categories are further indented and presented in italics.

In a *highly simplified* version of test design, the number of test questions or score points that come from each sub-sub-category is clearly specified to reflect the relative importance of each category. For example, if every sub-sub-category were considered equally important, a reasonable test design

might specify that every sub-sub-category be measured using two test questions, resulting in the following hypothetical summative test design:

Content Category	# of Items
Operations & Algebraic Thinking	**6**
Write and interpret numerical expressions	4
Use parentheses, brackets, or braces…	*2*
Write simple expressions that record calculations…	*2*
Analyze patterns and relationships	2
Generate…numerical patterns…given rules…	*2*
Number & Operations in Base Ten	**14**
Understand the place value system	8
Recognize [digit values increase tenfold when one place… left]	*2*
Explain patterns in…multiplying by powers of 10…	*2*
Read, write, and compare decimals to thousandths	*2*
Use place value understanding to round decimals to any place	*2*
Perform operations…to hundredths	6
Fluently multiple multi-digit whole numbers…	*2*
Find whole-number quotients of whole numbers…	*2*
Add, subtract, multiply, and divide decimals to hundredths…	*2*
Number & Operations—Fractions	**14**
Use equivalent fractions…to add and subtract fractions	4
Add and subtract fractions with unlike denominators…	*2*
Solve [fraction word problems by comparison…]	*2*
Apply and extend…multiplication and division	10
Interpret a fraction [as a division problem]…	*2*
[Extend whole number] multiplication to…fractions…	*2*
Interpret multiplication as scaling (resizing)…	*2*
Solve…problems [with] multiplication of fractions…	*2*
[Extend division to involve unit fractions]	*2*
Measurement & Data	**10**
Convert like measurement units [in the same] system	2
Convert among different sized measurement units…	*2*
Represent and interpret data	2
Make a line plot to display [data with fractional units]…	*2*
Geometric measurement: understand…volume	6
Understand volume as an attribute of solid figures…	*2*
Measure volumes by counting unit cubes…	*2*
Relate volume to [multiplication and division]…	*2*
Geometry	**8**
Graph points on the coordinate plane to solve…	4
Use [two] perpendicular lines…to define a coordinate…	*2*
Represent… points in the first quadrant…	*2*
Classify two-dimensional figures…on…properties	4
[Know category] attributes [apply] to all sub-categories…	*2*
Classify…figures in a hierarchy based on properties	*2*
Total	**52**

A *mini-summative interim assessment design* is intended to reasonably replicate the summative assessment experience with the exception of being shorter. For example, on an interim assessment with five testing opportunities, this could be accomplished by measuring each content standard with 1 rather than 2 items, giving the following mini-summative interim assessment design, making each interim assessment half as long as the summative assessment:

Content Category	# of Items on Interim Assessment				
	1	2	3	4	5
Operations & Algebraic Thinking	**3**	**3**	**3**	**3**	**3**
Write and interpret numerical expressions	2	2	2	2	2
Use parentheses, brackets, or braces…	*1*	*1*	*1*	*1*	*1*
Write simple expressions that record calculations…	*1*	*1*	*1*	*1*	*1*
Analyze patterns and relationships	1	1	1	1	1
Generate…numerical patterns…given rules…	*1*	*1*	*1*	*1*	*1*
Number & Operations in Base Ten	**7**	**7**	**7**	**7**	**7**
Understand the place value system	4	4	4	4	4
Recognize [digit values increase tenfold when one place… left]	*1*	*1*	*1*	*1*	*1*
Explain patterns in…multiplying by powers of 10…	*1*	*1*	*1*	*1*	*1*
Read, write, and compare decimals to thousandths	*1*	*1*	*1*	*1*	*1*
Use place value understanding to round decimals to any place	*1*	*1*	*1*	*1*	*1*
Perform operations…to hundredths	3	3	3	3	3
Fluently multiple multi-digit whole numbers…	*1*	*1*	*1*	*1*	*1*
Find whole-number quotients of whole numbers…	*1*	*1*	*1*	*1*	*1*
Add, subtract, multiply, and divide decimals to hundredths…	*1*	*1*	*1*	*1*	*1*
Number & Operations—Fractions	**7**	**7**	**7**	**7**	**7**
Use equivalent fractions…to add and subtract fractions	2	2	2	2	2
Add and subtract fractions with unlike denominators…	*1*	*1*	*1*	*1*	*1*
Solve [fraction word problems by comparison…]	*1*	*1*	*1*	*1*	*1*
Apply and extend…multiplication and division	5	5	5	5	5
Interpret a fraction [as a division problem]…	*1*	*1*	*1*	*1*	*1*
[Extend whole number] multiplication to…fractions…	*1*	*1*	*1*	*1*	*1*
Interpret multiplication as scaling (resizing)…	*1*	*1*	*1*	*1*	*1*
Solve…problems [with] multiplication of fractions…	*1*	*1*	*1*	*1*	*1*
[Extend division to involve unit fractions]	*1*	*1*	*1*	*1*	*1*
Measurement & Data	**5**	**5**	**5**	**5**	**5**
Convert like measurement units [in the same] system	1	1	1	1	1
Convert among different sized measurement units…	*1*	*1*	*1*	*1*	*1*
Represent and interpret data	1	1	1	1	1
Make a line plot to display [data with fractional units]…	*1*	*1*	*1*	*1*	*1*
Geometric measurement: understand…volume	3	3	3	3	3
Understand volume as an attribute of solid figures…	*1*	*1*	*1*	*1*	*1*
Measure volumes by counting unit cubes…	*1*	*1*	*1*	*1*	*1*
Relate volume to [multiplication and division]…	*1*	*1*	*1*	*1*	*1*
Geometry	**4**	**4**	**4**	**4**	**4**
Graph points on the coordinate plane to solve…	2	2	2	2	2
Use [two] perpendicular lines…to define a coordinate…	*1*	*1*	*1*	*1*	*1*
Represent… points in the first quadrant…	*1*	*1*	*1*	*1*	*1*
Classify two-dimensional figures…on…properties	2	2	2	2	2
[Know category] attributes [apply] to all sub-categories…	*1*	*1*	*1*	*1*	*1*
Classify…figures in a hierarchy based on properties	*1*	*1*	*1*	*1*	*1*
Total	**26**	**26**	**26**	**26**	**26**

Multiple interim assessments built to this design would have different sets of test questions, but with the same emphasis on each of the content categories as on the summative assessment.

Modular interim assessment designs are different, however. Modular designs are intended to focus in on strategically selected subsets of the content standards (typically selected to represent potential moderate-sized units of instruction). Therefore, modular interim assessment designs are not similar to the summative test design. For example, in a highly simplified approach, each of the five broadest content categories could be selected as the focus for each of five interim assessment modules, giving

the following modular interim assessment design of approximately the same length as the mini-summative designs:

Content Category	# of Items on Interim Assessment				
	1	2	3	4	5
Operations & Algebraic Thinking	27				
Write and interpret numerical expressions	18				
Use parentheses, brackets, or braces…	*9*				
Write simple expressions that record calculations…	*9*				
Analyze patterns and relationships	9				
Generate…numerical patterns…given rules…	*9*				
Number & Operations in Base Ten		28			
Understand the place value system		16			
Recognize [digit values increase tenfold when one place… left]		*4*			
Explain patterns in…multiplying by powers of 10…		*4*			
Read, write, and compare decimals to thousandths		*4*			
Use place value understanding to round decimals to any place		*4*			
Perform operations…to hundredths		12			
Fluently multiple multi-digit whole numbers…		*4*			
Find whole-number quotients of whole numbers…		*4*			
Add, subtract, multiply, and divide decimals to hundredths…		*4*			
Number & Operations—Fractions			28		
Use equivalent fractions…to add and subtract fractions			8		
Add and subtract fractions with unlike denominators…			*4*		
Solve [fraction word problems by comparison…]			*4*		
Apply and extend…multiplication and division			20		
Interpret a fraction [as a division problem]…			*4*		
[Extend whole number] multiplication to…fractions…			*4*		
Interpret multiplication as scaling (resizing)…			*4*		
Solve…problems [with] multiplication of fractions…			*4*		
[Extend division to involve unit fractions]			*4*		
Measurement & Data				25	
Convert like measurement units [in the same] system				5	
Convert among different sized measurement units…				*5*	
Represent and interpret data				5	
Make a line plot to display [data with fractional units]…				*5*	
Geometric measurement: understand…volume				15	
Understand volume as an attribute of solid figures…				*5*	
Measure volumes by counting unit cubes…				*5*	
Relate volume to [multiplication and division]…				*5*	
Geometry					28
Graph points on the coordinate plane to solve…					14
Use [two] perpendicular lines…to define a coordinate…					*7*
Represent… points in the first quadrant…					*7*
Classify two-dimensional figures…on…properties					14
[Know category] attributes [apply] to all sub-categories…					*7*
Classify…figures in a hierarchy based on properties					*7*
Total	27	28	28	25	28

The benefit of a modular interim assessment design is that it can provide much more granular and instructionally useful information because there are enough items measuring fine-grained categories of content to inform broad (not day-to-day) instructional and/or remedial decisions.

APPENDIX E: DETAIL ON ISSUES IN SUB-SCORE REPORTING

Subscores serve as achievement reports on subsets of the full set of knowledge and skill represented by a total score. For example, many English language arts summative assessments produce a total score for English language arts, subscores for at least reading and writing, and often finer-grained subscores for topics such as informational and literary reading. Similarly, a mathematics test typically yields an overall math score and potential subscores in topics such as numbers and operations, algebraic reasoning, measurement and geometry, and statistics and probability. One of the greatest challenges in current large-scale summative assessment design is to create tests that are no longer than necessary to produce a very reliable total score (e.g., 5th grade mathematics) while yielding adequately reliable subscores to help educators and others gain more instructionally-relevant information than gleaned from just the total score.

Unfortunately, there is a little known aspect of educational measurement (outside of measurement professionals) that large-scale tests are generally designed to report scores on a "unidimensional" scale. This means that the 5th grade math test, for example, is designed to report overall math performance, but not to tease out differences in performance on things like geometry or algebra because the only questions that survive the statistical review processes are those that relate strongly to the total score of overall math. If the test was designed to include questions that better distinguish among potential subscores, the reliability (consistency) of the total score would be diminished. There are "multidimensional" procedures that can be employed to potentially produce reliable and valid subscores, but these are much more expensive to implement and complicated to ensure the comparability of these subscores and the total score across years. The National Assessment of Educational Progress (NAEP) is the one example of a well-known assessment designed to produce meaningful results at the subscore level, but NAEP has huge samples to work with and more financial resources and psychometric capacity at its disposal than any state assessment. In other words, it is not realistic at this time to consider moving away from a unidimensional framework for Wyoming's next statewide summative assessment, which means that the subscores will unfortunately be much less reliable estimates of the total score than useful content-based reports. This is true for essentially all commercially-available interim assessments as well so that in spite of user reports that they like assessment X or Y because it produces fine-grain subscores useful for instructional planning, any differences in subscores are likely due to error rather than anything educationally meaningful.

In spite of this widely-held knowledge by measurement professionals, every state assessment designer knows that they need to produce scores beyond the total score otherwise stakeholders would complain they are not getting enough from the assessment. Recall that producing very reliable total scores is critical for accountability uses of statewide assessments and, all things being equal, the reliability is related to the number of questions (or score points) on a test . Therefore, most measurement experts recommend having at least 10 score points for each subscore with to achieve at least some minimal level of reliability, so that statewide summative tests tend to get longer to accommodate subscore reporting. Therefore, one way to lessen the time required on the statewide summative assessment is to focus the summative assessment on reporting the total score and use the optional modules for districts that would like more detailed and accurate information about particular aspects of the content domain.

APPENDIX F: POTENTIAL QUALIFYING PRODUCTS/VENDORS

The Task Force put a premium on ensuring assessment quality, practical usefulness of assessment data, and on state-provided assessments not being exclusive to Wyoming. At the same time, the Task Force and the State Board of Education at its September 23, 2015 meeting expressed concern about whether the recommendations in this report may unreasonably reduce the number of potential qualified bidders. While the Task Force presents these companies as potential bidders, this in no way means that the company would either respond to a Wyoming RFP or that they would be able to meet the requirements of the RFP**.** Any potential Wyoming assessment vendor would have to provide evidence that their product can meet the requirements outlined in the RFP.

Language Arts and Mathematics

Table 6.1 below presents the potential companies and products would be likely or possibly available for Language Arts and Mathematics. This information is based on the knowledge of the two facilitators as a result of their work in other states and knowledge of the industry.

Table 6.1. *Likely and possibly qualifying products.*

Source	Type of Source	Status as of Spring 2015
ACT Aspire	Test Vendor	Administered in 2015 in two (2) states
Data Recognition Corporation	Test Vendor	Ready for use
Educational Testing Service	Test Vendor	Under development
Measured Progress	Test Vendor	Under development
PARCC	Consortium of States	Administered in 2015 in eleven (11) states
Smarter Balanced	Consortium of States	Administered in 2015 in eighteen (18) states
University of Kansas	State University	Administered in 2015 in two (2) states
Utah	State sells test items	Administered in 2015 in two (4) states

Based on Table 6.1, it appears that there are sufficient sources of likely and possibly qualifying products to assure that there is adequate and competitive bidding. We list in red some potential sources in Table 6.1 even though (1) no documentation is currently available for the products they have developed or are in the process of developing, and (2) no other state is currently using products from those sources for statewide summative assessment. We include these potential sources because by the time a request for proposals (RFP) is issued, these vendors may have adequate documentation and their products may have been adopted by at least one other state.

Finally, for Language Arts and Mathematics there are a few additional important considerations about collaboration with each potential source that may be probed in an RFP and in scoring bids on the RFP. Wyoming must consider the degree of control it wants in any new assessment system. Several of the potential products—such as ACT Aspire, University of Kansas, and Utah—would afford Wyoming very little, if any, control over the assessment program. On the other hand, if Wyoming becomes a governing member of an assessment consortium (PARCC or Smarter Balanced), it may have a limited amount of influence over the nature of the assessment system. In either case, Wyoming may extend its influence by convincing other states of the importance of its position and together with other states recommend a change to the assessment program. It is unknown to what degree DRC, ETS, and Measured Progress would afford clients control over their

products, but they would be proprietary products over which final decisions would rest with the vendors.

Second, the division of labor differs across potential assessment providers. In the case of ACT Aspire, PARCC, University of Kansas, and likely DRC, ETS, and Measured Progress, the assessment provider is solely responsible for product development and for test administration, scoring, and reporting; and the state is responsible for overseeing contract performance. Smarter Balanced is responsible for product development and monitoring consistency across member states and states are responsible for procuring a state-specific vendor for test administration, scoring, and reporting and for monitoring the contract performance of that vendor. On the other hand, PARCC manages all assessment activities centrally. States such as Florida, Tennessee, and Arizona have purchased the rights to use Utah test items in 2015, but there is no cross-state collaboration beyond that financial transaction.

Science

Science is addressed separately because whereas there is considerable similarity of the Wyoming state standards in Language Arts and Mathematics to those of many other states, the Wyoming state standards in science are unique. Therefore, there may or may not be sources with qualified products (meaning that an exclusive Wyoming science assessment may be needed). The potential assessment options available for science will depend on the new science content standards adopted by the Wyoming State Board of Education.

Of the sources listed in Table 6.1, ACT Aspire, Utah, and the University of Kansas offer science assessments. The DRC, ETS, and Measured Progress products may include science assessments when they become available. PARCC and Smarter Balanced products do not include science assessments.

The Task Force recommended keeping the existing science assessment until new Wyoming science standards have been adopted, but that the RFP issued for a new assessment system include requirements to immediately begin development of a new science assessment consistent with the recommendations in this report when the new Wyoming state science standards are adopted. They further recommended that collaboration with other states with sufficiently similar science standards be investigated as a first option. Finally, the Task Force recommended that depending on the instructional shifts required by any new science standards, the state may choose to adjust the timing of a new science assessment to best accommodate the required instructional shifts.