Oklahoma Assessment Report: Oklahoma State Department of Education Recommendations for House Bill 3218

Prepared for the Oklahoma State Department of Education (OSDE) and Oklahoma State Board of Education (OSBE) by the National Center for the Improvement of Educational Assessment, Inc.

Draft: October 24, 2016

By:
Juan D'Brot, Ph.D.
Erika Hall, Ph.D.

with contributions from
Scott Marion, Ph.D.
Joseph Martineau, Ph.D.

Contents

Executive Summary

The Oklahoma Legislature directed the State Board of Education (OSBE) to evaluate Oklahoma's current state assessment system and make recommendations for its future. As a result, the Oklahoma State Department of Education (OSDE) held regional meetings across the state and convened the Oklahoma Assessment and Accountability Task Force to deliberate over many technical, policy, and practical issues associated with implementing an improved assessment system. The 95 Task Force members met four times between August 4 and October 18, 2016. This report presents the results of those deliberations in the form of recommendations from the OSDE to the State Board.

Purpose of this Report

This report addresses the requirements stated in House Bill 3218, provides an overview of key assessment concepts, describes the role of the Task Force, and presents the recommendations made by the OSDE. Additionally, this report provides considerations relevant to the recommendations made by the State Department, which are presented in the full body of the report.

House Bill 3218

In June of 2016, Oklahoma Governor Mary Fallin signed House Bill 3218 (HB 3218), which relates to the adoption of a statewide system of student assessments. HB 3218 required the OSBE to study and develop assessment recommendations for the statewide assessment system. The House Bill specifically tasks the OSBE, in consultation with representatives from the Oklahoma State Regents for Higher Education, the Commission for Educational Quality and Accountability, the State Board of Career and Technology Education, and the Secretary of Education and Workforce Development, to study and develop assessment requirements. Additionally, HB 3218 requires the State Board to address accountability requirements under ESSA, which will be presented in a separate report for accountability. This report focuses specifically on the assessment requirements of HB 3218, which include the degree to which the Oklahoma assessment

- aligns to the Oklahoma Academic Standards (OAS);
- provides a measure of comparability among other states;
- yields both norm-referenced and criterion-referenced scores;
- has a track record of statistical reliability and accuracy; and
- provides a measure of future academic performance for assessments administered in high school.

Collecting Feedback from Regional Engage Oklahoma Meetings and the Oklahoma Task Force

Prior to convening Oklahoma's Assessment and Accountability Task Force, the OSDE held regional meetings at Broken Arrow, Sallisaw, Durant, Edmond, Woodward, and Lawton. These meetings yielded responses on various questions addressing the desired purposes and types of assessments. This regional feedback was incorporated in the discussions with the Oklahoma Assessment and Accountability Task Force. The Task Force included 95 members who represented districts across the state, educators,

parents, business and community leaders, tribal leaders, and lawmakers. Additionally, members from the Oklahoma State Regents for Higher Education, the Commission for Educational Quality and Accountability, the State Board of Career and Technology Education, and the Secretary of Education and Workforce Development were also represented on the Task Force. For a complete list of Task Force members, please refer to Appendix A of this report.

On four separate occasions the members of the Task Force met with experts in assessment and accountability to consider each of the study requirements and provide feedback to improve the state's assessment and accountability systems. Two of those experts also served as the primary facilitators of the Task Force: Juan D'Brot, Ph.D., from the National Center on the Improvement of Educational Assessment (NCIEA) and Marianne Perie, Ph.D., from the University of Kansas' Achievement and Assessment Institute. These meetings occurred on August 4 and 5, September 19, and October 18, 2016. At each meeting, the Task Force discussed the elements of HB 3218, research and best practices in assessment and accountability development, and feedback addressing the requirements of HB 3218. This feedback was subsequently incorporated into OSDE's recommendations to the OSBE.

Key Summative Assessment Recommendations

Oklahoma's Assessment and Accountability Task Force and the OSDE recognized that assessment design is a case of optimization under constraints[1]. 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 and prioritize desired characteristics and intended uses for a new Oklahoma statewide summative assessment for OSDE to consider.

Upon consolidating the uses and characteristics, the facilitators returned to the Task Force with draft goals for the assessment system. The Task Force provided revisions and input to these goals. Facilitators then presented the final goals to the Task Force. Once goals were defined, the desired uses and characteristics were clarified within the context of the Task Force's goals. The members of the Task Force agreed to the following goals for OSDE to consider for Oklahoma's assessment system:

1. Provide instructionally useful information to teachers and students with appropriate detail (i.e., differing grain-sizes for different stakeholder groups) and timely reporting;
2. Provide clear and accurate information to parents and students regarding achievement and progress toward college- and career-readiness (CCR) using an assessment that is meaningful to students;
3. Provide meaningful information to support evaluation and enhancement of curriculum and programs; and
4. Provide information to support federal and state accountability decisions appropriately.

Following discussion of the Oklahoma assessment system's goals, the Task Force worked with the facilitators to articulate feedback for the grade 3-8 and high school statewide summative assessments.

---

1 See Braun (in press).

This feedback was subsequently incorporated into the OSDE's recommendations to the State Board. These recommendations are separated into those for grades 3-8 and those for high school.

Recommendations for Assessments in Grades 3-8

The feedback provided by the Task Force and subsequently incorporated by the OSDE for grades 3-8 can be grouped into four categories: Content Alignment and Timing, Intended Purpose and Use, Score Interpretation, and Reporting and State Comparability. The OSDE's recommendations are presented below.

Content Alignment and Timing

- Maintain the focus of the new assessments on the Oklahoma Academic Standards (OAS) and continue to administer them at the end of grades 3 through 8; and
- Include an adequate assessment of writing to support coverage of the Oklahoma English Language Arts (ELA) standards.

Intended Purpose and Use

- Ensure the assessment can support calculating growth for students in at least grades 4-8 and explore the potential of expanding growth to high school depending on the defensibility of the link between grade 8 and high school assessments and intended interpretations; and
- Ensure the assessment demonstrates sufficient technical quality to support the intended purposes and current uses of student accountability (e.g., promotion in grade 3 based on reading and driver's license requirements on the grade 8 ELA assessments).

Score Interpretation

- Provide a measure of performance indicative of being on track to CCR, which can inform preparation for the Oklahoma high school assessment;
- Support criterion-referenced interpretations (i.e., performance against the OAS) and report individual claims including but not limited to scale score[2], Lexile[3], Quantile[4], content cluster[5], and growth[6] performance; and
- Provide normative information to help contextualize the performance of students statewide such as intra-state percentiles.

---

[2] A scale score (or scaled scores) is a raw score that has been transformed through a customized set of mathematical procedures (i.e., scaling and equating) to account for differences in difficulty across multiple forms and to enable the score to represent the same level of difficulty from one year to the next.

[3] A score developed by MetaMetrics that represents either the difficulty of a text or a student's reading ability level.

[4] A score developed by MetaMetrics that represents a forecast of or a measure of a student's ability to successfully work with certain math skills and concepts.

[5] A content cluster may be a group of items that measure a similar concept in a content area on a given test.

[6] Growth can be conceptualized as the academic performance of the same student over two or more points in time. This is different from improvement, which is change in performance over time as groups of students matriculate or when comparing the same collection of students across time (e.g., Grade 3 students in 2016 and Grade 3 students in 2015).

Reporting and State Comparability

- Support aggregate reporting on claims including but not limited to scale score, Lexile, Quantile, content cluster, and growth performance at appropriate levels of grain-size (e.g., grade, subgroup, teacher, building/district administrator, state); and
- Utilize the existing National Assessment of Educational Progress (NAEP) data to establish statewide comparisons at grades 4 and 8. NAEP data should also be used during standard setting[7] activities to ensure the CCR cut score is set using national and other state data.

Recommendations for Assessments in High School

The feedback provided by the Task Force and subsequently incorporated by the OSDE can be grouped into four categories: Content Alignment and Timing, Intended Purpose and Use, Score Interpretation, and Reporting and State Comparability. The OSDE's recommendations are presented below.

Content Alignment and Timing

- Use a commercial off-the-shelf college-readiness assessment (e.g., SAT, ACT) in lieu of state-developed high school assessments in grades 9 or 10; and
- Consider how assessments measuring college-readiness can still adequately address assessment peer review requirements, including but not limited to alignment.

Intended Purpose and Use

- Ensure the assessment demonstrates sufficient technical quality to support the need for multiple and differing uses of assessment results.
- Explore the possibility of linking college-readiness scores to information of value to students and educators (e.g., readiness for post-secondary, prediction of STEM readiness, remediation risk); and
- Ensure that all students in the state of Oklahoma can be provided with a reliable, valid, and fair score, regardless of accommodations provided or the amount of time needed for a student to take the test. Ensure that scores reflecting college-readiness can be provided universally to the accepting institution or employer of each student.

Score Interpretation

- Support criterion-referenced interpretations (i.e., performance against the OAS) and report individual claims appropriate for high school students;
- Provide evidence to support claims of CCR. These claims should be (1) supported using theoretically related data in standard setting activities (e.g., measures of college-readiness and other nationally available data) and (2) validated empirically using available post-secondary data linking to performance on the college-readiness assessment; and
- Provide normative information to help contextualize the performance of students statewide such as intra-state percentiles.

---

[7] The process through which subject matter experts set performance standards, or cut scores, on an assessment or series of assessments.

Reporting and State Comparability
- Support aggregate reporting on claims at appropriate levels of grain-size for high school assessments (e.g., grade, subgroup, teacher, building/district administrator, state); and
- Support the ability to provide norm-referenced information based on other states who may be administering the same college-ready assessments, as long as unreasonable administration constraints do not inhibit those comparisons.

Key Considerations for Summative Assessment Recommendations

While the Task Force addressed a targeted set of issues stemming from HB 3218, the facilitators were intentional in informing Task Force members of three key areas that must be considered in large-scale assessment development and/or selection:

1. **Technical quality**, which serves to ensure the assessment is reliable, valid for its intended use, and fair for all students;
2. **Peer Review**, which serves as a means to present evidence of technical quality; and
3. **Accountability**, which forces the issue of intended purpose and use.

In the time allotted, the Task Force was not able to consider all of the constraints and requirements necessary to fully expand upon their feedback to the OSDE. The facilitators worked to inform the Task Force that the desired purposes and uses reflected in their feedback would be optimized to the greatest extent possible in light of technical- and policy-based constraints[8]. As historically demonstrated, we can expect that the OSDE will continue to prioritize fairness, equity, reliability, and validity as the agency moves forward in maximizing the efficiency of Oklahoma's assessment system. A more detailed explanation of the context and considerations for adopting OSDE's recommendations is provided in the full report below.

Conclusion

The conversations that occurred between Task Force members, assessment and accountability experts, and the OSDE resulted in a cohesive set of goals for an aligned comprehensive assessment system which includes state and locally-selected assessments designed to meet a variety of purposes and uses. These goals are listed on page 9 of this report. The feedback provided by the Task Force and the recommendations presented by the OSDE, however, are focused only on Oklahoma's statewide summative assessments.

While the OSDE's recommendations can be grouped into the four categories of (1) Content Alignment and Timing, (2) Intended Purpose and Use, (3) Score Interpretation, and (4) Reporting and State Comparability, it is important to understand how these recommendations address the overarching requirements outlined in HB 3218.

Alignment to the OAS. Summative assessments used for accountability are required to undergo peer review to ensure the assessments are reliable, fair, and valid for their intended uses. One such use is to measure student progress against Oklahoma's college- and career-ready standards. The Task Force and

---

8 See Braun (in press).

department believe it is of vital importance that students have the opportunity to demonstrate their mastery of the state's standards. However, there is also a perceived need to increase the relevance of assessments, especially in high school. The Task Force and OSDE believe a state-developed set of assessments for grades 3-8 and a college-readiness assessment in high school would best support teaching and learning efforts in the state.

Comparability with other states. Throughout feedback sessions, Task Force meetings, and OSDE deliberations, the ability to compare Oklahoma performance with that of other states was considered a valuable feature of the assessment system. However, there are tensions among administration constraints, test design requirements, and the strength of the comparisons that may make direct comparisons difficult. Currently, Oklahoma can make comparisons using statewide aggregated data (e.g., NAEP scores in grades 4 and 8, college-readiness scores in grade 11), but is unable to support comparisons at each grade. Task Force feedback and OSDE recommendations suggest leveraging available national comparison data beyond its current use and incorporating it into assessment standard setting activities. This will allow the OSDE and its stakeholders to determine CCR cut scores on the assessment that reflect nationally competitive expectations.

Norm-referenced and criterion-referenced scores. Based on Task Force feedback, the OSDE confirmed that reported information supporting criterion-referenced interpretations (e.g., scale score, Lexile, Quantile, content cluster, and growth performance) are valuable and should continue to be provided in meaningful and accessible ways. Additional feedback and OSDE's recommendations note that norm-referenced interpretations would enhance the value of statewide summative assessment results by contextualizing student learning and performance. By working with a prospective vendor, the OSDE should be able to supplement the information provided to stakeholders with meaningful normative data based on the performance of other Oklahoma students.

Statistical reliability and accuracy. The technical quality of an assessment is an absolute requirement for tests intended to communicate student grade-level mastery and for use in accountability. The Standards for Educational and Psychological Testing[9] present critical issues that test developers and test administrators must consider during assessment design, development, and administration. While custom state-developed assessments require field testing and operational administration to accumulate evidence of statistical reliability and accuracy, the quality of the processes used to develop those assessments can be easily demonstrated by prospective vendors and the state. In contrast, off-the-shelf assessments should already have evidence of this and the state can generalize their technical quality if the assessment is given under the conditions defined for the assessment. Thus, the technical quality of an assessment is a key factor in ensuring assessment results are reliable, valid, and fair.

Future academic performance for assessments administered in high school. As noted earlier in the report, there is a clear value in high school assessment results being able to predict future academic performance. Based on OSDE's recommendation of using a college-readiness assessment in high school, the state and its prospective vendor should be able to determine the probability of success in early post-

---

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

secondary academics based on high school assessments. However, the state and its prospective vendor should amass additional Oklahoma-specific evidence that strengthens the claims of likely post-secondary success. This can be supported both through standard setting activities and empirical analyses that examine high-school performance based on post-secondary success.

The recommendations made to the OSDE in the previous section offer relatively fine-grain suggestions that can be interpreted through the lens of the HB 3218 requirements. These recommendations also reflect the Task Force's awareness of the three areas of technical quality, peer review requirements, and accountability uses, which were addressed throughout deliberations. Through regional meetings and in-depth conversations with the Task Force, the OSDE was able to critically examine the feedback provided and present recommendations to support a strong statewide summative assessment that examines the requirements of HB 3218 and seeks to maximize the efficiency of the Oklahoma assessment system in support of preparing students for college and careers.

Limitations of this Report

The OSDE and Task Force acknowledged that there are many other assessments that comprise the Oklahoma assessment system, including the Alternative Assessment on Alternate Achievement Standards (AA-AAS), the English Language Learner Proficiency Assessment (ELPA), and the many assessments that make up the career and technical assessments. However, the Task Force did not address these assessments in this report for two main reasons. First, the focus placed on the Task Force was to address the requirements of HB 3218 specific to the state summative assessment. While the goals defined by the Task Force go beyond the scope of the House Bill, they are important in framing OSDE's recommendations specific to the statewide summative assessment. Second, the time frame for making these recommendations and issuing this report was compressed. The OSDE devoted considerable effort in a short amount of time to arrive at these recommendations through regional feedback meetings and by convening the Task Force within the specified deadline. Therefore, it may be prudent for the OSDE to examine more specific aspects of this report with small advisory groups that include representation from the original Task Force.

Introduction

The Oklahoma Legislature directed the State Board of Education (OSBE) to evaluate Oklahoma's current state assessment system and make recommendations for its future. As a result, the Oklahoma State Department of Education (OSDE) held regional meetings across the state and convened the Oklahoma Assessment and Accountability Task Force 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 OSDE's recommendations to the State Board.

Purpose of this Report

As part of the response to House Bill 3218, the OSBE was tasked with studying a variety of requirements for Oklahoma's assessment and accountability system. This report addresses the requirements stated in House Bill 3218, provides an overview of key assessment concepts, describes the role of the Task Force, and presents the recommendations made by the OSDE. Additionally, this report provides considerations relevant to the recommendations made by the OSDE.

House Bill 3218

In May of 2016, the Oklahoma Legislature approved House Bill 3218 (HB 3218), which relates to the adoption of a statewide system of student assessments. HB 3218 required for the OSBE to study and develop assessment recommendations for the statewide assessment system.

The House Bill specifically tasks the OSBE, in consultation with representatives from the Oklahoma State Regents for Higher Education, the Commission for Educational Quality and Accountability, the State Board of Career and Technology Education, and the Secretary of Education and Workforce Development, to study assessment requirements and develop assessment recommendations. Additionally, HB 3218 requires the State Board to address accountability requirements under ESSA, which is presented in a separate report for accountability. The House Bill study notes the following requirements should be examined by the State Board for both assessment and accountability:

- A multi-measures approach to high school graduation;
- A determination of the performance level on the assessments at which students will be provided remediation or intervention and the type of remediation or intervention to be provided;
- A means for ensuring student accountability on the assessments which may include calculating assessment scores in the final or grade-point average of a student; and
- Ways to make the school testing program more efficient.

The House Bill also specifies additional requirements for assessment that the Board should examine as part of the study. These include an assessment that

- aligns to the Oklahoma Academic Standards (OAS);
- provides a measure of comparability among other states;
- yields both norm-referenced and criterion-referenced scores;

- has a track record of statistical reliability and accuracy; and
- provides a measure of future academic performance for assessments administered in high school.

Convening the Oklahoma Assessment and Accountability Task Force

In response to the HB 3218 requirements, the OSDE convened an Assessment and Accountability Task Force that included representatives from the those noted on page 20 of the House Bill: students, parents, educators, organizations representing students with disabilities and English language learners, higher education, career technology education, experts in assessment and accountability, community-based organizations, tribal representatives, and business and community leaders. For a complete list of Task Force members, please refer to Appendix A of this report.

The role of the Task Force was to deliberate over the assessment and accountability topics required in the House Bill and provide feedback that the OSDE would incorporate into their recommendations to the State Board. The Task Force was comprised 95 members who met with experts in assessment and accountability to consider each of the study requirements and make recommendations to improve the state's assessment and accountability systems. Two of those experts also served as the primary facilitators of the Task Force: Juan D'Brot, Ph.D., from the National Center on the Improvement of Educational Assessment (NCIEA) and Marianne Perie, Ph.D., from the University of Kansas' Achievement and Assessment Institute.

The Task Force met four times to discuss best practices in assessment and accountability and to provide feedback informing OSDE's recommendations to the State Board. These meetings occurred on August 4, August 5, September 19, and October 18, 2016. Throughout these meetings, the Task Force discussed HB 3218, the role of the Task Force, research and best practices in assessment and accountability development, and feedback addressing the requirements of HB 3218. This feedback was subsequently incorporated into OSDE's recommendations to the OSBE.

Feedback from Regional Meetings and the Oklahoma Task Force

Prior to convening Oklahoma's Assessment and Accountability Task Force, the OSDE held regional meetings at Broken Arrow, Sallisaw, Durant, Edmond, Woodward, and Lawton. These meetings yielded responses on various questions addressing the desired purposes and types of assessments. This regional feedback was incorporated into the discussions with the Oklahoma Assessment and Accountability Task Force. Additional information on House Bill 3218 can be found on OSDE's website: http://sde.ok.gov/sde/hb3218.

The Task Force includes 95 members who represent districts across the state, educators, parents, and lawmakers (for a complete list of Task Force members, please refer to Appendix A of this report) and met four times to address the assessment. The August meeting served primarily as an introduction to the requirements of the House Bill and to the issues associated with assessment and accountability design. Task Force members were also introduced to the *Every Student Succeeds Act* (ESSA), a bipartisan measure that reauthorized the *Elementary and Secondary Education Act* (ESSA), and ESSA's requirements for statewide educational systems. The August meeting also served as a foundational

meeting that allowed the Task Force members to identify the primary goals of the assessment system. The September meeting served as an opportunity to clarify the goals of the Task Force and provide specific feedback that directly addressed the House Bill requirements. The October meeting was used to finalize the feedback from the Task Force and discuss next steps for the OSDE to develop recommendations for the OSBE.

Throughout the four meetings, Task Force members engaged in discussion that addressed the varied uses, interpretations, and values associated with the state's assessment system. These discussions were used to establish and refine the Task Force's feedback, which were subsequently incorporated into the OSDE's recommendations. The final recommendations are presented in the section titled OSDE Recommendations for Oklahoma's Assessment Recommendations, which can be found in the full report.

Considerations for Developing an Assessment System

Before presenting OSDE's recommendations in response to House Bill 3218, we first provide some critical definitions and necessary context.

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 of teaching strategies and 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 there are many different uses of assessment results, however for many uses the distinction between categories is blurred. For example, many of the appropriate uses of assessment introduced below may fall into both broad categories. We present a further distinction of assessments based on the appropriate use of those assessments below. These distinctions include formative, summative, and interim assessments.

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[10] because of the direct relevance to the Task Force's work. The facilitators provided a similar overview to the Task Force members to focus feedback

---

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

on the statewide summative assessment. We define and outline the appropriate uses of the three types of assessment below.

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[11] 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;
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.

---

[11]See Sadler (1989).

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 standardized summative assessments may include school and district accountability, curriculum/program evaluation, monitoring educational trends, and informing policymakers and other stakeholders. Depending on their alignment to classroom instruction and the timing of the administration and results, summative assessments may also 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. These assessments are often labeled "mid-terms," "final projects," "final papers," or "final exams" in middle and high school grades. Elementary school classrooms have similar types of summative assessments but they tend not to be referenced using a consistent label. 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. Depending on the conditions of assessments, results may be comparable within and across classrooms, schools, districts, or even states.

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

- 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 assessment 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.

The Role and Timing of Assessments in Relation to Standards and Instruction

Throughout conversations with the Assessment and Accountability Task Force, the facilitators defined and described the assessments types and uses presented here to ensure members had a shared understanding of assessment. To address the specific requirements of HB 3218, the Task Force only focused on the role and uses of *summative* assessments—specifically, the state summative assessment for accountability. To further explore the role of state summative assessments, the Task Force spent time discussing the role and timing of these assessments in the educational system.

Given the backwards-looking nature of the information gleaned from statewide summative assessments and their potential uses (e.g., evaluate achievement, monitor progress over time, support accountability), it is important to understand how these assessments follow standards and instruction. However, after-the-fact assessment results can be used to inform adjustments to curriculum that may lead to revisions in instruction. That is, once standards are developed and adopted, curriculum aligned to those standards is implemented, which helps inform teachers' instruction to those standards.

The statewide summative assessment must also be aligned to those standards to inform educators whether students are making progress against grade-level expectations. Depending on the results of the assessments, educators then determine whether any adjustments to curriculum or instruction are necessary to support student learning. However, the assessment is dependent on the state standards and great efforts are taken to determine the facets of the standards that are most appropriate to assess. This process is described in more detail in the next section.

The Assessment Development Process

As described to the Task Force, the assessment development process must begin with a clarification of the uses and purposes of the assessment. In the case of Oklahoma's state summative assessment, the assessments must provide evidence of student proficiency of grade-level standards, inform progress toward college- and career-readiness (CCR), and support student and school accountability. A detailed description of the major goals established in light of the Task Force's suggested uses is provided in the OSDE Recommendations section of this report.

In order to appropriately frame the OSDE's recommendations, it is important to consider the general steps that are necessary to develop an assessment. Those steps include, but are not necessarily limited to the following[12]--depending on the uses of the assessment:

1. Develop assessment specifications, which are based upon: the state's academic standards, detailed specifications about the learning objectives that support the standards, and the rules dictating requirements for test content, format, and accessibility for all students;
2. Develop and review assessment materials, which include item development guides, scoring rubrics, graphic design requirements, a verification of content and standard alignment, and score report requirements;

---

[12] Adapted from DRC|CTB (2016).

3. Conduct pilot tests, usability studies (to ensure ease of use by students and educators), tryout studies (to confirm consistent and accurate scoring if relevant), and bias and sensitivity reviews (to ensure content is validly and fairly represented for all students);

4. Conduct field tests to determine how well items are performing, that items effectively represent the content being assessed, and that items can be accessed fairly and appropriately by all students;

5. Produce final assessment materials, which include final test versions, reports for educators and students, and supporting information/data that helps contextualize test results to those consuming reports from the test such as administrative manuals and interpretative guides;

6. Administer, score, and report student performance using the final version of the tests; and

7. Engage in ongoing evaluation of the assessment system to ensure the assessment is meeting the goals of the system and to determine if any refinements or revisions to improve its quality and effectiveness are needed.

While these can be considered a general set of steps for assessment development, there may be additional or fewer steps depending on the intended uses of the assessment results. Although this report focuses only on Oklahoma's summative assessment, there are additional components of an assessment system that may provide a more comprehensive view of student performance and school quality (e.g., locally-selected assessments, assessments common across districts, or classroom developed assessments and formative practices). Those additional components may include all, a subset, or additional steps than those listed here.

OSDE Recommendations for Oklahoma's Assessment

Oklahoma's Assessment and Accountability Task Force and the OSDE recognized that assessment design is a case of optimization under constraints[13]. In other words, there may be many desirable purposes, uses, and goals for assessment, but some of them 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 determining assessment design and specifications. Therefore, a critical early activity of the Task Force was to identify and prioritize desired characteristics and intended uses for a new Oklahoma statewide summative assessment for OSDE to consider.

It is important to note that the Task Force recognized that Oklahoma's assessment system should have a wider set of goals, but the feedback in response to HB 3218 should be focused around the statewide summative assessment. The following section describes the process through which the Task Force established goals and provided feedback to the OSDE. This feedback was incorporated into OSDE's recommendations to the State Board, which is included later in this section.

---

13 See Braun (in press).

Assessment Goals based on Desired Characteristics and Uses

Task Force members initially were asked to ignore constraints, and identify their highest priority purposes for assessment and their desired uses and characteristics of assessment results. Task Force members, working in small groups, identified their highest priority uses, and shared their thoughts with other smaller groups. After the first meeting, the proposed uses and characteristics were consolidated and returned to the Task Force as a set of draft goals for the assessment system. The Task Force provided revisions and feedback to these goals. Facilitators then presented the final goals to the Task Force for confirmation. Once goals were defined, the desired uses and characteristics were articulated within the context of the Task Force's feedback to the required study points of the House Bill. The members of the Task Force agreed to the following goals for OSDE to consider for Oklahoma's assessment system:

1. Provide instructionally useful information to teachers and students with appropriate detail (i.e., differing grain-sizes for different stakeholder groups) and timely reporting;
2. Provide clear and accurate information to parents and students regarding achievement and progress toward CCR using an assessment that is meaningful to students;
3. Provide meaningful information to support evaluation and enhancement of curriculum and programs; and
4. Provide information to support federal and state accountability decisions appropriately.

An important outcome of this process is that no single type of assessment (formative, interim, or summative) can accommodate all of the uses and characteristics represented in the Task Force's goals. In fact, to accomplish the full set of desired uses and characteristics, a system of assessments would be required that span across assessment types (formative, interim, and summative) and levels (classroom, district, and state). This can be accomplished by combining state and local assessments to create a coherent system and eliminating unnecessary assessment. Furthermore, this is in addition to the more process-oriented formative assessment practices that are synonymous with instruction. However, those needs exceed the scope of examination required as part of the response to HB 3218. The OSDE should continue to work with the State Board and those representatives on the Task Force to address the needs stated by the Task Force that are beyond the scope of the statewide summative assessment.

OSDE Recommendations: Addressing Intended Goals

In developing recommendations for a new state summative assessment, the OSDE considered the Task Force's deliberation of issues presented in HB 3218. As a reminder, the areas to be studied specific to the statewide assessment included

- ensuring alignment to the OAS
- providing a measure of comparability among other states
- demonstrating a track record of statistical reliability and accuracy
- yielding both norm-referenced and criterion-referenced scores
- providing a measure of future academic performance for assessments administered in high school

To address these areas, Task Force members were asked to respond to specific questions that focused on each of the above-stated points, any relevant development or administration constraints that may need to be considered, and the intended and unintended consequences that might be associated with the assessment's uses.

Task Force members, working in small groups, reflected upon each of the major areas presented in the House Bill and shared their thoughts with other smaller groups. Facilitators compiled and consolidated input into feedback distinguished by grades 3-8 and high school assessments. Upon consolidating feedback, the facilitators returned to the Task Force with draft feedback statements for 3-8 and high school assessments. The Task Force provided revisions and edits to these feedback statements. Facilitators then presented the final feedback statements to the Task Force for confirmation. This feedback was then reviewed by the OSDE and incorporated into recommendations for the State Board.

What follows is a brief description of the recommendations grouped by category. The recommendations are separated into recommendations for grades 3-8 assessments and recommendations for high school assessments.

Recommendations for 3-8 statewide assessments

The feedback provided by the Task Force and subsequently incorporated by the OSDE for grades 3-8 can be grouped into four categories: Content Alignment and Timing, Intended Purpose and Use, Score Interpretation, and Reporting and State Comparability. Following each set of recommendations, a brief discussion on the context of and considerations for adopting these recommendations is provided.

Content Alignment and Timing

The following recommendations are presented for Content Alignment and Timing:

- Maintain the focus of the new assessments on the Oklahoma State Standards and continue to administer them at the end of grades 3 through 8; and
- Include an adequate assessment of writing to support coverage of the Oklahoma English Language Arts (ELA) standards.

The Task Force members made it apparent that the assessments in grades 3-8 should maximize the amount of instruction available to students by administering the assessments at the end of each grade. Additionally, the Task Force recognized that the Oklahoma ELA standards included expectations of writing for students and that the assessment should reflect those standards. The OSDE should explore ways in which they can continue to support educators and administrators in ensuring the assessment is administered in the most efficient manner to support learning opportunities for students.

Intended Purpose and Use

The following recommendations are presented for Intended Purpose and Use:

- Ensure the assessment can support calculating growth for students in at least grades 4-8 and explore the potential of expanding growth to high school depending on the defensibility of the link between grade 8 and high school assessments and intended interpretations;

- Ensure the assessment demonstrates sufficient technical quality to support the intended purposes and current uses of student accountability (e.g., promotion in grade 3 based on reading and driver's license requirements on the grade 8 ELA assessments).

The Task Force recognized the need for the assessment to communicate progress toward CCR, but that students may differ in their degree of progress toward CCR. As a result, the Task Force believed that it is important for the assessment to support the calculation of growth across years and potentially growth to standard (i.e., the required growth to reach or maintain grade-level expectations). While this is something that the OSDE is already considering, the Department should explore the multiple options available in calculating growth that may or may not require the use of vertical scales to inform educators of student progress over time.

Additionally, Task Force members were aware of the potentially conflicting intended purposes and uses of the assessment at grades 3 and 8. That is, using a single assessment as both a signal for CCR and as a signal for minimum competency can lead to mixed messages. While the OSDE currently uses a subscore specific to grade 3 for reading (i.e., Reading Sufficiency Act Status), it will be important to examine how the assessments are used in policy to identify potential systematic problems. The OSDE should continue exploring how policy decisions can help mitigate any unintended consequences associated with using assessments signaling CCR for student accountability.

Score Interpretation

The following recommendations are presented for Score Interpretation:

- Provide a measure of performance indicative of being on track to CCR, which can inform preparation for the Oklahoma high school assessment;
- Support criterion-referenced interpretations (i.e., performance against the OAS) and report individual claims including but not limited to scale score, Lexile, Quantile, content cluster, and growth performance; and
- Support normative information to help contextualize performance of students statewide using something such as intra-state percentiles.

The Task Force deliberated for some time regarding how scores should be interpreted. The two key areas of discussion included interpretations in support of progress toward CCR and interpretations to help contextualize performance. With regard to CCR interpretations, clearly articulating how students perform against the state standards was critical. Furthermore, because the OAS are reflective of students being college and career ready upon graduation from high school, the grade-level interpretations should reflect whether students are on-track for CCR (assuming the cut score for grades 3-8 is informed using data that reflects CCR-like expectations). However, sufficient information should be reported at the individual level to help students and educators understand progress against the state standards. This contextualization should extend to providing within-state normative information that may include percentiles of performance, like-student performance, or like-school performance data. The OSDE should explore the types of within-state normative information their prospective vendors could provide to the public through reporting.

Reporting and State Comparability

The following recommendations are presented for Reporting and State Comparability:

- Support aggregate reporting on claims including but not limited to scale score, Lexile, Quantile, content cluster, and growth performance at appropriate levels of grain-size (e.g., grade, subgroup, teacher, building/district administrator, state); and
- Utilize the existing National Assessment of Educational Progress (NAEP) data to establish statewide comparisons at grades 4 and 8. NAEP data should also be used during standard setting[14] activities to ensure the CCR cut score is set using national and other state data

The Task Force also wrestled with the best way to support statewide reporting and comparisons to other states. It was evident to Task Force members that the same information reported at the student level should be reported in the aggregate. Specifically, information made available to students and their guardians should be aggregated (at the school, district and state-level) and provided to educators, administrators, and the public. The OSDE should continue to explore meaningful ways to report information clearly and publically when working with their prospective vendor.

How to support state by state comparisons was less straightforward. Members generally agreed that there was significant value in understanding how Oklahoma students perform in comparison to students in other states. There was less agreement, however, with regard to the level of granularity necessary to support those comparisons. That is, some Task Force members believed that comparisons would be most valuable at each grade (and in some cases by student); whereas other members believed comparisons were sufficient at the state level.

Upon further examination of this issue, the facilitators noted the technical requirements necessary to make state to state comparisons at varying units of analysis (e.g., student, subgroup, school, grade, district, state). Once the Task Force members became aware of the additional requirements (e.g., embedded field-test items, additional testing time, cost, similar testing administration conditions, use of nationally-normed tests) and the potential limitations of the interpretations based on various approaches, the perceived value of fine-grained comparisons diminished. Ultimately, Task Force members generally agreed that the *system* of assessments should support state-to-state comparisons of performance. That is, the statewide summative assessment may not serve that purpose, but other assessments in Oklahoma's assessments system (e.g., NAEP) are intended to serve this purpose.

Additionally, the information gleaned from Oklahoma's participation in NAEP can be extended to inform nationally-relevant expectations of student performance on the statewide summative assessment. This can be done by leveraging existing methodologies[15] using NAEP data that can be applied to Oklahoma's standard setting activities. This process can inform standard setting participants of how Oklahoma student performance compares to other states across the country. The OSDE should explore the

---

[14] The process through which subject matter experts set performance standards, or cut scores, on an assessment or series of assessments.

[15] See Jia, Phillips, Wise, Rahman, Xu, Wiley, & Diaz (2014) and Phillips (2009).

inclusion of national comparison data into standard setting activities with their prospective vendor and determine the level of rigor to which Oklahoma's CCR cut score should be aligned.

Recommendations for Assessments in High School

The feedback provided by the Task Force and subsequently incorporated by the OSDE can be grouped into four categories: Content Alignment and Timing, Intended Purpose and Use, Score Interpretation, and Reporting and State Comparability. Following each set of recommendations, a brief discussion on the context of and considerations for adopting these recommendations is provided.

Content Alignment and Timing

The following recommendations are presented for Content Alignment and Timing:

- Use of a commercial off-the-shelf college-readiness assessment (e.g., SAT, ACT) in lieu of state-developed high school assessments in grades 9 or 10; and
- Consider how assessments measuring college-readiness can still adequately address assessment peer review requirements, including but not limited to alignment.

Building off of the conversation in grades 3-8, the Task Force recognized the inherent value in signals of CCR. To that end, the Task Force members believed strongly that the state should consider the adoption of a commercial off-the-shelf college-readiness assessment. However, Task Force members were made aware that large-scale statewide assessments must adequately pass peer review requirements[16]. One of these requirements includes demonstrating that statewide assessments demonstrate sufficient alignment to the full range of the State's grade-level academic content standards[17].

The statewide summative assessment has to support several purposes. For example, Oklahoma's high school assessment must be aligned to the standards that students are taught by the year students are assessed (e.g., 11th grade), should reflect evidence of student learning in state's the accountability system, and serve as a signal of CCR. While an off-the-shelf college-readiness assessment will readily provide evidence of claims of college-readiness, it may be more difficult to amass evidence the assessment sufficiently reflects the OAS to support claims of grade-level mastery and progress toward Oklahoma's conceptualization of CCR. As a result, the OSDE will need to explore the degree to which different off-the-shelf college-readiness assessments will demonstrate sufficient alignment and what, if any, augmentation may be necessary to satisfy peer review requirements. To that end, the OSDE should continue to be involved in thoughtful discussion with other states and contacts familiar with peer review requirements. This will help inform expectations of prospective vendors with regard to alignment and additional peer review requirements for college-readiness assessments.

---

[16]Peer review requirements are requirements that have been developed by the U.S. Department of Education that support ESSA's requirement that that each State annually administer high-quality assessments in at least reading/language arts, mathematics, and science that meet nationally recognized professional and technical standards. Peer review involves states receiving feedback from external experts and the Department on the assessments it is using to meet ESEA requirements.

[17]See U.S. Department of Education (2015)

Intended Purpose and Use

The following recommendations are presented for Intended Purpose and Use:

- Ensure the assessment demonstrates sufficient technical quality to support the need for multiple and differing uses of assessment results;
- Explore the possibility of linking college-readiness scores to information of value for students and educators (e.g., readiness for post-secondary, prediction of STEM readiness, remediation risk); and
- Ensure that all students in the state of Oklahoma can be provided with a reliable, valid, and fair score, regardless of accommodations provided or the amount of time needed for a student to take the test. Ensure that scores reflecting college-readiness can be provided universally to the accepting institution or employer of each student

Like the recommendations presented in grades 3-8, Task Force members were aware of the challenges associated with using assessments for multiple purposes. Given the critical focus placed on signals of CCR for high school students, unintended consequences may be best avoided through the operationalization of the accountability system to ensure schools are recognized for progress in student learning. The OSDE should continue working to avoid potential negative unintended consequences in developing an ESSA accountability system.

One of the potentially negative unintended consequences that the Task Force discussed was associated with college-readiness scores and information of value. A primary reason why so many Task Force members were interested in the use of an off-the-shelf college-readiness assessment was the immediate value it added to students by providing a score that would be recognized by post-secondary institutions as an indicator of readiness. However, Task Force members were aware of the current challenges associated with providing an institution-recognized score to those students who received accommodations or if the assessment administration conditions were markedly different from those required by an off-the-shelf provider. Thus, it is important for the OSDE to ensure that advocacy viewpoints are reflected in conversations with prospective vendors to support the provision of reliable, valid, and fair scores to all students in the state of Oklahoma.

It is important to note that a small minority (i.e., two of the 95-member Task Force) believed it would be valuable to have a grade-level assessment aligned to the OAS rather than an off-the-shelf college-readiness assessment.

Score Interpretation

The following recommendations are presented for Score Interpretation:

- Support criterion-referenced interpretations (i.e., performance against the OAS) and report individual claims appropriate for high school students;
- Provide evidence to support claims of CCR. These claims should be (1) supported using theoretically related data in standard setting activities (e.g., measures of college readiness and other nationally available data) and (2) validated empirically using available post-secondary data linking to performance on the college-readiness assessment; and

- Provide normative information to help contextualize the performance of students statewide such as intra-state percentiles.

Like the recommendations for grades 3-8, the Task Force discussed the most important interpretations that should be supported for the high school assessments. Given the recommendations under Intended Purpose and Use, it should come as no surprise that Task Force members prioritized claims of CCR. However, claims of student performance should also reflect progress against the state standards. Like recommendations for grades 3-8, sufficient information should be reported at the individual level to help students and educators understand progress against the state standards, which may include within-state normative information. The OSDE should explore the types of within-state normative information their prospective vendors could provide to the public through reporting.

Aligned with the previous set of recommendations for high school, the OSDE will need to work with their prospective vendor to ensure that the high school assessment can support both a CCR and standards-based claim for students. These CCR-based claims should also be further validated using empirical evidence within the state of Oklahoma and using any available national data depending on the vendor.

Reporting and State Comparability
The following recommendations are presented for Reporting and State Comparability:

- Support aggregate reporting on claims at appropriate levels of grain-size for high school assessments (e.g., grade, subgroup, teacher, building/district administrator, state); and
- Support the ability to provide norm-referenced information based on other states who may be administering the same college-ready assessments, as long as unreasonable administration constraints do not inhibit those comparisons.

The feedback provided by the Task Force for statewide reporting was similar to those for grades 3-8. That is, aggregate reporting should reflect the same types of information that are provided at the individual level and aggregate information should be provided to educators, administrators, and the public in meaningful and easily accessible ways.

Given the Task Force's suggestion to adopt an off-the-shelf college-readiness assessment, Task Force members recommended that the OSDE work to support state-to-state comparisons. The availability of students across states potentially being administered the same items and test forms (i.e., depending on the selected vendor) allows for the possibility of direct comparisons of college-readiness. However, the Task Force members recognized the potential challenges that might be associated with changes in test administration practices that may be required to support fair administration for all students in Oklahoma. In other words, national comparisons were believed to be important, but those comparisons of CCR should not require unreasonable administration constraints. The OSDE should ensure that any prospective vendor be very clear in the kinds of comparisons that can be supported when considering Oklahoma-specific administration practices.

Key Areas of Importance to Consider

While the Task Force addressed a targeted set of issues stemming from House Bill 3218, the facilitators were intentional in informing Task Force members of three key areas of importance that must be considered in large-scale assessment development:

1. **Technical quality**, which serves to ensure the assessment is reliable, valid for its intended use, and fair for all students;
2. **Peer Review**, which serves as a means to present evidence of technical quality; and
3. **Accountability**, which forces the issue of intended purpose and use.

In the time allotted the Task Force was not able to consider all of the constraints and requirements necessary to fully expand upon their feedback to the OSDE. The facilitators worked to inform the Task Force that the desired purposes and uses reflected in their feedback would be optimized to the greatest extent possible in light of technical- and policy-based constraints[18]. As historically demonstrated, we can expect that the OSDE will continue to prioritize fairness, equity, reliability, and validity as the agency moves forward in maximizing the efficiency of Oklahoma's assessment system.

Conclusion

The conversations that occurred between Task Force members, assessment and accountability experts, and the OSDE resulted in a cohesive set of goals for an aligned comprehensive assessment system which includes state and locally-selected assessments designed to meet a variety of purposes and uses. These goals are listed on page 9 of this report. The feedback provided by the Task Force and the recommendations presented by the OSDE, however, are focused only on Oklahoma's statewide summative assessments.

While the OSDE's recommendations can be grouped into the four categories of (1) Content Alignment and Timing, (2) Intended Purpose and Use, (3) Score Interpretation, and (4) Reporting and State Comparability, it is important to understand how these recommendations address the overarching requirements outlined in HB 3218.

Alignment to the OAS. Summative assessments used for accountability are required to undergo peer review to ensure the assessments are reliable, fair, and valid for their intended uses. One such use is to measure student progress against Oklahoma's college- and career-ready standards. The Task Force and department believe it is of vital importance that students have the opportunity to demonstrate their mastery of the state's standards. However, there is also a perceived need to increase the relevance of assessments, especially in high school. The Task Force and OSDE believe a state-developed set of assessments for grades 3-8 and a college-readiness assessment in high school would best support teaching and learning efforts in the state.

Comparability with other states. Throughout feedback sessions, Task Force meetings, and OSDE deliberations, the ability to compare Oklahoma performance with that of other states was considered a

---

18 See Braun (in press).

valuable feature of the assessment system. However, there are tensions among administration constraints, test design requirements, and the strength of the comparisons that may make direct comparisons difficult. Currently, Oklahoma can make comparisons using statewide aggregated data (e.g., NAEP scores in grades 4 and 8, college-readiness scores in grade 11), but is unable to support comparisons at each grade. Task Force feedback and OSDE recommendations suggest leveraging available national comparison data beyond its current use and incorporating it into assessment standard setting activities. This will allow the OSDE and its stakeholders to determine CCR cut scores on the assessment that reflect nationally competitive expectations.

Norm-referenced and criterion-referenced scores. Based on Task Force feedback, the OSDE confirmed that reported information supporting criterion-referenced interpretations (e.g., scale score, Lexile, Quantile, content cluster, and growth performance) are valuable and should continue to be provided in meaningful and accessible ways. Additional feedback and OSDE's recommendations note that norm-referenced interpretations would enhance the value of statewide summative assessment results by contextualizing student learning and performance. By working with a prospective vendor, the OSDE should be able to supplement the information provided to stakeholders with meaningful normative data based on the performance of other Oklahoma students.

Statistical reliability and accuracy. The technical quality of an assessment is an absolute requirement for tests intended to communicate student grade-level mastery and for use in accountability. The Standards for Educational and Psychological Testing[19] present critical issues that test developers and test administrators must consider during assessment design, development, and administration. While custom state-developed assessments require field testing and operational administration to accumulate evidence of statistical reliability and accuracy, the quality of the processes used to develop those assessments can be easily demonstrated by prospective vendors and the state. In contrast, off-the-shelf assessments should already have evidence of this and the state can generalize their technical quality if the assessment is given under the conditions defined for the assessment. Thus, the technical quality of an assessment is a key factor in ensuring assessment results are reliable, valid, and fair.

Future academic performance for assessments administered in high school. As noted earlier in the report, there is a clear value in high school assessment results being able to predict future academic performance. Based on OSDE's recommendation of using a college-readiness assessment in high school, the state and its prospective vendor should be able to determine the probability of success in early post-secondary academics based on high school assessments. However, the state and its prospective vendor should amass additional Oklahoma-specific evidence that strengthens the claims of likely post-secondary success. This can be supported both through standard setting activities and empirical analyses that examine high-school performance based on post-secondary success.

The recommendations made to the OSDE in the previous section offer relatively fine-grain suggestions that can be interpreted through the lens of the HB 3218 requirements. These recommendations also reflect the Task Force's awareness of the three areas of technical quality, peer review requirements, and

---

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

accountability uses, which were addressed throughout deliberations. Through regional meetings, advisory group meetings, input in response to posted questions, and in-depth conversations with the Task Force, the OSDE was able to critically examine the feedback provided and present recommendations to support a strong statewide summative assessment that examines the requirements of HB 3218 and seeks to maximize the efficiency of the Oklahoma assessment system in support of preparing students for college and careers.

DRAFT

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.* National Council on Measurement in Education. Washington, DC.

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

Data Recognition Corporation | CTB (2016). Designing assessment systems: A primer on the test development process. Retrieved September 1, 2016, from https://ctb.com/ctb.com/control/assetDetailsViewAction?currentPage=3&articleId=895&assetType=article&p=library

Jia, Y., Phillips, G., Wise, L.L., Rahman, T., Xu, X., Wiley, C., Diaz, T.E. (2014). *2011 NAEP-TIMSS Linking Study: Technical Report on the Linking Methodologies and Their Evaluations* (NCES 2014-461). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

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.

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.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). Knowing What Students Know: The Science and Design of Educational Assessment. Washington, DC. Retrieved September 21, 2016, from http://www.nap.edu/openbook.php?record_id=10019&page=R1.

Phillips, G. W (2009), *The Second Derivative: International Benchmarks in Mathematics for U.S. States and School Districts.* American Institutes for Research. Washington, DC.

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 October 5, 2016, from
http://www.cehd.umn.edu/NCEO/onlinepubs/synthesis44.html.

U.S. Department of Education (2015). *Non-regulatory guidance for states for meeting requirements of the Elementary and Education Act of 1965, as amended*. U.S. Department of Education. Washington, D.C.

Wiley, E. C. (2008). *Formative Assessment: Examples of Practice*. Retrieved October 1, 2016, from http://ccsso.org/documents/2008/formative_assessment_examples_2008.pdf.

DRAFT

Appendix A: Task Force Representation

Name	Organization	Title
Hofmeister, Joy	State Dept. Education	State Superintendent of Public Instruction
Dunlap, Katie Dr.	State Dept. Education	Deputy Superintendent of Assessment and Accountability
Tamborski, Michael	State Dept. Education	Executive Director of Accountability
Walker, Craig	State Dept. Education	Executive Director of State Assessments
Barnes, Lynn	Oklahoma City Public Schools	Sr. Executive Dir of Curriculum & Federal Programs
Bax, Benjamin	American Federation of Teachers	Field Representative
Baxter, Leo J.	State Board of Education of Oklahoma	Board Member
Bendick, Debbie Dr.	Edmond Public Schools	Assoc. Superintendent
Best, Mary	American Federation of Teachers	President
Bishop, Katherine	Oklahoma Education Association	Vice President
Blanke, Debbie Dr.	Oklahoma State Regents for Higher Education	Academic Affairs
Burchfield, Rocky	Fairview Public Schools	Superintendent
Burk, Jana	Tulsa Public Schools	Executive Director of Teacher/Leadership Effectiveness Initiative
Bushey, Brent	Oklahoma Public School Resource Center	Executive Director
Buswell, Robert	Office of Educational Quality and Accountability	Director of Educational Accountability
Caine, Ann	Oklahoma State School Boards Association	Director of Education Leadership
Capps, Staci	Byng Public Schools	Curriculum Director/Grant Developer
Casey, Dennis Rep.	Oklahoma House Representatives	Oklahoma House Representative
Charney, Randee	Research Associate	Schusterman Family Foundation
Choate, Tony	Chickasaw Nation	Media Relations
Cobb, Rick	Mid-Del Schools	Superintendent
Condit, Donnie Rep.	Oklahoma House of Representatives	Oklahoma House Representative
Cook, H. Gary Dr.	University of Wisconsin	Associate Scientist, Expert in Assessment and Accountability, E.L.L.
Cooper, Donna	Choctaw Nicoma Park Schools	Asst. Superintendent
D'Brot, Juan Dr.	Center for Assessment	Senior Associate, Expert in Assessment and Accountability
DeBacker, Terri Dr.	University of Oklahoma College of Education	Assoc. Dean

Name	Organization	Title
Dossett, J.J. Sen.	Oklahoma Senate	Oklahoma Senator
Dugan, Drew	Greater Oklahoma City Chamber	Vice President
Dunlop, Janet Dr.	Broken Arrow Public Schools	Assoc. Superintendent
Dunn, Kathy	Mid-Del Schools	Asst. Superintendent for Teaching and Learning
Elam, Mary Dr.	Oklahoma City Public Schools	Senior Research Associate, Planning, Research, and Evaluation Dept.
Fedore, Stephen	Tulsa Public Schools	Director of Data Quality and Data Use
Flanagan, William	State Board of Education of Oklahoma	Board Member
Font, Raul	Latino Community Dev Agency	CEO/Executive Director
Ford, John Sen.	Oklahoma Senate	Oklahoma Senator
Foster, Becki	Oklahoma Department of Career and Technology Education	Associate State Director for Curriculum, Assessment, Digital Delivery and Federal Programs
Franks, Cathryn	State Board of Education of Oklahoma	Board Member
Fulton, Lisa	Ada City Schools	District Test Coordinator
Garn, Gregg A. Dr.	University of Oklahoma	Dean of Education
Grunewald, Angela	Edmond Public Schools	Executive Director of Elementary Education
Guerrero, Julian Jr.	Tribal Education Dept. National Assembly (TEDNA)	Project Director, Native Youth Community Project
Heigl, Brenda	Oklahoma Parent Teacher Association	President
Henke, Katie Rep.	Oklahoma House of Representatives	Oklahoma House Representative
Hernandez, Kristy	Moore Public Schools	Director of Student Services
Hime, Shawn	Oklahoma State School Boards Association	Executive Director
Hooper, Tony	Lawton Public Schools	Director of Accountability and Assessment
House, Sharon	Oklahoma Parents Center, Services for Families of Children with Disabilities	Executive Director
Hutchinson, Tony	Oklahoma State Regents for Higher Education	Strategic Planning Analysis Workforce and Economic Dev
Keating, Daniel	State Board of Education of Oklahoma	Board Member
Lepard, Jennifer	Oklahoma State Chamber	V.P. of Government Affairs
Lester, Erin	Tulsa Public Schools	Director of Educational Indicators
Lora, Aurora	Oklahoma City Public Schools	Superintendent
Love, Courtney	Oklahoma Virtual Charter Academy	Operations Manager
Mack, Marcie	Oklahoma Department of Career	State Director

Name	Organization	Title
	and Technology Education	
McDaniel, Tracy	KIPP Charter Oklahoma City	Founding School Leader & Principal
Monies, Jennifer	Oklahoma Educated Workforce Initiative	Executive Director
Mouse, Melanie Dr.	Putnam City Schools	Asst. Superintendent
Muller, Lisa Dr.	Jenks Public Schools	Asst. Superintendent
Nollan, Jadine Rep.	Oklahoma House Representatives	Oklahoma House Representative
Ogilvie, Clark	Owasso Public Schools	Superintendent
Owens, Beecher	Mannford HS	2016 Graduate
Owens, Rick	Lawton Public Schools	Secondary Education
Owens, Ryan	CCOSA	Co-Executive Director/General Counsel; Director Legislative Services
Parks, Tammy	Howe Public Schools	PDC Coordinator
Parrish, Jim	Choctaw Nation	Executive Director of Education
Pennington, David	Ponca City Public Schools	Superintendent
Perie, Marianne Dr.	University of Kansas	Director Achievement and Assessment Institute; Expert in Assessment and Accountability
Pittman, Anatasia Sen.	Oklahoma Senate	Oklahoma Senator
Polk, Jamie	Lawton Public Schools	Asst. Superintendent
Price, Bill	State Board of Education of Oklahoma	Board Member
Priest, Alicia	Oklahoma Education Association	President
Reavis, Madison	Muskogee HS	2016 Graduate
Riggs, Ruthie	Edmond Public Schools	Assoc. Superintendent
Roberts, Kuma	Tulsa Regional Chamber	Education Program Manager
Roberts, Sarah	Inasmuch Foundation	Senior Program Officer
Rogers, Rep. Michael	Oklahoma House of Representatives	Oklahoma House Representative
Roman Nose, Quinton	Tribal Education Departments National Assembly (TEDNA)	Executive Director, Board of Directors
Ross, Robert	Inasmuch Foundation & State Board of Education of Oklahoma	Board of Directors, Board Member
Sadler, Kimberly	Oklahoma Department of Career and Technology Education	Associate State Director for Curriculum, Assessment, Digital Delivery and Federal Programs
Shirley, Natalie	OK Governor's Office	Secretary of Education and Workforce Development
Simmons, Shirley Dr.	Norman Public Schools	Asst. Superintendent
Shouse, Jerrod	Owner	Shouse Consulting
Sly, Gloria Dr.	Cherokee Nation	Education Liaison Education Services
Stanislawski,	Oklahoma Senate	Oklahoma Senator

Name	Organization	Title
Gary Sen.		
Stoycoff, Zack	Tulsa Regional Chamber	Government Affairs Director
Tatum, Sheryl	Oklahoma Virtual Charter Academy	Head of School
Taylor, Etta	Oklahoma Parent Teacher Association	President Elect
Thompson, Shannon	Moore Public Schools	Dean of Academics
Thomsen, Todd Rep.	Oklahoma House Representatives	Oklahoma House Representative
Tinney, Ginger	Professional OK Educators	Executive Director
Trent, Sean	Mid-Del Schools	Executive Director of Academic Services & Technology
Viles, Susan	Woodward Schools	District Test Coordinator/RSA Test Coordinator
Weeter, Richard Dr.	Oklahoma City Public Schools	Executive Director of Planning, Research, and Evaluation Dept.
Woodard, Johanna Dr.	Owasso Public Schools	Coordinator of Academic Services
Woodard, Petra	Millwood Public Schools	High School Principal
Yunker, Jake	Oklahoma Governor's Office	Deputy Policy Director

Appendix B: 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 ELA summative assessments produce a total score for ELA, 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., grade 5 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 the grade 5 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 Oklahoma's next statewide summative assessment, which means 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 in spite of user reports 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 they need to produce scores beyond the total score otherwise stakeholders would complain they are not getting enough from the assessment. Recall, 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 to achieve at least some minimal level of reliability, so 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.