Considerations for Streamlining Nevada's K–12 Assessments: Addendum to the 2018/2019 Assessment Reports

Submitted to:
Peter Zutz

Administrator of Assessment, Data, & Accountability Management

Submitted by:
WestEd

February 15, 2019

WestEd
WestEd.org

Preface

During 2018 and early 2019, WestEd produced a series of reports for the Nevada Department of Education (NDE) that examined the state's assessment system through multiple lenses. WestEd researchers evaluated assessments in terms of how well the assessments achieved their desired objectives. Items from science assessments at grades 5, 8, and high school were examined for alignment to the Nevada Academic Content Standards for Science. Likewise, items from the Nevada Alternate Assessment (NAA) for English language arts and mathematics at grades 3 through 8 and 11 were examined for alignment to the Nevada Academic Content Standards. Furthermore, for three statewide assessment programs — ACT, Nevada Science Assessments, and the NAA program — WestEd reviewed validity and reliability documentation according to the critical elements of technical quality specified in the U.S. Department of Education (ED) assessment peer review guidelines. Finally, Nevada's assessment system was placed into the broader context of the assessment systems of all other states. Across these reports, WestEd presented findings that inform the NDE and policymakers about options for improving the effectiveness and efficiency of the state assessment system. In February 2019, NDE requested that WestEd pull together considerations for improving and streamlining Nevada's K–12 assessments into a single report. This addendum is the result of that synthesis.

Addendum Outline

In this addendum, we summarize current research to present six characteristics of high-quality assessment systems. These six criteria provide a useful framework for evaluating the quality of any state and local assessment system. Next, we describe a set of seven research-based principles useful for informing the selection of individual assessments and assessment strategies that collectively make up a high-quality system of assessments. We reviewed Nevada's statewide system of assessments against these principles to inform specific recommendations for improving the quality of individual state assessments and for streamlining and improving the system's efficiency. Our recommendations are presented below the principle to which they apply. These recommendations focus attention on efficiency while ensuring that other elements of a high-quality assessment system are not compromised.

Limitations

A state's role in supporting balanced assessment systems can vary widely. While all states must ensure that districts comply with the federal Every Student Succeeds Act (ESSA) requirements, the extent to which states support or intervene in local assessment decision-making is often a function of federal or state law, the state department's interpretation of the law through state code, or stakeholder input that drives statewide support decisions. Marion (2018) indicates that balanced assessment systems are often best addressed at the local level, due to limits on state authority and districts' need to oversee districtwide assessments.

Because WestEd's evaluation focused on Nevada's statewide assessment system, our recommendations for streamlining the system pertain only to assessments currently required by the state. This report provides guidance that districts may find useful for achieving balance within their local assessment systems; however, district-specific recommendations for streamlining assessment is outside the scope of this project.

Essential Characteristics of a High-Quality Assessment System

An effective system of assessments should adhere to a set of principles that inform the selection of individual assessments needed to address specific needs. Additionally, decisions about whether or not to include individual assessments must be considered within the context of the system as a whole. For example, an interim reading assessment can be administered at the beginning of a school year to screen students who may be struggling and require additional support. Follow-up assessment may be necessary to diagnose and address specific issues preventing a struggling reader from progressing at an expected pace. However, what if the child's teacher already has sufficient information from prior assessment data to reliably diagnose and address this student's needs? In this case, further diagnostic testing may be redundant and unnecessary, regardless of the quality and usefulness of the results generated from additional testing.

In the first part of this review, we focus on the whole system, using research to develop a framework for evaluating systems of assessments. The framework describes six characteristics of effective systems of assessments, which can be used to guide decisions about who, when, and how often to assess, which assessment to use, and how to know whether additional assessment is needed. In the second part of this review, we describe a set of principles that are useful for making assessment selection decisions when information gaps emerge or when an agency (state or district) or school is considering whether to adopt or discard a particular assessment to create a more streamlined system.

Coherence

Coherence occurs when a system's component parts work together in a logical and consistent way to produce a clear and efficient whole. To achieve coherence, curriculum, instruction, and assessment must be built upon and grounded in a robust theory of student learning. Theories explaining how students learn are used to develop a well-articulated learning model and learning progressions that describe how knowledge and skills build on one another as students are exposed to new instruction and experiences (Chattergoon & Marion, 2016; Conley, 2018). These learning progressions can, in turn, be used to inform decisions to ensure strong alignment across curriculum and curriculum support materials, instructional strategies, and assessment. Ideally, such decisions produce classroom-level instructional systems in which curriculum, instruction, and assessment are fully integrated, as opposed to separate and disconnected tasks. Coherence is achieved when decisions about curriculum, instruction, and assessment occur in tandem: decisions about curriculum and instruction are informed by and align with the outcomes that students should produce and that assessments will measure.

Comprehensiveness

A comprehensive system includes the full range of assessment approaches needed to accomplish a multitude of purposes for a multitude of groups. Because assessments have different purposes and stakeholders have different needs, a wide range of assessments must be available to meet the needs of diverse stakeholders (e.g., parents, teachers, leaders, policymakers, students). Moreover, each assessment should contribute unique and essential information, resulting in a complete understanding of who students are, how they learn, and what they know and can do (Conley, 2018; Sigman & Mancuso, 2017).

In a comprehensive system, the information needed by stakeholders at one level of the system should be useful to stakeholders at other levels. For example, state assessment results used by state department staff to monitor achievement gaps and identify the lowest performing schools should also be useful for districts and schools to monitor achievement performance and growth, evaluate school programs, and inform curricular and instructional decisions. A recent article published by the Center on Standards and Assessment Implementation at WestEd (Sigman & Mancuso, 2017) provides a helpful summary of the types of assessment needed in any assessment system to support educational decision-making. These

include four broad categories of assessment: formative, diagnostic, interim/benchmark, and summative.

Balance

Comprehensive systems of assessment incorporate the full range of assessments to address various purposes. A balanced system is one in which assessments and the information produced from them are available and ready for use by the right education professionals (for whom), in the right proportion (how often), at the right time (when), and for the right purpose (for which students or student subgroups).

Balance implies that each assessment should be used to optimize its utility within an integrated system. To achieve balance, stakeholders must understand how and when to use multiple measures to accomplish a specific purpose, as well as how these measures fit within an integrated system of improvement. Across levels of the system, stakeholders understand why each assessment in the system is necessary, how they should be used, who needs to use the results, for what purpose, and how assessment use may affect or influence others' decisions within the system.

Assessments are most helpful when they are used to address the limited set of purposes for which they were designed. Educators may occasionally be tempted to use assessments in ways that overreach their intended purposes, which can lead to frustration, inefficiencies, and inaccurate interpretation of results. For instance, end-of-course exams provide useful results for assessing the quality of curriculum, examining instructional rigor and consistency across classrooms, grades, and subgroups, or evaluating instructional programs. They are not useful for informing a teacher's instruction, particularly when results are often not available until after the school year ends. Similarly, a principal who reviews benchmark results with teachers quarterly to determine which students need additional support is making good use of benchmark assessment information. But a teacher who reviews last year's state test scores weekly to group students for Wednesday's phonics lesson will quickly become frustrated.

Efficiency

Efficiency is achieved when stakeholders have access to the full array of assessment tools and training to achieve their objectives, *and* when redundant, unused, and untimely assessments are eliminated from the system (Chattergoon & Marion, 2016; Conley & Darling-Hammond, 2013).

Assessment practice in the No Child Left Behind (NCLB) era presents an unfortunate example of inefficiencies that can emerge when one type of assessment is given more weight than it deserves. Since 2001, federal policy began to measure a school's success on annual standardized test results, relying on test results in mathematics and English language arts. The high stakes associated with these tests (under NCLB, lower than expected performance resulted in a "failing" label) influenced a narrowing of the curriculum and led many schools and teachers to focus excessively on test-taking skills and teaching to test items, as opposed to teaching for learning. Interim assessments that touted to predict state assessment results were introduced, along with supplemental diagnostic and weekly progress monitoring assessments for students who struggled to understand grade-level material. Often these interim and diagnostic assessments were not well-aligned to new college and career ready standards, which minimized their usefulness for guiding instruction. Formative and performance-based assessment strategies useful for guiding day-to-day curriculum and instructional decisions and assessing higher order skills took a back seat to test preparation. Districts and schools increasingly administered redundant tests (i.e., multiple exams in the same subject to the same students) because not all results yielded data by item, grade, subject, student, or school — prompting additional exams to get data at the desired level of granularity (Hart et al., 2015).

Educators at different levels of the system must work together to make sure that, as new measurement approaches are introduced, the system remains efficient, keeping balance in check. To achieve this,

decision makers at the state and local levels must understand the roles of various assessments, as well as their limitations. State policy makers can support assessment literacy locally through professional development and technical assistance. When such collaboration and capacity are in place, redundant, untimely, and unused assessments can be quickly identified and eliminated (or adjusted) to sustain balance and alignment across the system.

Alignment

An aligned system supports assessment balance (the right category of assessment is available and used at the right time for the right purpose) and assessment use *within the classroom, across levels of the system, and across the grade level continuum*. Alignment requires stakeholders at all levels of the system have access to a unique set of assessment tools (e.g., tests, rubrics, self-assessment tools), training, and technology to support student learning. Misalignment occurs when one link in the chain does not have the information, training, or support needed, which, in turn, will inevitably compromise the strength of the entire chain. For instance, while a teacher relies heavily on "in the moment" formative assessment practices to determine what individual students know and can do, principals and district staff rely more heavily on standardized information to examine subgroup performance (e.g., classrooms, grades, students with IEP's, minority subgroups) or identify students who may need more individualized support. Both types of assessment information are critical, and without either, the system does not work properly.

To maximize alignment, structures must be in place at each level to ensure that the proper assessment tools, resources, and training are in place to support teaching and learning. This necessitates a common set of objectives and coordinated communication within and across agencies (SEA, Regional Centers, LEAs, schools). Analytic tools developed by the SEA or LEA are most useful when they address the highest-priority needs in curriculum and instruction, accountability, gifted, special education, English learners (ELs), and school transformation departments. Additionally, the training provided to schools should be coordinated to ensure training across departments is aligned, administered with fidelity, at the proper dosage, and for the intended audience. Training in formative assessment may be appropriate for all teachers, while specific training to support the use of progress monitoring tools may be appropriate as a supplementary training for teachers who provide tier 3 supports. To ensure efficient dissemination of these tools and strategies, departments within the SEAs and LEAs must align their objectives and collectively determine what tools and training are necessary to support school and classroom needs (Childress, Elmore, & Grossman, 2006).

A ***theory of action*** is a prerequisite for a well-aligned assessment system. As Chattergoon and Marion (2016) suggest, "a set of assessments, even if they cohere, will not fulfill the intended purposes if the information never reaches the intended user." A theory of action guides implementation of the assessment system by specifying the system's purpose, resources, and inputs needed to produce intended outputs and outcomes. A well-articulated theory of action ensures that objectives are aligned across agencies and that staff at all levels are working efficiently to produce intended outcomes.

Flexibility

The extent to which an assessment system can achieve balance depends largely on the policy requirements that drive assessment decisions. Policies that are rigid tend to encourage standardization and a one-size-fits-all approach to assessment. Some degree of standardization is necessary and helpful. For example, stakeholders need standardized results to make valid comparisons, monitor growth, encourage fairness (e.g., standardizing accommodations for students with shared disabilities), and address inequities (e.g., using standardized results to make funding decisions). But too much standardization creates the opposite effect: increasing inefficiencies (over-testing by using overlapping assessments), inhibiting fairness, promoting inequities, and potentially stifling innovation.

ESSA (NCLB's replacement, which was signed into law in late 2015) offers more flexibility to states and addresses concerns about increases in standardized testing, narrowing of test methods, and narrowing of skills and abilities taught and tested under NCLB (Conley & Darling-Hammond, 2013). ESSA allows states to replace their state assessment with a nationally recognized, locally selected assessment (such as the ACT or SAT) to meet high school testing requirements; replace one summative assessment with multiple interim assessments that result in a single summative score; utilize performance assessment (portfolios, projects, or extended performance tasks) to "partially" measure higher order thinking skills; and set target limits on testing time. These flexibilities are designed to influence innovation and promote improved balance in state and local assessment systems. For example, states have more flexibility to introduce test methods, such as performance assessments, which can measure complex constructs embedded in states' new college and career readiness standards. This is especially important because traditional tests (including those with technology-enhanced items) cannot measure many important standards needed to succeed in college and careers. Conley and Darling-Hammond (2013) include specific examples of higher-order standards adopted in most states:

- Conducting extended research using multiple forms of evidence
- Communicating ideas – discussing or presenting orally or in multimedia formats
- Collaborating with others to define or solve a problem
- Planning, evaluating, and refining solution strategies
- Using mathematical tools and models in science, technology, and engineering contexts

More rigorous standards demand that assessment systems incorporate more flexible assessment methods to fully integrate curriculum, instruction, and assessment and support more efficient assessment practice. Striking the right balance between flexibility (experimenting with and scaling new or innovative assessments) and structure and consistency (retaining traditional assessment regimes) can be difficult. But states and districts willing to pilot and adopt alternative approaches to assessment will be much more likely to address common assessment challenges and sustain high-quality assessment systems.

Principles to Inform the Selection of Assessments and Assessment Strategies

In this section, we present a set of seven principles to inform the selection of individual assessments and assessment strategies that collectively make up a high-quality system of assessments. We reviewed Nevada's statewide system of assessments against these principles to inform specific recommendations for improving the quality of individual state assessments and for streamlining and improving the system's efficiency. Our recommendations are presented below the principle to which they apply. Recommendations to inform a more streamlined system of statewide assessments are included under principle #1, 3, 6, and 7. These recommendations focus attention on efficiency while ensuring that other elements of a high-quality assessment system are not compromised.

For any assessment system to work properly, each individual assessment within the system must meet specific criteria for serving the purpose it is intended to serve. That requires examining the quality of each individual assessment to ensure it produces valid, reliable, meaningful, and necessary information. For instance, is the assessment technically sound? Does it produce valid and reliable information to support the use for which it is intended? Does it provide information in a reasonable amount of time (Council of Chief State School Officers, 2015)?

In 2015, ED released the Testing Action Plan (TAP) fact sheet. The plan was released, in part, to address a crisis of over-testing students that was emerging across the country (CCSSO, 2015; Hart et al., 2015). The TAP included a set of seven principles to support state and district leaders in ensuring that their students take high-quality and thoughtfully selected assessments as part of a comprehensive system (Sigman & Mancuso, 2017). According to the TAP, assessments must be:

1. Worth taking
2. High-quality
3. Time-limited
4. Fair — and supportive of fairness — in equity and educational opportunity
5. Fully transparent to students and parents
6. Just one of multiple measures
7. Tied to improved learning

The seven principles outlined in the TAP provide a helpful framework for selecting the right set of assessments to include in a comprehensive and balanced system. Below is a summary of each principle, which can be used to inform the evaluation and selection of assessments to include in the larger system. Additionally, these principles can be used to evaluate, compare, keep, and/or discard existing assessments and thereby create a more efficient system.

1. **Worth taking:** An assessment should be aligned to content and skills a student is learning and should measure the same complex work students do in an effective classroom. It should also provide useful data to inform a student's learning needs and guide instruction.
2. **High-quality:** Assessments should measure knowledge and skills against state-developed college and career ready standards. Collectively over time, assessments should cover the full range of relevant state standards, elicit complex student demonstrations or applications of knowledge, provide valid and reliable results for all students, and provide an accurate measure of student growth.
3. **Time-limited:** States and school districts should carefully consider the extent to which each assessment serves a unique and essential role in the learning process. No child should spend more than two percent of her classroom time taking standardized tests. Test preparation strategies should be discouraged and limited, and low-quality test preparation strategies must be eliminated.
4. **Fair:** Assessments should include accessibility and accommodations for students with disabilities and ELs to accurately reflect what students really know and can do. The same assessments of core subjects (reading, writing, science, math) should be administered consistently statewide so that teachers and leaders have a clear picture of which students and/or schools are meeting expectations and which students and/or schools need additional support and interventions to succeed. Although most states administer standardized assessments annually, New Hampshire is experimenting with administering personalized competency assessments more frequently, with statewide standardized assessments administered every two to four years as a periodic check to ensure all students are making adequate progress. States and districts should also ensure that assessments are only used for the purposes for which they were intended and designed.

5. **Fully transparent:** States and districts should ensure that every parent gets understandable information about the assessments their students are taking. Information on any test students are required to take should include (1) the purpose, (2) the source of the requirement, (3) when results will be provided to parents and students, (4) how educators will use the results, and (5) how parents can use the results to help their child. Parents should also receive assessment results in a timely manner.
6. **Just one of multiple measures:** No single assessment should ever be the sole factor in making educational decisions about a student, educator, or school. Measures such as achievement, behavior, school climate, and others can provide a comprehensive understanding of students' needs and how schools are doing. Observations, surveys, and contributions to the school community can also be used to ensure a comprehensive evaluation of performance.
7. **Tied to improved learning:** The vast majority of assessments should be used to improve teaching and learning. Assessment outcomes should be used to identify what students know and to guide additional teaching, supports, or interventions that will help students master challenging material.

Recommendations for Streamlining Nevada's Statewide System of Assessments

WestEd applied these seven principles to Nevada's statewide assessment system and generated considerations for either modifying or eliminating the use of individual assessments to create a more streamlined system. These considerations are offered below.

The Nevada Department of Education administers, reports, and uses data from a variety of assessments to ensure compliance with federal ESSA laws. Although limited for instructional purposes, these assessments provide essential information the state needs for identifying the lowest-performing schools in need of comprehensive school improvement (CSI schools); identifying subgroup performance gaps and schools in need of targeted school improvement (TSI schools); directing resources for school turnaround and continuous improvement; evaluating the effectiveness of state and district programs; and informing school improvement strategies. In considering how the state can streamline their statewide system of assessments, WestEd inventoried assessments administered by the state, identified whether each is required or voluntary, and determined which were necessary to meet ESSA guidelines. Table 1 below summarizes this information using information from WestEd's national inventory of statewide assessments conducted for Nevada as part of this project.[1] Results highlight assessments that could potentially be eliminated to create a more streamlined assessment system.

As Table 1 shows, five of the ten assessments included in Nevada's statewide assessment system are required to meet ESSA accountability laws; two — SBAC Interims, and EOC Exams — are provided to support district planning and instruction, but the state does not require these assessments be administered. Three assessments — Brigance, MAP, and CTE — are required by the state to comply with state law.

[1] WestEd submitted a State of the States Report to the Nevada Department of Education in January 2019. The report presents a comparison of assessment models used across states to meet federal ESSA requirements.

Table 1: Assessments That Comprise Nevada's Statewide System of Assessments

Assessment	State Required	Federally Required	Notes
Smarter Balanced Assessment Consortium (SBAC) annual assessment	Yes	Yes	Required for ESSA accountability
ACT	Yes	Yes	Required for ESSA accountability
Nevada Science Assessments (grades 5, 8, and HS)	Yes	Yes	Required for ESSA accountability
WIDA assessments	Yes	Yes	Required for ESSA accountability
Nevada Alternate Assessment (NAA)	Yes	Yes	Required for ESSA accountability
Brigance Early Childhood Screens	Yes	No	Nevada requires students to be assessed upon entrance to kindergarten to identify individual student needs and track progress during kindergarten to grade 3, specifically regarding a student's literacy level.
Measures of Academic Progress (MAP)[2]	Yes	No	Nevada requires MAP administration as a part of the Read by Grade Three (RBG3) program (SB 391, effective July 1, 2015). This statute was designed to dramatically improve student achievement by ensuring all students will be able to read proficiently by the end of the 3rd grade.
Career and technical education (CTE) assessments	Yes	No	CTE exams measure the skill attainment of students who have completed a program course sequence. These assessments are required by Nevada Administrative Code 389.800.
End-of-course (EOC) assessments	No	No	As of the 2017–18 school year, Nevada no longer required EOCs for federal accountability purposes.
Smarter Balanced Assessment Consortium (SBAC) interim assessments	No	No	SBAC Interim Assessments are voluntary assessments at grades 3 through 8 provided to districts to support school planning and instruction.

To streamline assessments, the NDE can consider the following options:

1. Reduce the overall number of assessments required and used
2. Reduce the number of interim assessments administered at the district level
3. Encourage districts to implement more balanced assessment systems

[2] Nevada requires MAP administration at grades K–3. Although several Nevada districts implement MAP at higher grade levels, results and recommendations related to MAP in this report are specific to the K–3 MAP assessments.

The following recommendations attempt to leverage these three options as Nevada considers options for creating a more streamlined system of assessments at the state level and promoting more streamlined systems of assessment at the local level.

Findings from WestEd's assessment evaluation report suggest that the EOC Exams are of limited usefulness and overlap with other assessments currently administered at the local level. Additionally, interview results from district assessment directors suggest that several districts using the SBAC interims may also be administering one or more interim assessments that overlap in their respective purposes. And although Brigance, MAP, and CTE are required by state law, opportunity exists at the local level to streamline these assessments to increase assessment efficiency across Nevada districts.

Encourage districts to review existing summative tests in high school and eliminate EOC exams where appropriate.

Districts should consider administering only one common end-of-course exam that can be used for local accountability, evaluation, and school improvement purposes. Districts that are currently administering common end-of-course exams should make sure that these exams (1) adequately cover standards across the full range of knowledge levels and (2) meet minimum standards of technical quality (including minimum thresholds for establishing a range of reliability and validity criteria), and (3) provide useful results for district and school stakeholders. Districts with end-of-course exams that meet these criteria should consider eliminating state EOCs. The NDE can support this initiative by identifying and endorsing one or more summative assessments that can be used to assess EOC performance in high school course work, with priority given to core subject areas.

Encourage districts to eliminate the use of multiple interim assessments.

Districts need one interim assessment, but many include multiple interim assessments, particularly in grades 3–8.

Explore options for reducing the number of assessments required for the Read by Grade 3 Program.

The state may consider identifying and endorsing a list of interim and/or summative assessments and providing minimum cut scores for meeting Read by Grade 3 requirements. Such a list would likely reduce the number of assessments administered in districts that prefer one high-quality literacy assessment over another (e.g., MAP vs. SBAC interims).[3] Alternatively, if local educators are concerned about over-testing, the state could consider eliminating the Read by Grade 3 testing requirements in grades K–2. The state agency (e.g., NDE) could be charged with overseeing training to ensure that districts are monitoring students' literacy progress using one or more valid assessment tools that align with Read by Grade 3 objectives.

Consider eliminating the use of the Brigance Screens as a required assessment.

Although it is essential to assess kindergarten readiness and monitor progress, districts use a variety of tools to address this purpose. Instead of requiring Brigance, the state could consider vetting and endorsing a list of high-quality assessment tools that schools could use to monitor kindergarten readiness and monitor progress. As described above, endorsing a list of assessments provides districts with more options, making it easier for them to eliminate the Brigance if a similar high-quality alternative is more widely accepted and used.

[3] Currently, MAP must be administered twice in kindergarten and three times in grades 1 through 3.

Provide resources for NDE to support districtwide assessment literacy training and implement balanced assessment systems at the local level.

The high-stakes nature of most required annual summative achievement assessments, such as the SBAC and annual science assessments — combined with the considerable time it takes to administer them — may prompt Nevada educators to want more from statewide assessment results than the assessments may be designed to deliver. For example, DTD survey and interview results suggest that district educators want assessment data from these summative tests (as currently designed) to be more granular to inform instruction. Findings also indicate that educators want to reduce the amount of time and the number of test items required to administer these tests. These two findings — wanting shorter tests while, at the same time, wanting more granular results that are both reliable and valid — cannot be simultaneously addressed in a single annual statewide achievement test. Furthermore, annual summative achievement tests are most useful for evaluating school performance and student progress, and to inform schoolwide decisions about curriculum, instruction, and programming; these tests are not designed to support ongoing instructional decision-making.

These findings suggest the need for (1) a more balanced assessment system within school districts; (2) assessment literacy training for district leaders and staff to understand different types of assessments (i.e., formative, interim, diagnostic, summative) and how they should be used within the context of a balanced assessment system to support instruction and continuous improvement; and (3) stronger communication and coordination between NDE and Nevada school districts to clarify state assessment requirements and to clearly delineate the roles and responsibilities of NDE from those of local districts. To support these suggestions, NDE should consider how departments might restructure existing resources to focus on improving assessment system balance, data literacy training, and communication strategies. Additionally, the state legislature may consider allocating resources for NDE to hire or train staff and build internal capacity to effectively address these needs.

References

Chattergoon, R., & Marion, S. (2016). Not as easy as it sounds: Designing a balanced assessment system. *State Education Standard*, *16*(1), 6–9.

Childress, S., Elmore, R., & Grossman, A. (2006). How to manage urban school districts. *Harvard Business Review 11,* 55–68.

Conley, D. T. (2018). *The promise and practice of next generation assessment*. Cambridge, MA: Harvard Education Press.

Conley, D. T., & Darling-Hammond, L. (2013). *Creating systems of assessment for deeper learning.* Stanford, CA: Stanford Center for Opportunity Policy in Education.

Council of Chief State School Officers (CCSSO). (2015). Comprehensive statewide assessment systems: A framework for the role of the state education agency in improving quality and reducing burden.

Hart, R., Casserly, M., Uzzell, R., Palacios, M., Corcoran, A., and Spurgeon, L. (2015). Student testing in America's great city schools: An inventory and preliminary analysis. Washington DC: Council of Great City Schools. Retrieved September 2018 from https://www.cgcs.org/cms/lib/DC00001581/Centricity/Domain/87/Testing%20Report.pdf

Marion, S. F. (2018). The opportunities and challenges of a systems approach to assessment. *Educational Measurement: Issues and Practice, 37*(1), 45–48. Retrieved from https://onlinelibrary.wiley.com/doi/full/10.1111/emip.12193

New Hampshire Department of Education. (2013). Enriching New Hampshire's assessment and accountability systems through the quality performance assessment framework. New Hampshire: Author.

Sigman, D., & Mancuso, M. (2017). Designing a comprehensive assessment system. San Francisco, CA: WestEd.

U.S. Department of Education. (2015) Fact sheet: Testing action plan. Retrieved from https://www.ed.gov/news/press-releases/fact-sheet-testing-action-plan