Sound Basic Education for All

An Action Plan for North Carolina

WestEd

In collaboration with Learning Policy Institute and
The William & Ida Friday Institute for Educational Innovation

North Carolina's Statewide Assessment System

How Does the Statewide Assessment System Support Progress Toward Meeting the *Leandro* Requirements?

WestEd
WestEd.org

LEARNING POLICY INSTITUTE
Research. Action. Impact.

THE WILLIAM & IDA FRIDAY INSTITUTE
FOR EDUCATIONAL INNOVATION

Contents

Background

Context

A high-quality assessment system that provides useful and timely data on student growth and proficiency is an integral component for ensuring a sound basic education for all students. Results from high-quality assessments, coupled with a thoughtfully designed accountability system, can provide valuable information about students' academic progress and inform stakeholders when policies and practices are not working as intended. A high-quality assessment system needs to serve multiple purposes reflecting the needs of multiple stakeholder groups, providing crucial information to support progress toward a sound basic education across all levels of the broader education system.

A high-quality assessment system includes different types of assessments and processes, each used for different purposes at different levels of the system (i.e., state, district, school, and classroom). A single assessment is able to address only a limited number of purposes. Adopting a set of high-quality assessments alone, without understanding how they fit within a larger system that addresses a specific set of goals and purposes, creates inefficiencies that impede the needs of all stakeholder groups. For example, a district may identify and adopt a high-quality set of interim assessments that, on their own, achieve their intended purpose of helping the district's educators better understand whether their students are on track in their learning progress. However, if these interim assessments do not align with the existing statewide assessment system, they may create incoherence or imbalance, which can negatively impact student achievement. To prevent overassessment, the system must be planned carefully, which will likely entail making difficult decisions about which assessments to keep and which to do away with; ensuring staff work within the system; and training staff to understand the purposes of individual assessments, what they are designed to do, and how they complement and support other assessment practices within the system.

To meet such a diverse set of needs and purposes, an assessment system must be carefully planned, implemented, and monitored. Ongoing input and open communication across state and local education agencies and stakeholders, especially students and teachers, are necessary. Stakeholders and representatives within state and local education agencies are expected to bring their own unique experiences and perspectives about what assessments work best to address a given purpose. When implemented well, results generated from a high-quality assessment system will improve stakeholder decisions. Better decisions produce opportunities for deeper and more meaningful instruction, leading to holistic improvements in students' development and achievement.

This paper synthesizes recent literature to describe the essential characteristics of a high-quality assessment system and provides the approach, findings, and conclusions of the researchers' evaluation of North Carolina's existing assessment system.

Literature Review

The literature review establishes the framework that the researchers utilized to establish the approach and research questions, evaluate the findings, and draw conclusions about North Carolina's current assessment system. Based on current research about assessments and assessment processes, we identify the complex and diverse needs of stakeholders, describe six characteristics of high-quality assessment systems, and describe a set of principles that can be used to evaluate assessments when the state, a district, or a school is considering whether to adopt or implement a new assessment or modify existing assessments.

Complex and Diverse Needs of Stakeholders

A high-quality assessment system must work in concert with curriculum and instruction to serve multiple purposes for different stakeholders. Although not an exhaustive list, the examples below illustrate the complex and diverse needs of different stakeholder groups (Jobs for the Future, 2018).

- » **Students:** The system should support school environments in which students take personal responsibility for their academic, social, and emotional growth. To do this, students need to understand the long- and short-term expectations of their education system and develop strategies and habits of mind necessary to reach and exceed those expectations. A robust assessment system works in concert with curriculum and instruction to support students as they learn how to set and adjust learning goals, monitor progress, think metacognitively, and apply creative solutions to fulfill their potential.

- » **Families:** Parents and caregivers want to know whether their child is thriving academically, socially, and emotionally. They want their child's teachers and school staff to understand their child's strengths and weaknesses and to rely on objective evidence to ensure that he or she is on track for success in school and in life. Parents often rely on normative data to understand how their child is performing relative to his or her peers and to the broader population. Ongoing standardized interim/benchmark and summative standards-based assessments, such as end-of-year tests, keep parents informed about their child's progress, helping them advocate for additional services when needed. Ability tests, diagnostic tests, and progress-monitoring data are helpful for parents, both to understand when special services are warranted and to monitor their child's progress.

- » **Teachers:** Teachers need access to classroom and interim/benchmark assessments to check for student understanding, differentiate appropriately, diagnose learning challenges, monitor progress, and facilitate learning. In addition, they need ongoing access to training, coaching, and resources to support real-time teaching and learning strategies.

» **District and school leaders:** District and school leaders rely on standardized interim/benchmark and summative measures to direct resources toward specific subgroups of students or students who may be falling behind. They need access to multiple sources of data to evaluate curriculum and instruction, facilitate teachers' and students' growth, monitor performance, identify achievement gaps, and inform school and district improvement plans.

» **Staff from state departments of education:** Staff from state departments of education must be able to collect, analyze, and report assessment data quickly and transparently to multiple stakeholder groups in different ways and for different purposes. They also need access to integrated data systems so that assessment data can be used in combination with other data to direct resources and support, evaluate policies and programs, and drive continuous improvement.

» **Policymakers:** Policymakers and their research staff need access to longitudinal data to track district and school health. A centralized data repository that includes common elements to link individual student assessment data to student information systems and cross-agency data can be helpful for conducting research to address pressing policy questions, such as the cost/benefit of new initiatives and the effectiveness of school programs and pilots.

Essential Characteristics of a High-Quality Assessment System

High-quality assessment systems are coherent, comprehensive, balanced, efficient, aligned, and flexible. These characteristics are not mutually exclusive; they are inextricably linked and provide a holistic approach to assessment.

Coherence

Coherence occurs when a system's component parts work together in logical and consistent ways to produce a clear and efficient whole. To achieve coherence, the "big three" categories — curriculum, instruction, and assessment — must be built upon a robust theory of student learning. Coherence is achieved when decisions about curriculum, instruction, and assessment occur in conjunction with each other. In other words, decisions about curriculum and instruction should be informed by and align with the outcomes that students should produce and that assessments will measure.

Theories explaining how students learn are used to develop a well-articulated learning model and learning progressions that describe the way in which knowledge and skills build on one another as students are exposed to new instruction and experiences (Chattergoon & Marion, 2016; Conley, 2018). Learning progressions describe the typical pathways and milestones that students follow toward mastery of a concept, skill, or practice, and they can be used to ensure strong alignment across curriculum and support materials, instructional strategies, and assessment. Ideally, learning progressions are built upon the state learning standards and support the production of classroom-level instructional systems in which curriculum, instruction, and assessment are fully integrated, as opposed to separated and disconnected.

High-quality assessment systems measure clear and concrete learning progressions and standards, in which knowledge, skills, and expectations become increasingly more advanced within and across grade levels. Assessments must move beyond measurement of low-level content knowledge, such as recollection and reproduction of facts, to measure deeper learning outcomes, such as critical thinking, creativity, problem solving, collaboration, social-emotional intelligence, resilience, resourcefulness, and skills that foster a growth mindset (Dweck, 2016; Darling-Hammond, Wilhoit, & Pittenger, 2014). Assessments must provide frequent and meaningful feedback that enables teachers to provide personalized instruction for students and enables students to take greater control over their learning process. They shift their expectations as they accept greater responsibility for their own learning, develop at their own pace, and demonstrate their knowledge across a range of cognitively complex tasks and through a variety of interests.

A theory of action is a prerequisite for a well-designed and high-quality 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 a coherent assessment system by specifying the system's purpose, the resources, and the inputs needed to produce intended outputs and outcomes.

Comprehensiveness

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 (students, families, teachers, local and state leaders, and policymakers). 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 and Mancuso, 2017). A comprehensive system addresses the assessments that are needed to fulfill a multitude of purposes for a multitude of groups.

Comprehensive assessment systems should incorporate the full range of measurement approaches to produce the evidence needed to support education decision-making (National Research Council, 2001). Teachers rely on formative assessment processes and interim/benchmark assessments to inform instructional practice and improve student understanding, whereas district staff and staff from state departments of education need information that can be used for accountability purposes to identify high-performing schools and schools that need additional supports to better serve the needs of their students.

Sigman and Mancuso (2017) provide a helpful summary of the types of assessment needed in a comprehensive assessment system to support education decision-making for various stakeholders. Table 1 describes the four broad categories of assessment: formative, diagnostic, interim/benchmark, and summative.

Table 1. Categories of Assessment

Type of assessment	Description of assessment type
Formative assessments	Formative assessments are planned and ongoing. The process is used by all students and teachers during learning and teaching to elicit and use evidence of student learning to improve student understanding of intended disciplinary learning outcomes as well as to support students to become self-directed learners (Council of Chief State School Officers, 2019). The information that is collected is finely grained, providing a level of detail about the current status of student learning in relation to lesson goals. Its main purpose is to inform real-time teaching and learning, using evidence and feedback to move learning forward by adjusting learning strategies, goals, and/or next instructional steps.
Diagnostic assessments	Although many assessments may be considered diagnostic, traditionally and formally, diagnostic tests are generally used when students are demonstrating difficulties in learning, and results may assist in diagnosing strengths and needs. Because of the diagnostic nature of these assessments, they are often administered by specially trained education personnel.
Interim/benchmark assessments	Interim, or benchmark, assessments are generally administered by teachers at key points in time for one or both of two purposes: (1) to evaluate what students have learned in relation to mid-term goals and (2) to predict students' performance on specific standards assessed by the state's end-of-year summative assessment. Interim assessments may be administered under standardized or nonstandardized conditions, depending on purpose. Results may provide teachers with an early warning signal about those students who are falling behind in their learning and may benefit from targeted assistance to help them learn content prior to end-of-year testing. For leaders, results indicate whether students are on track in meeting learning goals and can inform decisions about curricular adjustments and professional learning needs.
Summative assessments	Summative assessments provide information about students' achievement of academic content standards following a longer period of instruction, such as a full semester or school year. Examples of summative assessment include final course exams developed by a teacher and an end-of-year or end-of-course assessment developed by a state or by a multistate consortium. State-developed summative assessments are administered in a standardized manner so that each student across the state can demonstrate his or her achievement under the same testing conditions. Results from summative measures can be used for grading and reporting purposes, policy and program decisions, and decisions about resource allocation and professional learning priorities.

Although tests of content knowledge and cognitive strategies reveal a great deal about college and career readiness, other data sources are needed to develop a holistic understanding of what students know and can do and thereby support their development. To improve performance, both students and teachers need to know what students know, why they perform as they do, and how who they are influences their desire to learn and their interests in certain areas (Conley, 2018). Thus, comprehensive systems must extend beyond scale scores and achievement levels (e.g., Levels 1–5) to gather and report information about who students are and how they learn. This entails the use of noncognitive and behavioral measures, as well as the use of multiple collection methods and sources, including surveys, observations, logs or self-reports, and self-evaluations. Sources of information include students, teachers, school staff, and parents/caregivers. Used in combination, noncognitive data

collected from a wide range of methods and sources can yield helpful insights into students' interests, learning strategies, motivation, and other factors that play a big role in the learning process (Gaertner & Roberts, 2017; Gaertner, Conley, & Stoltz, 2016; Gaertner & McClarty, 2015, 2016).

Balance

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

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 each assessment should be used, the purpose of each assessment, who needs to use results, and how their 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 example, end-of-year (summative) assessments provide useful information for understanding student achievement, including subgroup performance and gaps, assessing the quality of curriculum, and evaluating instructional programs. However, results from end-of-year assessments provide limited data for informing a teacher's instruction because data from summative assessments are less granular (i.e., results are reported at the topic or domain level, not by individual learning standard). Further, results are often reported after the school year ends, which only helps planning for the following school year and cohort of students. Similarly, a principal who reviews benchmark results quarterly with teachers 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 on a weekly basis to group students for a lesson on a specific topic will quickly become frustrated.

In some assessment systems, the emphasis on statewide standardized testing is so heavy that stakeholders try to use the information to address purposes for which these assessments are not intended. Schools that spend substantial time preparing for, administering, and taking these assessments may be tempted to use the results to inform too many decisions. Substantial time and money are devoted to developing and administering standardized tests, so it may be difficult to accept that results will not provide the depth of information schools and teachers need to diagnose learning deficiencies, understand students' misconceptions, or even evaluate performance for important content domains. However, expecting more from these tests (e.g., content or cognitive domain scores) would require even longer tests that produce redundant information, much of it more efficiently collected using interim/benchmark assessments.

The recent focus on developing more personalized approaches to education fundamentally changes how assessment should happen. Personalization enables students to demonstrate what they know and can do at their own pace. It also pushes students to demonstrate their development of higher-order skills such as problem solving, communication, and evaluation. An assessment system that supports personalized learning may benefit from dynamic assessments that can be administered on demand, at greater frequency, and with better information to shape and inform the learning process for individual students. Performance-based assessment, portfolio assessment, and assessments that extend beyond simple multiple-choice items are helpful tools for measuring individual growth in complex domains. In addition, as schools and teachers shift their instruction to focus more on an individual's learning progress, personalized instruction demands a shift in emphasis from "assessment *of* learning" to "assessment *for* learning" (i.e., more emphasis on formative assessment and less emphasis on summative and interim/benchmark assessments) (Heritage, 2017; Jobs for the Future, 2018, Stiggins, 2017). In this way, shifting time and energy toward formative assessment will better inform learning and may reduce teachers' concerns about overtesting students. Figure 1, adapted from Darling-Hammond, Wilhoit, and Pittenger (2014), is a helpful visual for understanding the shift in emphasis from summative to formative assessment.

Figure 1: Relative Emphasis of Assessment Type in a Balanced Assessment System
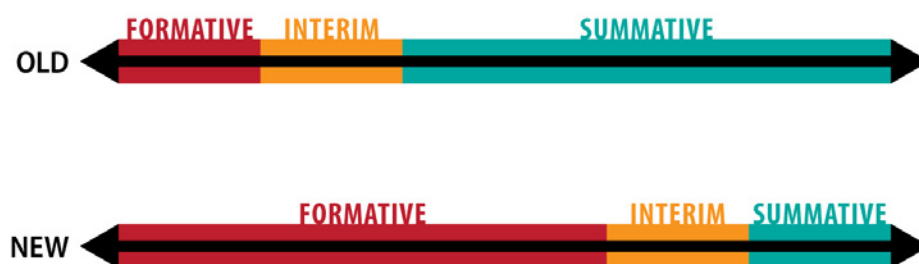
As states and districts consider how to address the balance of assessments in their existing systems, they will need to consider the purposes of each assessment, how each assessment will be used, for whom each assessment will provide the most benefit, the costs to implement the assessment, and the potential benefits to individual stakeholders (Council of Chief State School Officers, 2015). In addition, as new assessment tools are introduced, states will need to review their purpose and use in relation to existing assessments. These reviews are essential for maintaining an efficient system (more on efficiency below); however, they often lead to difficult decisions, particularly when a new assessment tool or strategy is designed to replace one that is well liked but outdated or misaligned.

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). For example, state assessment results used by staff from state departments of education to monitor achievement gaps and identify the lowest-performing schools should also be useful for districts and schools to monitor student performance and growth, evaluate school programs, and inform curricular and instructional decisions.

Assessment practice in the No Child Left Behind (NCLB) era presented an unfortunate example of inefficiencies that can emerge when one type of assessment carries more weight than it should. In 2001, federal policy began to measure a school's success on annual standardized assessment results, relying on results from mathematics and English language arts assessments. The high stakes associated with these tests — under the NCLB Act, 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 the test" as opposed to teaching for learning. Interim and benchmark assessments designed 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 the newer college- and career-readiness standards, which minimized the assessments' usefulness for guiding instruction. The result was that during the NCLB era, formative and performance-based assessment strategies useful for guiding day-to-day curriculum and instructional decisions and for assessing higher-order skills took a back seat to test prep. In 2015, the Council of Great City Schools (the Council) examined testing in the nation's largest urban school districts (Hart, Casserly, Uzzell, Palacios, Corcoran, & Spurgeon, 2015). Among their findings, the Council reported the following:

» About 39% of districts reported having to wait between two and four months before final state test results were available at the school level, thereby minimizing their utility for instructional purposes. In addition, results did not come back until after the conclusion of the school year.

» Multiple redundant exams were sometimes given in the same subjects and grades to the same students because not all results yielded data by item, grade, subject, student, or school — thereby prompting districts to give another exam to get data at the desired level of granularity.

» Districts used standardized assessments for purposes other than those for which they were designed. Some of these applications were state-recommended or state-required policies, and some originated locally.

» District tests were sometimes not well aligned with college- and career-readiness standards and often did not assess student mastery of specific content.

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. As a result, redundant, untimely, and unused assessments can be identified and eliminated or adjusted to sustain balance and alignment across the system.

Alignment

An aligned assessment system is one that supports the use of assessment within the classroom, across levels of the system, and across the grade-level continuum so that what is taught and measured leads to college- and career-ready citizens (Sigman & Manusco, 2017). Alignment requires that stakeholders at all levels of the system have access to a set of assessment tools (e.g., tests, rubrics, self-assessment tools), training, and technology to support student learning. For example, although a teacher may rely heavily on formative assessment practices to determine what individual students know and can do, principals and district staff rely more heavily on results from standardized tests to examine subgroup performance or flag students who may need more individualized support. Both types of assessment information are critical, and without both, the system doesn't 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 (e.g., college- and career-readiness standards) and coordinated communication within and across agencies (state departments of education, regional support centers, districts, and schools). To ensure efficient dissemination of these tools and strategies, departments within the state education agencies and local education agencies (LEAs) must align their objectives and collectively determine the tools and training that are necessary to support school and classroom needs (Childress, Elmore, & Grossman, 2006).

Alignment also refers to the connection between curriculum, instruction, and assessment, where each component aligns to the others as well as to the state's learning standards. An aligned assessment system reinforces the connection between curriculum, instruction, and assessment, thereby enhancing the learning process and improving outcomes for all students.

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, and address inequities. However, too much standardization has the opposite effect: increasing inefficiencies (e.g., overtesting by using redundant assessments), inhibiting fairness, promoting inequities, and potentially stifling innovation.

The 2016 Every Student Succeeds Act (ESSA), which replaced the NCLB Act, offers states more flexibility and addresses concerns about increases in standardized testing, narrowing of test methods, and narrowing of skills and abilities taught and tested under the NCLB Act (Conley & Darling-Hammond, 2013). ESSA includes the new Innovative Assessment Demonstration Authority (IADA), which allows states to operate innovative assessment systems, providing the flexibility to evaluate new assessment practices. Examples of such innovative practices include replacing one summative assessment with multiple interim assessments that result in a single summative score, also known as a through-grade or through-course assessment model, and utilizing performance assessment (e.g., portfolios, projects, extended performance tasks) to "partially" measure higher-order thinking skills.

These flexibilities are designed to influence innovation and promote improved balance in state and local assessment systems. In theory, providing states with more flexibility would allow them to introduce assessment methods that can measure complex constructs embedded in states' 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 through discussion, oral presentation, and/or 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 to support more efficient assessment practice. Striking the right balance between flexibility, structure, and consistency can be difficult. But states and districts willing to pilot and adopt alternative approaches to assessment will likely develop the capacity to address challenges and sustain high-quality assessment systems.

Principles to Inform the Selection of Assessments and Assessment Strategies

A high-quality assessment system should adhere to a set of principles that inform the selection of individual assessments to address specific needs and purposes. In addition, decisions about whether to include individual assessments must be considered within the context of the system as a whole. For any assessment system to work properly, each individual assessment within the system must meet specific criteria for serving its intended purpose(s), and the quality of each individual assessment must be examined to ensure it produces valid, reliable, meaningful, and necessary information. For example, these questions should be answered (Council of Chief State School Officers, 2015): 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?

In 2015, the U.S. Department of Education released the Testing Action Plan fact sheet. The plan was released, in part, to address a crisis of overtesting students that was emerging across the country (U.S. Department of Education, 2015; Hart, Casserly, Uzzell, Palacios, Corcoran, & Spurgeon, 2015). The Testing Action Plan included a set of seven principles, described in Table 2, to support state and district leaders in ensuring that their students take high-quality and thoughtfully selected assessments as part of a balanced and comprehensive system (Sigman & Mancuso, 2017).

Table 2. Principles for Assessment From the U.S. Department of Education Testing Action Plan

Principle	Description
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.
High quality	Assessments should measure knowledge and skills against state-developed college- and career-readiness 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.
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 2% of classroom time taking standardized tests. Low-quality test-preparation strategies must be eliminated.
Fair	Assessments should include accessibility and accommodations for students with disabilities and English learners to accurately reflect what students really know and can do. The same assessments of core subjects (reading, writing, science, mathematics) 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 schools need additional support and interventions to succeed.
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.
Just one of multiple measures	No single assessment should ever be the sole factor in making educational decisions about a student, an educator, or a 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 be used to ensure a comprehensive evaluation of performance.
Tied to improved learning	Most 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.

These principles can be used to evaluate, compare, keep, and/or discard existing assessments, thereby creating a more coherent, comprehensive, balanced, efficient, and aligned assessment system.

Approach

The researchers utilized two key methods to evaluate the existing statewide assessment system in North Carolina: interviews with numerous stakeholders and an extensive review and analysis of publicly available documentation related to the current assessment and accountability systems.

Research Questions

The goal of the study was to collect information regarding the usefulness of assessments in North Carolina and to elicit ideas for how the state may better support the implementation of a high-quality assessment system. The evaluation focused on the following research questions:

» How are statewide assessment results used by districts and schools to inform student and school improvement and close educational opportunity and achievement gaps?

» What technical assistance, training, resources, and support does the North Carolina Department of Public Instruction (NCDPI) provide to help districts and schools interpret and use statewide assessment results for student and school improvement?

» What gaps, if any, exist in the statewide assessment system? How do districts address or compensate for the gaps in the statewide assessment system?

Note that these research questions also supported the findings in the separate companion report describing the accountability system, *North Carolina's Statewide Accountability System: How to Effectively Measure Progress Toward Meeting the* Leandro *Tenets* (Learning Policy Institute, 2019).

Data Sources

The researchers conducted interviews with several state-level stakeholders, including staff from the NCDPI, and with county-level administrators who oversee assessment and accountability for their districts. A total of 12 interviews were conducted, 6 with state-level stakeholders and 7 with administrators from county offices of education representing a range of district characteristics. The interview protocols are provided in Appendix A.

In addition, the researchers reviewed key court documents and numerous documents from the NCDPI website, including, but not limited to the following:

» North Carolina ESSA Consolidated State Plan approved by the North Carolina State Board of Education and submitted to the U.S. Department of Education in September 2017

» *Testing Transparency — Statewide Administration of the Testing Program*, a report to the North Carolina General Assembly in December 2018

» *Interactive Local Testing Report*

» *North Carolina Testing Program Test Development Process: Item, Selection, and Form Development* (updated July 2017)

» Assessment/test specifications for the various state-required assessments

» Technical reports for the various state-required assessments

» Assessment briefs for the various state-required assessments

» Assessment briefs for the NC Check-Ins

» State testing results (Green Book) for 2017–18

Also, the researchers surveyed principals from across North Carolina. Two questions in the survey focused on the use of assessments in North Carolina. Appendix B provides the two assessment-related questions that were included on the principal survey.

Findings

North Carolina has a rich history of testing and accountability, beginning with the implementation of the statewide testing program in 1992–93 and the ABCs of Public Education in 1996–97. The statewide tests and the ABCs of Public Education established high education standards and accountability for all schools in North Carolina. The ABCs of Public Education model was replaced in 2012–13 by the (current) READY accountability model, which transitioned the alignment of the statewide assessments to college- and career-readiness standards. According to the NCDPI's READY Accountability Briefs (2013, 2016, 2017a), the READY initiative established the following:

» The North Carolina Standard Course of Study (NCSCOS) for all subjects and grade levels, which focuses on the critical, most essential skills and knowledge that students need at the next grade level and after high school

» End-of-grade (EOG) and end-of-course (EOC) assessments aligned to the NCSCOS, including more rigorous open-ended questions and real-world applications that require students to express their ideas clearly with supporting facts

» An accountability model that measures how well schools are doing to ensure that students are career and college ready upon high school graduation

In conjunction with the READY initiative, the North Carolina General Assembly passed legislation requiring A–F school performance grades, which are based on each school's achievement and growth scores on the EOG and EOC assessments. The effort to implement more rigorous standards and measures was intended to help students be ready for anything they choose to pursue after high school graduation (North Carolina Department of Public Instruction, 2013).

The researchers collaborated with the Learning Policy Institute to evaluate North Carolina's assessment and accountability systems. Although statewide assessment systems are inextricably linked to states' accountability systems, the remainder of this section details the researchers' findings related to the existing statewide assessment system.[1]

---

1   The findings and conclusions related to the accountability system are included in a separate report, *North Carolina's Statewide Accountability System: How to Effectively Measure Progress Toward Meeting the* Leandro *Tenets* (Learning Policy Institute, 2019).

Coherence and Alignment

The NCDPI states on its website that classroom instruction is a partnership between the state and local educators. The state sets the standards through the NCSCOS, and LEAs determine which curriculum and instructional materials to use. This approach is typical, as most states do not develop or require specific curricula for statewide use. However, there are no detailed resources that describe vetted or endorsed curricula that align to the NCSCOS. Districts and schools choose their curricula. Although the researchers didn't fully evaluate the breadth and depth of the curricula used across the state, central office staff that were interviewed suggested that the support and resources available to educators are variable across districts (e.g., larger LEAs with more resources tend to develop their own curriculum guides, resources, and training to support instruction, whereas smaller LEAs must rely on limited instructional resources).

Although no curricular materials are vetted or endorsed by the state, the NCDPI provides instructional support materials through its website. Unpacking-standards and cross-walk documents are provided for each content area. The unpacking-standards documents provide a more detailed explanation of the standards, designed to help educators better understand the student performance expectations articulated by the standards. The cross-walk documents describe the relationship between current and past standards. Some content areas include graphic organizers and glossaries of key terms. Vertical progressions, which describe the relationship of standards across the grade levels, are provided for English language arts and mathematics. Instructional frameworks that outline and sequence clusters of related standards, including suggested durations of time for instruction, are also provided for mathematics. To support the administration of the NC Check-Ins (the optional state-provided interim/benchmark assessments; see Comprehensiveness, pp. 16–18), the NCDPI provides test specifications that describe which standards are assessed on each NC Check-In.

Improving education outcomes for all students requires an extensive effort to strengthen the coherence and alignment between curriculum, instruction, and assessment. North Carolina's theory of action, as stated in its ESSA plan, is focused on creating an adaptive and personalized learning environment for every student. More specifically, North Carolina's guiding principle is to continue to transform from industrial-age practices of providing all students and educators with the same inputs and opportunities to digital-age practices in which all students and educators have access to unique learning experiences based upon their individual needs and aspirations (North Carolina Department of Public Instruction, 2017a).

Although North Carolina's theory of action is commendable, there is little evidence within the remainder of the ESSA plan, or elsewhere, that indicates that the statewide assessment system is aligned to the theory of action. The ESSA plan references frequent, on-demand, competency-based assessment, requiring the capacity to administer assessments at different times and for different students throughout the year and to elicit evidence of proficiency at a granular level (e.g., the standard level). Through its application to the IADA (see Flexibility, pp. 19–20), the NCDPI further suggests interest in moving toward a personalized and competency-based assessment system. However, the current assessment system uses a summative (end-of-year) assessment model, and the NCDPI does not provide a clearly defined plan to support the scalability of a personalized learning environment for all students in North Carolina. As the state transitions toward increased personalization, ensuring coherence and alignment of curriculum, instruction, and assessment will be critical to the success of its vision.

Comprehensiveness

ESSA requires that states annually assess students in English language arts/reading and mathematics in grades 3–8 and once in high school. It also requires one science assessment per grade span (i.e., once in grades 3–5, once in grades 6–8, and once in high school). In addition, English language proficiency assessments are required for all English learners in K–12. ESSA also stipulates that all students should be tested, that appropriate accommodations should be provided to students when needed, and that the use of alternate assessments aligned to alternate academic achievement standards should be limited to 1% of all tested students in the state.

The existing statewide assessment system in North Carolina complies with federal requirements set forth by ESSA. Table 3 summarizes the state-required tests in North Carolina for the 2018–19 school year.[2] State-required assessments are developed by or acquired by the NCDPI, and they are administered to all students in North Carolina.

Table 3. North Carolina Required Testing for the 2018–19 School Year

Grade	English language arts/reading	Mathematics	Science	Other	English language learners (ELLs)
3	Beginning-of-Grade 3; EOG	EOG			ACCESS for ELLs®
4	EOG	EOG			ACCESS for ELLs®
5	EOG	EOG	EOG		ACCESS for ELLs®
6	EOG	EOG			ACCESS for ELLs®
7	EOG	EOG			ACCESS for ELLs®
8	EOG	EOG; NC Math 1	EOG		ACCESS for ELLs®
9		NC Math 1			ACCESS for ELLs®
10	English II		Biology	Pre ACT®	ACCESS for ELLs®
11		NC Math 3		ACT®	ACCESS for ELLs®
12				ACT Workkeys®	ACCESS for ELLs®

The Kindergarten W-APT and WIDA Screener (grades 1–12) are required assessments for the identification of English learners; they are only administered once to determine eligibility.

In addition to the assessments listed in Table 3, there are several other state-provided assessments. State-provided assessments are either developed by the NCDPI or acquired by the NCDPI, but some are not required to be administered to all children in North Carolina. Table 4 summarizes the additional assessments provided by the NCDPI.

---

2 Table 3 is adapted from http://www.ncpublicschools.org/docs/accountability/1819reqtst.pdf

Table 4. Additional State-Provided Assessments

Name	Grades	Type
NCEXTEND1	3–8; 10–11	Alternate assessments (for EOG and EOC) required for students with significant cognitive disabilities
Alternate ACCESS for ELLs®	3–12	Alternate assessment (for ACCESS for ELLs®) required for students with significant cognitive disabilities and who are English learners
College and Career Readiness Alternate Assessment (CCRAA)	10–11	Alternate assessment (for PreACT® and ACT®) required for students with significant cognitive disabilities
NC Final Exams	4–12	Required EOG or EOC assessments for grades and subjects for which there is no federally mandated assessment
NC Check-Ins	3–8	Optional interim assessments aligned to EOG assessments
mCLASS®: Reading 3D™	K–3	Required benchmark assessments administered three times per year
Formative Instructional and Assessment Tasks for Mathematics	K–2	Optional assessment tasks
NC KEA Formative Assessment Process	K	Kindergarten entry assessment

The *NCEXTEND1*, Alternate ACCESS for ELLs®, and CCRAAs are the required alternate assessments aligned to alternate academic achievement standards for students with significant cognitive disabilities.

The NC Final Exams are required EOG or EOC assessments for grades and subjects for which there is no federally mandated assessment (e.g., social studies, high school English and mathematics courses other than English II, NC Math 1, and NC Math 3). Results from the NC Final Exams contribute to measures of student growth for teachers, but they are not used in the READY accountability system. For high school courses, results contribute to at least 20% of students' final course grades. LEA central office staff and educators that we interviewed suggested that the NC Final Exams are of limited utility, citing lack of actionable results (e.g., scale score, achievement levels) to support improvements in teaching and learning.

The NC Check-Ins are optional English language arts/reading and mathematics interim/benchmark assessments. They are designed to align to the EOG and EOC assessments, sharing a common item bank. The NC Check-Ins can be administered three times per year, and the results provide immediate and actionable information, enabling teachers, students, and families to better understand progress on the grade-level standards and end-of-year goals. There is strong evidence that LEAs find the NC Check-Ins very useful. In interviews conducted by the researchers, central office staff from all seven districts praised the NC Check-Ins, and they indicated that they had phased out or were in the process of phasing out their existing interim/benchmark assessments, replacing them with the NC Check-Ins.

There are several assessments that are used in grades K–2 for which there are no federally mandated assessments other than English language proficiency assessments for English learners. The mCLASS®: Reading 3D™

benchmark assessments are required three times per year in K–3. The combination of mCLASS assessments with the required Beginning-of-Grade 3 and EOG assessments leads to an abundance of testing in grade 3, ostensibly to support the state's Read to Achieve legislative initiative. The kindergarten entry assessment, administered in the first 60 days of kindergarten, is designed to support the formative assessment process and used to inform instructional planning. Aggregate data are collected by the NCDPI at the end of the assessment window, but the data are not used for accountability purposes.

The assessment directors from the seven county offices of education that were interviewed suggested that the state summative assessment data lack the detail and specificity (i.e., information on how well students performed in relation to specific academic standards) that would enable educators to modify instruction for individual students. However, state summative assessments are not designed to provide detailed data at the standards level because they are administered at the end of the year and are designed to assess the breadth of the standards. Further, the state summative assessments include a limited number of items (due to constraints on administration time), which minimizes the amount of detail that can be reported reliably. Reporting state summative assessment results at the standards level would require longer assessments (i.e., more assessment items), which would increase testing time and be less efficient to administer. Summative assessments are intended to be high-level snapshots of student progress and proficiency, whereas interim/benchmark assessments provide more detailed information on students' progress toward specific learning goals and standards.

The principal survey results suggested a different take on the usefulness of state summative assessment data, indicating that the data are useful for the following:

» School improvement (83% somewhat or strongly agreed, *n* = 585)

» Professional development planning (77% somewhat or strongly agreed, *n* = 587)

» Understanding student strengths (78% somewhat or strongly agreed, *n* = 585)

» Understanding student weaknesses (76% somewhat or strongly agreed, *n* = 586)

Principals also reported that state assessment data were easy to access (75% somewhat or strongly agreed, *n* = 586) and easy to understand (74% somewhat or strongly agreed, *n* = 583). Principals reported the strongest disagreement with the timeliness of data from the state assessments — 54% of principals (*n* = 583) somewhat or strongly agreed that data are available in a timely manner. The full results from the assessment-related questions on the principal survey are provided in Appendix B.

In summary, the statewide assessments provided by the NCDPI are comprehensive and comply with federal requirements. The NCDPI provides summative and interim/benchmark assessments that can be used to measure student progress and proficiency and inform decisions about policies and programs. LEAs and educators provide additional interim/benchmark assessments, diagnostic assessments (when needed), and formative assessment. This approach to establishing a comprehensive system is typical; however, it is important to monitor the use of assessments across the state to ensure balance and efficiency between state-required assessments and the additional assessments chosen and administered by LEAs and educators.

Balance and Efficiency

Many LEAs in North Carolina require the administration of assessments that are in addition to the state-required and state-provided assessments. The availability of commercially developed assessments is expansive. Commercial assessments, which claim to be aligned to the state standards, are typically marketed to districts for use as interim/benchmark, classroom, or formative assessments. Therefore, it is not surprising that many LEAs in North Carolina supplement the state-provided assessments with commercial or locally developed assessments.

A concerted effort is being made by state leaders and policymakers to better understand the use of assessments at the local level. In December 2018, the North Carolina State Board of Education and the NCDPI provided their first report to the state's General Assembly, per G.S. §115C-174.12 (d), describing the status of locally required assessments in North Carolina districts. This report, in conjunction with the *Interactive Local Testing Report*, is intended to provide greater transparency on the use of locally required assessments across the state. Subsequent sections of the same general statute include similar requirements that the North Carolina State Board of Education and the NCDPI produce a list and calendar for all state-required assessments by September 1 of each year (G.S. §115C-174.12 (e)).

Flexibility

The NCDPI, the North Carolina State Board of Education, and the state's General Assembly are involved in ongoing efforts to improve the assessment system in North Carolina. Beginning in 2014 with the appointment of the Task Force on Summative Assessment and an analysis of and report on its required assessments by external researchers, policymakers became keenly aware of the importance of balanced and efficient use of assessments in North Carolina (Guindon, Huffman, Socol, & Takahashi-Rial, 2014). The Task Force on Summative Assessment introduced the through-grade assessment model, which resulted in a proof-of-concept study and the development of the NC Check-Ins (North Carolina Department of Public Instruction, 2017).

The Task Force on Summative Assessment, the proof-of-concept study, and the development of the NC Check-Ins paved the way for the state's Innovative Assessment Demonstration Authority application. On December 14, 2018, the NCDPI applied to the U.S. Department of Education's IADA to explore an alternative assessment model, called a through-grade assessment model, as a possible replacement for its EOG assessments.[3] Through-grade assessment models utilize multiple interim assessments throughout the school year in lieu of a single summative assessment at the end of the year. In its application, the NCDPI proposed the development of the North Carolina Personalized Assessment Tool which would consist of three or four assessments that would be administered throughout the school year, replacing EOG assessment for each grade at the end of the school year. The overarching goal of this model is to provide teachers, students, and families with immediate and actionable data for guiding instruction during the school year. At the time this report was written, North Carolina's IADA application was pending approval from the U.S. Department of Education.

---

3 North Carolina's application and subsequent documentation can be found at https://www2.ed.gov/admins/lead/account/iada/index.html

North Carolina is one of four states to apply to the IADA for an innovative assessment pilot. In 2018, Louisiana and New Hampshire were approved for innovative assessment pilots under the IADA. Louisiana is attempting to create new humanities assessments with the goal of replacing its current English language arts and social studies assessments with the new humanities assessments. Louisiana is also attempting to include the use of passages from texts that have been covered throughout the school year in these new models so that the assessments more closely align to curriculum and instruction. New Hampshire is experimenting with administering performance-based assessments. Like North Carolina, Georgia applied in December 2018 for innovative assessment pilots, also proposing the use of a through-grade model, and its application has yet to receive approval. At the time this report was written, no other states have implemented a through-grade assessment model at the state level.

Quality

Given that the goal of the North Carolina Standard Course of Study is to prepare all students to become college and career ready, it is essential that the state-required and state-provided assessments measure the depth and breadth of the knowledge and skills defined in the NCSCOS (North Carolina Department of Public Instruction, 2018). College- and career-readiness standards and expectations, like those defined in the NCSCOS, require students to demonstrate complex reasoning and problem-solving skills and to communicate effectively. To adequately assess the knowledge and skills defined in the NCSCOS, it is important for assessments to include opportunities for students to demonstrate their abilities to reason, solve complex problems, and communicate effectively. Completing a thorough review of every state-required and state-provided assessment was beyond the scope of this evaluation; however, the assessment specifications documents and the alignment study report from the NCDPI website provide adequate information to assess the quality of the state assessments.

The NCDPI commissioned the Wisconsin Center for Education Research to conduct an independent evaluation of the alignment of the state assessments to the NCSCOS. Independent alignment studies are required under the federal peer review guidelines (U.S. Department of Education, 2018). The conclusion of the alignment report describes that the assessments for the most part are very well aligned, with marginally low alignment measures noted for three grade-level assessments (Smithson, 2015).

The assessment specifications reveal that the state summative assessments rely heavily on multiple-choice items. Heavy reliance on multiple-choice items lessens the cognitive demand of the assessment and deemphasizes complex reasoning and communication skills, which are key attributes of college- and career-readiness standards. Further, assessments that heavily rely on multiple-choice items tend to influence teachers' instructional decisions, often resulting in a focus on lower-level cognitive skills. The English language arts/reading assessments for grades 3–8 are entirely composed of multiple-choice items and do not assess writing standards. The English II (high school) assessment includes three constructed-response items, requiring the students to respond in a paragraph or less; the rest of the assessment is composed of multiple-choice items. The mathematics assessments for grades 3–8 include multiple-choice and gridded-response items. The high school mathematics assessments consist of multiple-choice, numeric-entry, and technology-enhanced items. The science assessments consist of multiple-choice and technology-enhanced items. The benefit of utilizing multiple-choice, gridded-response, numeric-entry, and technology-enhanced items is automated scoring, which minimizes the cost of development,

administration, and scoring and typically allows for reporting results more quickly. However, heavy reliance on these item types lessens the cognitive demand of the assessment and deemphasizes complex reasoning and communication skills, which are key attributes of college- and career-readiness standards. Further, assessments that heavily rely on multiple choice tend to influence teachers' instructional decisions, often resulting in a focus on lower-level cognitive skills.

Assessment results should be easy to understand. Therefore, students' scores on assessments are typically reported in conjunction with achievement levels. Achievement levels describe varying degrees of knowledge and skills demonstrated by students as determined by their scores on assessments. Achievement levels are critical to the interpretation of test results, providing stakeholders with clear expectations for what it means to meet or exceed a predetermined standard (i.e., proficiency). Further, well-defined achievement levels are essential to establishing meaningful and defendable cut scores, which identify the minimum score students need to obtain for specific levels of performance.

North Carolina utilizes five achievement levels (Levels 1–5) when reporting results for all state-required assessments that are utilized for accountability. Table 5 summarizes the achievement levels for the state-required assessments in North Carolina, as described by the Understanding the Five Achievement Levels document produced by the North Carolina State Board of Education and NCDPI in April 2014.

Table 5. Achievement Levels for State-Required Assessments

Achievement level	Meets on-grade-level proficiency standard	Meets college- and career-readiness standard
Level 5 – Superior Command of knowledge and skills	Yes	Yes
Level 4 – Solid Command of knowledge and skills	Yes	Yes
Level 3 – Sufficient Command of knowledge and skills	Yes	No
Level 2 – Partial Command of knowledge and skills	No	No
Level 1 – Limited Command of knowledge and skills	No	No

North Carolina's achievement levels ostensibly describe two levels of proficiency: on-grade-level proficiency (Level 3) and college and career readiness (Level 4). It is atypical for a state assessment program to utilize two standards of proficiency. Accountability results (i.e., school performance grades) are predicated upon students' achievement of Level 3 (on-grade-level achievement). The detailed description of Level 3 in the Achievement Level Information documents articulates the need for additional academic support to reach college and career readiness:

> *Students performing at this level have a sufficient command of grade-level knowledge and skills … but they may need academic support to engage successfully in this content area in the next grade level. They are prepared for the next grade level but are not yet on track for college and career readiness without additional academic support. (North Carolina State Board of Education, North Carolina Department of Public Instruction, 2014, p. 2)*

It is unclear why in North Carolina, grade-level and college- and career-readiness expectations are not synonymous and are not included together in one common proficiency level, rather than separated into two proficiency levels. All other documentation, including the READY Accountability Briefs, indicate that the goal of the NCSCOS is to prepare all students to become college and career ready. Undoubtedly, college and career readiness should be the standard all students in North Carolina should strive to achieve, and proficiency on the statewide assessments should reflect as much.

North Carolina's initial intention was to utilize four achievement levels: Limited Command, Partial Command, Solid Command, and Superior Command of the knowledge and skills in the North Carolina standards. The minimal proficiency standard would be the lowest score required to attain Solid Command. Standard setting — the process for setting cut scores — took place in July 2013 and was based on these four achievement levels. Standard setting is the industry-accepted practice for establishing cut scores that delineate the various achievement levels. This practice involves the convening of numerous educators from across the state to review the assessments and items and provide professional judgments to establish recommended cut scores. The standard-setting process used in North Carolina is well documented in technical reports on the NCDPI website.

After North Carolina completed its standard setting, a fifth achievement level — Level 3, Sufficient Command (representing on-grade-level proficiency) — was created and approved by the North Carolina State Board of Education. The methodology for establishing this achievement level involved subtracting one conditional standard error of measurement from the cut score delineating Solid Command (i.e., proficiency). This post-standard-setting creation of an additional level is atypical. And in the case of North Carolina, it created a scenario in which the error-bands associated with scores in Level 3 extend across three achievement levels (Levels 2, 3, and 4). Such a scenario is typically avoided when setting standards (cut scores). Peer reviewers also expressed this concern in the State Assessment Peer Review Notes for North Carolina (April and June 2016).[4]

---

4 https://www2.ed.gov/admins/lead/account/nclbfinalassess/nc5.pdf

Conclusions

Best Practices

Comprehensiveness and Flexibility

Overall, North Carolina's statewide assessment system complies with federal requirements. The English language arts/reading and mathematics general assessments for grades 3–8, the science general assessments in grades 5 and 8, and the English language arts/reading, mathematics, and science general assessments in high school all meet the U.S. Department of Education's assessment peer review requirements.[5]

The NC Check-Ins, the state-provided interim assessments aligned to the end-of-year assessments (i.e., the EOG and EOC assessments) that are freely available to all LEAs across North Carolina, have been well received, and their ongoing use should help reduce redundancy, provide actionable information to educators within the school year, and streamline the number of assessments used at the local level.

More recently, public discussion has focused heavily on the amount of time spent on assessments, and the NCDPI, the North Carolina State Board of Education, and the state's General Assembly are working together to better understand the use of local assessments and their impact on the amount of time spent on testing. The *Interactive Local Testing Report*, in conjunction with the official report *Testing Transparency — Statewide Administration of the Testing Program*, pursuant to S.L. 2017-57 (SB 257), Section 7.28A.(b), should continue to raise awareness around the number of assessments being administered at the local level and provide state and local stakeholders with the information needed to ensure balanced and efficient use of assessments.

Through its application to the Innovative Assessment Demonstration Authority, North Carolina has demonstrated its commitment to implementing high-quality assessments and assessment practices, capitalizing on an opportunity to explore alternative testing models and creating an opportunity to demonstrate that the through-grade assessment model is viable at the state level.

---

5 https://www2.ed.gov/admins/lead/account/nclbfinalassess/nc6.pdf

Major Needs and Challenges

Coherence and Alignment

Improving education outcomes for all students requires an extensive and collaborative effort at all levels to strengthen the connection between curriculum, instruction, and assessment. It is unreasonable to expect assessment results to improve without significant investment in educational resources, including high-quality curricular and instructional materials. Although the NCDPI provides instructional support materials via its website, there is a critical need for providing additional and ongoing support to LEAs to ensure the implementation of high-quality curricular and instructional materials across the state.

North Carolina's theory of action outlines four pillars of personalization. These four pillars call for students to have learner profiles, to pursue individualized learning paths, to follow competency-based progressions, and to have flexible but structured learning environments. These are ambitious objectives that require a significant shift in the structure of North Carolina's education system, including a significant investment in resources to support such a shift. More clarity is needed to describe how the state's assessment system supports the personalization of learning described in the theory of action, including how North Carolina plans to scale up personalization across the state and how the assessment system ensures a sound basic education to every student in North Carolina.

As described in North Carolina's ESSA theory of action and its IADA application, the relationship between personalized learning, the current state assessment system, and the proposed assessment system is vague. The theory of action describes "using real-time assessment strategies to inform classroom instruction, as opposed to using extensive, overbearing summative assessments as the main tools to inform instruction." Although this is certainly good assessment practice, it is not necessarily personalization, especially as described by the above-mentioned four pillars. The state's proposal to break up end-of-year assessments into several interim assessments would certainly allow for more immediate use of assessment data within the school year, which could lead to increased personalization of learning. However, it is also critical that educators be properly supported to understand how to use the assessment data to better personalize instruction. Asking educators to provide personalized instruction without providing significant professional development, including high-quality curricular and instructional resources, would be futile. Well-designed instructional support materials and a robust communication and dissemination strategy will be needed to fully support North Carolina's vision for more personalization and, ultimately, improved student outcomes as measured by its assessment system.

Balance and Efficiency

Although the reports on the use of local assessments and awareness of testing time are important first steps to ensuring balance and efficiency within the assessment system, the state could provide additional support to assist LEAs with assessment audits to identify possible redundancy of assessments. This additional support should help to ensure balance and efficiency throughout the assessment system. Further, promoting and supporting the use of the NC Check-Ins as interim/benchmark assessments would also support efforts to ensure balance and efficiency by potentially reducing the number of assessments at the local level. If LEAs choose to administer the

NC Check-Ins, they could reduce or discontinue the use of many of their local assessments, which would lead to reductions across the state in time spent testing. Further, more extensive use of the NC Check-Ins might allow the NCDPI to be more efficient with disseminating resources and support materials across the state.

Quality

High-quality assessments, as described by the U.S. Department of Education's Testing Action Plan, cover the full range of relevant state standards, elicit complex student demonstrations and applications of knowledge, provide valid and reliable results for all students, and provide an accurate measure of student growth. North Carolina's existing state-required assessments lack item types that measure the complex reasoning and communication skills that are aligned to the rigorous college- and career-readiness standards described in the NCSCOS. Other state assessment systems include more constructed- and extended-response items, including performance-based items and writing tasks, than currently exist in the North Carolina state summative assessments. The inclusion of items that require students to demonstrate application of their knowledge and skills should improve teaching and learning by emphasizing the importance of complex reasoning and communication skills.

When college- and career-readiness standards were adopted by most states, the level of rigor of the standards and the expectations of students increased sharply. The assessments aligned to these newer standards became more difficult than their predecessors, which in turn depressed proficiency rates. North Carolina originally planned and set cut scores for four achievement levels, in which the lower two levels described the need for additional academic support and the higher two levels described meeting or exceeding the proficiency standard. In March 2014, the North Carolina State Board of Education added a fifth achievement level describing on-grade-level proficiency to the existing proficiency levels (college- and career-readiness proficiency). However, stakeholders should be confident that achievement-level classifications translate to students' progress toward college and career readiness rather than describing a difference between grade-level proficiency and college- and career-readiness proficiency. A more coherent definition of proficiency and revisions to the achievement levels aligning grade-level expectations and college- and career-readiness expectations is needed to provide stakeholders with a clearer picture of student progress and proficiency.

Although assessment results on the state summative assessments provide important measures of student achievement and growth, these results are even more useful for educators when used in conjunction with other indicators of student progress. As North Carolina's assessment system continues to evolve, state-level policy-makers and decision-makers should promote the use of multiple indicators when making decisions about student achievement and progress toward providing all students in North Carolina with a sound basic education (Learning Policy Institute, 2019).

References

Chattergoon, R., & Marion, S. F. (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. S. (2006). *How to manage urban school districts*. Cambridge, MA: Harvard Business Review.

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. (2015). *Comprehensive statewide assessment systems: A framework for the role of the state education agency in improving quality and reducing burden*. Washington, DC: Author.

Council of Chief State School Officers. (2019). *Revising the definition of formative assessment*. (By the Formative Assessment for Students and Teachers [FAST] State Collaborative on Assessment and Student Standards [SCASS]). Washington, DC: Author.

Darling-Hammond, L., Wilhoit, G., & Pittenger, L. (2014). *Accountability for college and career readiness: Developing a new paradigm*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Dweck, C. (2016). *Mindset: The new psychology of success*. New York, NY: Penguin Random House LLC.

Gaertner, M., Conley, D. T., & Stoltz, P. (2016). *On track: Redefining readiness in education and the workplace*. New York: Pearson.

Gaertner, M., & McClarty, K. (2015). Performance, perseverance, and the full picture of college readiness. *Educational Measurement: Issues and Practice, 34*(2), 20–33.

Gaertner, M., & McClarty, K. (2016). The case for a middle-school college-readiness index. *Educational Measurement: Issues and Practice, 35*(3), 35–37.

Gaertner, M., & Roberts, R. (2017). More than a test score: Defining and measuring personal qualities. In K. McClarty, K. Mattern, & M. Gaertner (Eds.), *Preparing students for college and careers: Theory, measurement, and educational practice*. New York, NY: Routledge.

Guindon, M., Huffman, H., Socol, A. R., & Takahashi-Rial, S. (2014). *How much testing is taking place in North Carolina schools at grades K–12? An analysis of federal, state, and local required assessments*. Raleigh, NC: Financial and Business Services Internship Program. Retrieved from http://www.dpi.state.nc.us/docs/intern-research/reports/testing2014.pdf

Hart, R., Casserly, M., Uzzell, R., Palacios, M., Corcoran, A., & Spurgeon, L. (2015). *Student testing in America's great city schools: An inventory and preliminary analysis*. Washington, DC: Council of Great City Schools.

Heritage, M. (2017). Changing the assessment relationship to empower teachers and students. In K. McClarty, K. Mattern, & M. Gaertner (Eds.), *Preparing students for college and careers: Theory, measurement, and educational practice*. New York, NY: Routledge.

Jobs for the Future. (2018). *Ten principles for a high-quality system of assessments*. Boston, MA: Author. Retrieved from https://deeperlearning4all.org/wp-content/uploads/2018/02/10-principles.pdf

Learning Policy Institute. (2019). *North Carolina's statewide accountability system: How to effectively measure progress toward meeting the Leandro tenets*. Palo Alto, CA: Author.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.

North Carolina Department of Public Instruction. (2013). *READY accountability background brief*. Retrieved from http://www.ncpublicschools.org/accountability/reporting/

North Carolina Department of Public Instruction. (2016). *READY accountability background brief*. Retrieved from http://www.ncpublicschools.org/accountability/reporting/

North Carolina Department of Public Instruction. (2017a). *READY accountability background brief*. Retrieved from http://www.ncpublicschools.org/accountability/reporting/

North Carolina Department of Public Instruction. (2017b). *The North Carolina testing program summary report: 2015–16 proof of concept study*. Raleigh, NC: Author.

North Carolina Department of Public Instruction. (2018). *Final North Carolina state plan for the ESSA*. Retrieved from http://www.ncpublicschools.org/succeeds/

North Carolina Standard Course of Study. (2018). Retrieved from http://www.dpi.state.nc.us/curriculum/

North Carolina State Board of Education & North Carolina Department of Public Instruction. (2014). *Understanding the five achievement levels*. Retrieved from http://www.ncpublicschools.org/accountability/policies/briefs/

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

Smithson, J. L. (2015). *A report to the North Carolina Department of Public Instruction on the alignment characteristics of state assessment instruments covering grades 3–8 and high school in mathematics, reading, and science*. Madison, WI: Wisconsin Center for Education Research.

Stiggins, R. (2017). *The perfect assessment system*. Alexandria, VA: ASCD.

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

U.S. Department of Education. (2018). A *state's guide to the U.S. Department of Education's assessment peer review process*. Washington, DC: Author.

Appendix A — Interview Protocols

Interview Protocol and Questions
(North Carolina Department of Public Instruction Staff)

60 minutes semistructured

Introduction. Thank you for participating in today's interview. My name is ——. I work for WestEd. WestEd was appointed by the state's Supreme Court [the Court] as an independent consultant to develop an action plan to ensure all North Carolina children have an equal opportunity to obtain a sound basic education. Your feedback about your experiences will be used to inform the development of the action plan.

Confidentiality. What you tell us will be kept confidential in that we will not use your name/role or the school and district you work in, and we will not attribute any quotes to individuals. All the data we gather from interviews will be synthesized and summarized, and you will not be identifiable in our reports. However, we will be taking notes today. In the meantime, all files will be saved in a secure folder. Only members of the project team will have access to these notes.

Process. We are seeking your experience and observations, and we are also seeking your opinions, so please be as candid as possible. This is an informal session; think of it as a conversation. My role is to ask questions, listen, take notes, and keep you on track. We welcome all ideas, opinions, and points of view. If you are uncomfortable sharing something or think of something later today or tomorrow that you wish you had shared, please feel free to contact me. Are there any questions before we begin?

Purpose. This interview is being conducted to collect information regarding the usefulness of assessments used in North Carolina schools and to elicit ideas for how the state may better support the implementation of a high-quality assessment system. We also want to better understand how the North Carolina accountability and improvement system is used to identify struggling schools and inform school support and improvement efforts.

Statewide Assessments

What are the strengths and limitations of the statewide assessment system? How does the NCDPI support the districts' capacity for using the data from the statewide assessment system for continuous improvement?

» Which statewide assessments do schools find most and least useful for informing instruction and student support? How often are these assessments administered? How are these assessments used?

» What support (e.g., resources, training) does the NCDPI currently provide to improve how schools use assessment to inform teaching and learning and program improvement? What types of support are most helpful? What additional supports do districts and/or schools request?

» What are the challenges and/or obstacles that the NCDPI faces in supporting how schools administer and use assessments to inform teaching and learning and program improvement?

» To what extent are locally developed benchmark or interim assessments used? And to what extent do they provide similar or overlapping information compared with the statewide assessment system?

Accountability System

What are the strengths and limitations of the statewide accountability system? How does the NCDPI support the districts' capacity for using the data from the statewide accountability system for continuous improvement?

» How can the state's accountability system be improved to support continuous improvement and reduce achievement gaps among subgroups?

» What are the biggest changes in moving from the READY Accountability Model to the new accountability model under ESSA? What are the biggest advantages and disadvantages of the new ESSA model? What gaps, if any, in the new model need to be addressed?

» How are accountability results used to inform continuous improvement, particularly for comprehensive and targeted support to schools? (Probe to understand how results are used for each type of school.)

» To what extent does the state's accountability system correctly and reliably identify the schools that need comprehensive or targeted support? What, if any, improvements are needed to improve how schools are identified for comprehensive or targeted support?

» How transparent is the process for assigning school performance grades and identifying schools that need comprehensive or targeted support? How could the system be improved to be more transparent?

Additional Recommendations

» If you were asked to provide the Court with one assessment-related recommendation that would improve educational opportunities for the students of North Carolina, what would it be?

» Do you have any other comments that you would like to share regarding the statewide assessment and accountability systems in North Carolina?

Conclusion

Thank the participant for his/her time. Remind him/her that information will be kept confidential.

Interview Protocol and Questions
(District Staff: Assessment Leads and Superintendents)

60 minutes semistructured

Introduction. Thank you for participating in today's interview. My name is ——. I work for WestEd. WestEd was appointed by the state's Supreme Court [the Court] as an independent consultant to develop an action plan to ensure all North Carolina children have an equal opportunity to obtain a sound basic education. Your feedback about your experiences will be used to inform the development of the action plan.

Confidentiality. What you tell us will be kept confidential in that we will not use your name/role or the school and district you work in, and we will not attribute any quotes to individuals. All the data we gather from interviews will be synthesized and summarized, and you will not be identifiable in our reports. However, we will be taking notes today, and we will record if you do not have an objection. The purpose of the recording is to clarify our notes; we will destroy recordings upon completion of our summary. Further, all files will be saved in a secure folder. Only members of the project team will have access to these notes. Only members of the project team will have access to these notes and recordings. Is it okay for us to record this interview?

Purpose. This interview is being conducted to collect information regarding the usefulness of assessments used in North Carolina schools and to elicit ideas for how the state may better support the implementation of a high-quality assessment system. We also want to better understand how the North Carolina accountability system is used to identify struggling schools and inform district and school improvement efforts.

Process. We are seeking your experience and observations, and we are also seeking your opinions, so please be as candid as possible. This is an informal session; think of it as a conversation. My role is to ask questions, listen, take notes, and keep the conversation on track. We welcome all ideas, opinions, and points of view. If you are uncomfortable sharing something or think of something later today or tomorrow that you wish you had shared, please feel free to contact me. Are there any questions before we begin?

Statewide Assessments

What are the strengths and limitations of the statewide assessment system?

» Which statewide assessments do schools find most and least useful for informing instruction and student support? How often are these assessments administered? How are these assessments used?

» What support (e.g., resources, training) does the NCDPI currently provide to help districts and schools interpret and use assessment data to inform teaching and learning and program improvement? What types of support are most helpful? What additional support do districts and/or schools request?

» To what extent are locally developed benchmark or interim assessments used in your district? How does the district use these assessments? To what extent do they provide similar or overlapping information compared with the statewide assessment system?

Accountability System

What are the strengths and limitations of the statewide accountability system?

» How can the state's accountability system be improved to support continuous improvement and reduce achievement gaps among subgroups?

» What support (e.g., resources, training) does the NCDPI currently provide to help districts and schools interpret and use accountability indicators to inform school improvement? What types of support are most helpful? What additional support would be helpful?

» How does your district use accountability results to inform continuous improvement, particularly for comprehensive and targeted support to schools? (Probe to understand how results are used for each type of school.)

» To what extent does the state's accountability system correctly and reliably identify the schools that need comprehensive or targeted support? What, if any, improvements are needed to improve how schools are identified for comprehensive or targeted support?

» How transparent is the process for assigning school performance grades and identifying schools that need comprehensive or targeted support? How could the system be improved to be more transparent?

Additional Recommendations

» If you were asked to provide the NCDPI with one assessment-related recommendation to improve educational opportunities for the students of North Carolina, what would it be?

» Do you have any other comments that you would like to share regarding the statewide assessment and accountability systems in North Carolina?

Conclusion

Thank the participant for his/her time. Remind him/her that information will be kept confidential.

Appendix B — Principal Survey Questions and Results About the Assessment System

Indicate your agreement with the following statements about your school's data from the state assessments:

	n	Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree
Useful for informing school improvement planning	585	6.7%	10.4%	53.5%	29.4%
Useful for informing professional development needs	587	7.8%	15.3%	53.2%	23.7%
Useful for understanding student strengths	585	8.0%	14.4%	53.9%	23.8%
Useful for understanding student weaknesses	586	7.7%	16.4%	52.2%	23.7%
Useful for understanding the quality of teaching	585	8.2%	25.1%	48.9%	17.8%
Easy to access	586	7.7%	16.9%	55.6%	19.8%
Easy to understand	583	6.0%	19.6%	56.4%	18.0%
Available in a timely manner	582	26.1%	19.8%	41.2%	12.9%

In a typical month, how often do your teachers use interim/benchmark or formative assessment data to do the following?

	n	Less than once a month	Once or twice a month	Weekly or almost weekly	A few times per week
Identify instructional content to use in class	584	6.7%	32.0%	50.3%	11.0%
Tailor instruction to individual students' needs	583	3.4%	19.9%	53.0%	23.7%
Develop recommendations for additional instructional support	584	6.0%	35.5%	48.6%	9.9%
Form small groups of students for targeted instruction	583	5.7%	19.6%	45.8%	29.0%
Discuss an individual student's progress with a parent or guardian	581	14.1%	39.0%	37.7%	9.1%
Provide feedback to an individual or groups of students	583	3.8%	22.6%	50.3%	23.3%
Adjust content or instructional strategies to be taught in subsequent lessons	583	3.1%	18.4%	55.2%	23.3%
Meet with another teacher or specialist (e.g., instructional coach, school psychologist)	583	6.5%	26.6%	50.1%	16.8%